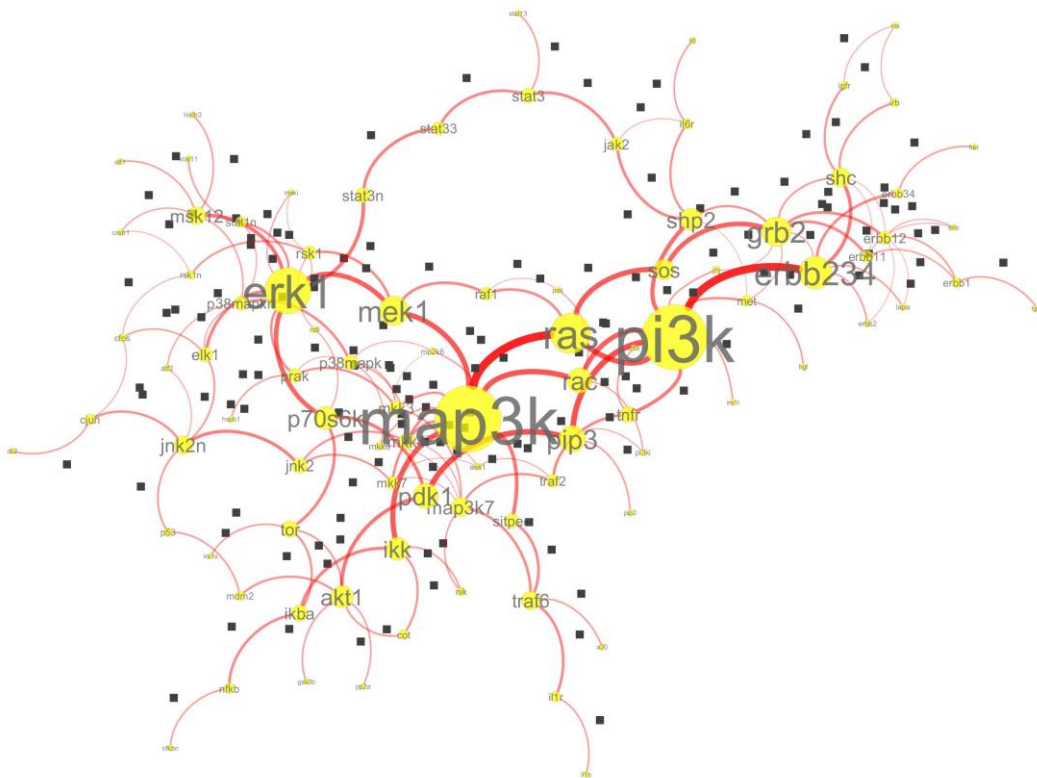




**Εθνικό Μετσόβιο Πολυτεχνείο**  
**Σχολή Μηχανολόγων Μηχανικών**  
Τομέας Μηχανολογικών Κατασκευών και Αυτομάτου Ελέγχου

**Εργαστήριο Εμβιομηχανικής και Συστημικής Βιολογίας**

**Διπλωματική Εργασία**



**«Μοντελοποίηση μονοπατιών μεταγωγής σήματος σε κύτταρα»**

Σταύρακας Βασίλειος

Επιβλέπων: Δρ. Λεωνίδα Αλεξόπουλος

Αθήνα, Οκτώβριος 2013

# Περιεχόμενα

<b>Ευχαριστίες</b> .....	6
<b>1. Περίληψη</b> .....	7
<b>2. Εισαγωγικά</b> .....	8
2.1 Η επιστήμη της Βιοϊατρικής Τεχνολογίας σήμερα .....	8
2.1.1 Βιοϊατρική Μηχανική ( <i>Biomedical Engineering</i> ) .....	10
2.1.2 Βιολογική Μηχανική ( <i>Biological Engineering</i> ) .....	17
2.2 Συστημική Βιολογία ( <i>Systems Biology</i> ) .....	23
2.2.1 Γενική Περιγραφή .....	23
2.2.2 Βασικά Στοιχεία .....	24
2.3 Η επιστήμη της Βιοπληροφορικής ( <i>bioinformatics</i> ) .....	27
2.3.1 Γενική Περιγραφή .....	27
2.3.2 Στόχος .....	29
2.3.3 Ερευνητικές Προσεγγίσεις .....	31
2.3.4 Βασικά Ερευνητικά Πεδία .....	32
2.4 Μονοπάτια Μεταγωγής Σήματος .....	33
2.4.1 Εισαγωγικά .....	33
2.4.2 Πρωτεΐνες .....	35
2.4.3 Σηματοδότηση κυττάρων .....	48
2.5 Μοντελοποίηση μονοπατιών μεταγωγής σημάτων .....	54
2.5.1 Εισαγωγή .....	54
2.5.2 Θεμελιώδης θεωρία δικτύων .....	55
2.5.3 Δίκτυα στη βιολογία .....	59
2.5.4 Προσανατολισμένα δίκτυα .....	62
2.5.5 Προσέγγιση κατά Boole για τη μοντελοποίηση σηματοδότησης δικτύων .....	78
2.5.6 Περιορισμοί μεθόδων .....	89
2.6 Πειραματικά δεδομένα στη συστημική βιολογία .....	91
2.6.1 Πρωτεϊνωματική .....	91
2.6.2 Μετρήσεις πρωτεϊνών .....	93
2.6.3 Τεχνική ELISA .....	100
2.7 Βελτιστοποίηση μονοπατιών σηματοδότησης με υπολογιστικές μεθόδους .....	106
2.7.1 Εισαγωγικά .....	106
2.7.2 Μέθοδος Ακέραιου Γραμμικού Προγραμματισμού για τη μοντελοποίηση μονοπατιών μεταγωγής σημάτων πρωτεϊνών ( <i>ILP</i> ) .....	106
2.7.3 Μέθοδος Μη-Γραμμικού Προγραμματισμού για την ποσοτική μοντελοποίηση μονοπατιών μεταγωγής σημάτων πρωτεϊνών .....	115
<b>3. Αναπτυσσόμενη Μέθοδος</b> .....	120
3.1 Εισαγωγικά .....	121
3.2 Μελέτη προβλήματος .....	122
3.3 Εφαρμοζόμενη μέθοδος .....	125
3.4 Σύγκριση Μεθόδων .....	134
<b>4. Αποτελέσματα</b> .....	136
4.1 Παρουσίαση αποτελεσμάτων .....	136
4.2 Σχολιασμός Αποτελεσμάτων .....	145
4.3 Παρουσίαση Αποτελεσμάτων με χρήση της μεθόδου του Ακέραιου Γραμμικού Προγραμματισμού ( <i>ILP</i> ) .....	149
4.4 Σύγκριση μεθόδων και αποτελεσμάτων .....	153
4.5 Σχολιασμός Παραρτήματος .....	155
<b>5. Ανακεφαλαίωση - Συμπεράσματα</b> .....	156
<b>Παράρτημα</b> .....	160
Main Programm .....	160
Subroutine Path_Construction .....	183
<b>Βιβλιογραφία</b> .....	184
Βιβλία – Παραδόσεις .....	184
Δημοσιεύσεις – Επιστημονικές Διατριβές .....	185
Ιστότοποι .....	187

## Περιεχόμενα Εικόνων

Εικόνα 1: Τομείς Βιοϊατρικής Μηχανικής .....	11 -
Εικόνα 2: Διαδικασία κατασκευής ιστού.....	12 -
Εικόνα 3: Κλάδοι της Εμβιομηχανικής.....	13 -
Εικόνα 4: Κλάδοι Βιολογικών Μετρήσεων .....	15 -
Εικόνα 5: Στάδια παρασκευής φαρμάκου .....	16 -
Εικόνα 6: Τομείς Βιολογικής Μηχανικής.....	18 -
Εικόνα 7: Κλάδοι Βιολογικών Μετρήσεων .....	20 -
Εικόνα 8: Μέθοδοι μέτρησης πρωτεϊνών .....	21 -
Εικόνα 9: Μετάδοση σήματος στο κύτταρο .....	25 -
Εικόνα 10: Μητρώο αναπαράστασης απόκρισης κυττάρου .....	29 -
Εικόνα 11: Ο τομέας της βιοπληροφορικής.....	31 -
Εικόνα 12: Δίκτυα μεταγωγής σημάτων .....	34 -
Εικόνα 13: Παραδείγματα λειτουργίας πρωτεϊνών .....	35 -
Εικόνα 14: Αμινοξέα στον ανθρώπινο οργανισμό .....	36 -
Εικόνα 15: Τμήμα πρωτεΐνης .....	37 -
Εικόνα 16: Σχηματισμός διμερούς πρωτεΐνης .....	38 -
Εικόνα 17: Δομή ενδιάμεσου ινιδίου.....	41 -
Εικόνα 18: Δομή μικροσωληνίσκου .....	42 -
Εικόνα 19: Δομή νηματίου ακτίνης .....	43 -
Εικόνα 20: Μηχανισμός μυϊκής συστολής .....	45 -
Εικόνα 21: Μεταγραφή / μετάφραση .....	47 -
Εικόνα 22: Ενδοκρινής επικοινωνία .....	49 -
Εικόνα 23: Παρακρινής επικοινωνία.....	50 -
Εικόνα 24: Νευρωνική επικοινωνία .....	50 -
Εικόνα 25: Επικοινωνία εξαρτώμενη από επαφή.....	51 -
Εικόνα 26: Διάδοση σήματος στο κύτταρο .....	52 -
Εικόνα 27: Μηχανισμός μετάδοσης σήματος στο κύτταρο .....	53 -
Εικόνα 28: Δύο αναπαραστάσεις του ίδιου γράφου .....	56 -
Εικόνα 29: Ζευγύον δέντρο γράφου .....	57 -
Εικόνα 30: Πίνακας γεινίαςης (adjacency matrix) .....	59 -
Εικόνα 31: Μη προσανατολισμένο δίκτυο .....	60 -
Εικόνα 32: Προσανατολισμένο δίκτυο .....	61 -
Εικόνα 33: Χαρακτηριστικά μεταβολικά δίκτυα.....	62 -
Εικόνα 34: Αναπαράσταση με λίστα γεινίαςης .....	63 -
Εικόνα 35: Ένα προσανατολισμένο δίκτυο .....	63 -
Εικόνα 36: Αναζήτηση κατά βάθος.....	64 -
Εικόνα 37: Αναζήτηση κατά πλάτος .....	66 -
Εικόνα 38: Αναζήτηση κατά βάθος σε ένα μεγαλύτερο δίκτυο .....	67 -
Εικόνα 39: Αναζήτηση κατά πλάτος σε ένα μεγαλύτερο δίκτυο.....	68 -
Εικόνα 40: Δάσος αναζήτησης κατά πλάτος .....	69 -
Εικόνα 41: Αποτέλεσμα κώδικα επίσκεψης.....	70 -
Εικόνα 42: Αρχικά στάδια της μεθόδου του Warshall.....	71 -
Εικόνα 43: Τελικά στάδια της μεθόδου του Warshall .....	71 -
Εικόνα 44: Ένα προσανατολισμένο δίκτυο με βάρος .....	72 -
Εικόνα 45: Αρχικά στάδια του αλγόριθμου του Floyd.....	73 -
Εικόνα 46: Τελικά στάδια του αλγόριθμου του Floyd .....	73 -
Εικόνα 47: Ένα προσανατολισμένο μη-κυκλικό δίκτυο (dag).....	74 -
Εικόνα 48: Αναζήτηση κατά πλάτος σε ένα dag.....	75 -
Εικόνα 49: Προσανατολισμένο δίκτυο και οι ισχυρά συνεκτικές συνιστώσες του .....	77 -
Εικόνα 50: Ένα απλό μοντέλο δικτύου κατά Boole .....	81 -
Εικόνα 51: Τα κύρια βήματα στη δυναμική μοντελοποίηση κατά Boole των βιολογικών συστημάτων.....	83 -
Εικόνα 52: Εφαρμογές της μοντελοποίησης κατά Boole στα βιολογικά συστήματα .....	86 -
Εικόνα 53: Διάγραμμα ροής της μεθοδολογίας Boole .....	87 -
Εικόνα 54: Εξαγωγή μητρώων Boole.....	88 -
Εικόνα 55: Ανακατασκευή δικτύων .....	89 -
Εικόνα 56: " Φάσμα " πρωτεϊνικής ανάλυσης.....	92 -
Εικόνα 57: Μέθοδος Western blot.....	95 -
Εικόνα 58: Βασικά χαρακτηριστικά μεθόδων συνάφειας.....	96 -
Εικόνα 59: Μέθοδος ELISA.....	97 -
Εικόνα 60: Μικροσφαιρίδιο.....	98 -
Εικόνα 61: Σύγκριση μεθόδων .....	99 -
Εικόνα 62: Αναπαράσταση της μεθόδου .....	100 -
Εικόνα 63: Αρχή λειτουργίας ELISA.....	101 -
Εικόνα 64: Αρχή λειτουργίας Multiplexed ELISA.....	102 -

Εικόνα 65: set micro-beads.....	- 103 -
Εικόνα 66: ELISA Μικροσφαιριδίων.....	- 103 -
Εικόνα 67: Υπόδειγμα συστήματος ανίχνευσης .....	- 104 -
Εικόνα 68: Πειραματικό και υπολογιστικό διάγραμμα ροής για την αξιολόγηση της επίδρασης των φαρμάκων.....	- 108 -
Εικόνα 69: Συγκεκριμένη κυτταρική τοπολογία με χρήση Ακέραιου Γραμμικού Προγραμματισμού ..	- 114 -
Εικόνα 70: Βελτιστοποίηση ενός μεσαίας κλίμακας μοντέλου με χρήση της NLP μεθόδου.....	- 119 -
Εικόνα 71: Παράδειγμα απεικόνισης τοπολογίας δικτύου .....	- 123 -
Εικόνα 72: Παράδειγμα πίνακα εξάρτησης δεδομένων .....	- 124 -
Εικόνα 73: Συνοπτική παρουσίαση εξεταζόμενου προβλήματος .....	- 124 -
Εικόνα 74: Παράδειγμα Κόμβου Σύγκρουσης.....	- 130 -
Εικόνα 75: Παράδειγμα αδυναμίας ένωσης κόμβων μέσω διαδρομής.....	- 131 -
Εικόνα 76: Τοπολογία δικτύου σχετικά μικρής κλίμακας.....	- 137 -
Εικόνα 77: Απεικόνιση δικτύου με το λογισμικό Cytoscape .....	- 138 -
Εικόνα 78: Στατιστικά στοιχεία δικτύου .....	- 138 -
Εικόνα 79: Ιστόγραμμα Shortest Paths .....	- 139 -
Εικόνα 80: Απεικόνιση κατανομής ακμών στο δίκτυο .....	- 140 -
Εικόνα 81: Εξαγόμενο τελικό δίκτυο.....	- 141 -
Εικόνα 82: Εξαγόμενο τελικό δίκτυο 2.....	- 142 -
Εικόνα 83: Τελικό εξαγόμενο δίκτυο 3.....	- 143 -
Εικόνα 84: Πειραματικά Δεδομένα .....	- 145 -
Εικόνα 85: Απεικόνιση τελικού δικτύου με χρήση μεθόδου ILP.....	- 149 -
Εικόνα 86: Τελικό δίκτυο με ελαχιστοποίηση της αντικειμενικής συνάρτησης.....	- 151 -
Εικόνα 87: Τελικό δίκτυο με μεγιστοποίηση της αντικειμενικής συνάρτησης.....	- 152 -

*Αφιερωμένο στην Αγγελική  
και στον Αποστόλη*

## **Ευχαριστίες**

Η παρούσα διπλωματική εργασία σηματοδοτεί το τέλος των πενταετών σπουδών μου στη σχολή Μηχανολόγων Μηχανικών ΕΜΠ. Με την ευκαιρία θα ήθελα να ευχαριστήσω όλους όσους με βοήθησαν το τελευταίο αυτό εξάμηνο για να φέρω εις πέρας και να ολοκληρώσω τη διπλωματική μου εργασία.

Αρχικά, θα ήθελα να στείλω χαιρετισμούς και πολλή αγάπη στα παιδιά του εργαστηρίου «Εμβιομηχανικής και Συστημικής Βιολογίας» του τμήματος Μηχανολόγων Μηχανικών του Εθνικού Μετσόβιου Πολυτεχνείου, για την υποστήριξη, τις συμβουλές και τη βοήθεια που μου έχουν παρέχει όλο αυτόν τον καιρό. Ιδιαίτερα, όμως, θα ήθελα να ευχαριστήσω το διδάκτορα Ιωάννη Μελά για την καθοδήγηση και υποστήριξή του. Η συμβολή του στην περάτωση του παρόντος ήταν ανεκτίμητη. Ακόμα, θα ήθελα να ευχαριστήσω πάρα πολύ το Θοδωρή Σακελλαρόπουλο για τις πολύτιμες συμβουλές, την παρέα και το χρόνο που αφιέρωσε για να με βοηθήσει στα υπολογιστικά κομμάτια.

Επίσης, νιώθω την ανάγκη να ευχαριστήσω τον επίκουρο καθηγητή, κύριο Λεωνίδα Αλεξόπουλο για την καθοδήγηση και την υποστήριξή του τον τελευταίο ενάμιση χρόνο. Η ευκαιρία που μου έδωσε να γίνω μέλος του εργαστηρίου ήταν πολυτιμότερη και με βοήθησε σημαντικά να προσθέσω χρήσιμες και σημαντικές γνώσεις στο δυναμικό μου. Η πόρτα του ήταν πάντα ανοιχτή για μένα για απορίες και διευκρινίσεις και οι πάντα εύστοχες παρατηρήσεις του και η παρότρυνση του για συστηματική και στοχευόμενη δουλειά με έκαναν καλύτερο.

Τέλος, θα ήθελα να ευχαριστήσω τους γονείς μου Αγγελική και Αποστόλη, την οικογένειά μου Θανάση, Αρετή, Άννα, Βαγγέλη, Θώμη και Πέτρο και τους φίλους μου, την άλλη μου οικογένεια, Ευγενία, Κωνσταντίνα, Κατερίνα, Χρήστο, Δημήτρη και Αννέζα για την χρόνια υποστήριξή τους και πίστη τους σε μένα. Χωρίς αυτούς τίποτα από όλα αυτά δεν θα είχε γίνει. Χωρίς αυτούς δεν θα είχα καταφέρει να πιστέψω στον εαυτό μου, όταν πραγματικά το χρειάστηκα. Η παρούσα διπλωματική αφιερώνεται σε όλους τους με απεριόριστη αγάπη και πολλά ευχαριστώ.

---

## 1. Περίληψη

Η μοντελοποίηση των μονοπατιών μεταγωγής σημάτων μεταξύ των κυττάρων είναι υψίστης σημασίας για την κατανόηση της λειτουργίας τους. Ο όρος αυτός αναφέρεται στη διαδικασία αναγνώρισης των σχέσεων που διέπουν τη διάδοση των σημάτων από τη μία πρωτεΐνη στην άλλη, εξηγώντας τελικώς τρόπους με τους οποίους τα κύτταρα αντιδρούν και αποκρίνονται σε παράγοντες του βιοχημικού τους μικρό-περιβάλλοντος. Τα μονοπάτια σηματοδότησης αποτελούνται από ένα σύνολο πρωτεϊνικών αλληλεπιδράσεων, οι οποίες αναγνωρίζονται μέσω πρωτεϊνικών διαδικασιών υψηλής απόδοσης και διατίθενται μέσω διαδικτυακών βάσεων δεδομένων. Κατά τη διάρκεια των τελευταίων χρόνων έχουν αναπτυχθεί αρκετές μέθοδοι για τη μοντελοποίηση τέτοιων μονοπατιών αποσκοπώντας στην περιγραφή των λειτουργιών των κυττάρων και στην επεξήγηση της σχέσης αιτιότητας που διακρίνει τις ασθένειες και τα συμπτώματά τους. Κύριος στόχος των μεθόδων αυτών είναι η μοντελοποίηση των βιολογικών δικτύων σηματοδότησης ως λογικά μοντέλα και η χρήση συμβατικών μορφών βελτιστοποίησης (Ακέραιος Γραμμικός Προγραμματισμός – ILP και Μη – Γραμμικός Προγραμματισμός – NLP) συνδυάζοντάς τις με φωσφοπρωτεομικά δεδομένα υψηλής απόδοσης για τη δημιουργία μοντέλων ικανών να προβλέψουν τους μηχανισμούς σηματοδότησης του εξεταζόμενου κυτταρικού τύπου. Η παρούσα εργασία αποστασιοποιείται από τις προαναφερθείσες μεθόδους και τη λογική της δομημένης βελτιστοποίησης και στοχεύει στην ανάπτυξη ενός συνοπτικού κώδικα, ο οποίος θα μπορεί να δέχεται τα παραπάνω δεδομένα και να εξάγει την τελική τοπολογία (κατά βέλτιστο ανά περίπτωση τρόπο), η οποία θα ανταποκρίνεται στα δεδομένα των μετρήσεων, λαμβάνοντας ταυτόχρονα υπ' όψιν της τους υπάρχοντες περιορισμούς που υπαγορεύονται από τη θεωρία της βιολογίας, τη διεθνή βιβλιογραφία και τα υπάρχοντα δεδομένα στο διεθνές δίκτυο βάσεων δεδομένων.

---

## 2. Εισαγωγικά

Τη σημερινή εποχή, για την εξέλιξη της έρευνας απαιτείται η συνεργασία επιστημόνων από διαφορετικά πεδία. Χαρακτηριστικό παράδειγμα άρτιας συνεργασίας μεταξύ δύο επιστημών αποτελεί η Βιοϊατρική Τεχνολογία, όπου η έρευνα στην ιατρική προωθείται με τη βοήθεια της μηχανολογίας. Η παρούσα επιστήμη διαιρείται σε αρκετούς κλάδους, αλλά αυτός που θα μας απασχολήσει στη παρούσα διπλωματική εργασία είναι αυτός της Συστημικής Βιολογίας και πιο συγκεκριμένα αυτός της Βιοπληροφορικής. Γενικά, ο συγκεκριμένος ερευνητικός τομέας, τα τελευταία χρόνια, κατέχει ένα ιδιαίτερα σημαντικό και ενδιαφέρον ρόλο στη διεθνή επιστημονική κοινότητα με πολλές προοπτικές εξέλιξης και πολλά πεδία εφαρμογής. Ο συνδυασμός των θεωρητικών γνώσεων της ιατρικής με τις πρακτικές εφαρμογές των γνώσεων της μηχανολογίας τον καθιστούν ιδιαίτερα χρήσιμο για την εξαγωγή κατάλληλων και χρήσιμων αποτελεσμάτων στην ανάπτυξη φαρμάκων και στην καταπολέμηση ασθενειών και παθήσεων.

### 2.1 Η επιστήμη της Βιοϊατρικής Τεχνολογίας σήμερα

Η *βιοϊατρική τεχνολογία* είναι μια επιστήμη με ιδιαίτερα ευρύ φάσμα εφαρμογών που επιδέχεται πληθώρα ορισμών. Τόσο στα ελληνικά, όσο και στη διεθνή ορολογία, υπάρχουν πολλαπλές ορολογίες που περιγράφουν ιδιαίτερα συγγενή πεδία, τα οποία μάλιστα συχνά αλληλεπικαλύπτονται. Για το λόγο αυτό οι ορισμοί και οι περιγραφές δεν είναι ιδιαίτερα αυστηροί και δεν τοποθετούνται στεγανά, αλλά κάθε επιστημονική δραστηριότητα που κινείται στα πλαίσια της εφαρμογής των μεθόδων της επιστήμης του μηχανικού σε βιολογικά συστήματα μπορεί να ενταχθεί στον τομέα της βιοϊατρικής τεχνολογίας.

Ως ορισμό για την βιοϊατρική τεχνολογία (biotechnological engineering/bioengineering) μπορούμε να αναφέρουμε εδώ ότι είναι η επιστήμη η οποία εφαρμόζει το σκεπτικό και τις μεθόδους της επιστήμης της βιολογίας (και δευτερευόντως της φυσικής, των μαθηματικών, της χημείας και της πληροφορικής) σε συνδυασμό με τις μεθοδολογίες ανάλυσης και σύνθεσης της επιστήμης του μηχανικού για να λύσει προβλήματα που σχετίζονται με βιολογικά συστήματα ([7]).

Σκοπός της βιοϊατρικής τεχνολογίας είναι να συνδράμει στην κατανόηση των βασικών αρχών λειτουργίας των βιολογικών συστημάτων και στην ανάπτυξη αποδοτικών τεχνολογιών, βασισμένων στη βιολογία, για να καλύψει έναν ευρύ φάσμα κοινωνικών αναγκών. Οι ανάγκες αυτές μπορεί να αφορούν τους τομείς της διάγνωσης, πρόληψης και θεραπείας ασθενειών, την ανάπτυξη νέων υλικών, συσκευών και διαδικασιών ακόμα και την αντιμετώπιση ευρύτερων περιβαλλοντικών προβλημάτων.

Δεδομένου ότι εντός της επιστήμης αυτής συμπεριλαμβάνεται ο τρόπος σκέψης και η προσέγγιση από την πλευρά του μηχανικού, στις συσκευές, τις



μεθόδους και τις διαδικασίες που αναπτύσσονται είναι σημαντικό να εμπεριέχεται πάντα και η προσπάθεια για την εύρεση της πλέον αποδοτικής και με το χαμηλότερο κόστος λύσης του προβλήματος. Ακόμα, ο μηχανικός γενικά δεν εστιάζει σε βάθος στην επιστήμη της βιολογίας, αλλά μοντελοποιώντας το πρόβλημα που του δίνεται, μπορεί να αντιμετωπίζει το εξεταζόμενο κάθε φορά βιολογικό σύστημα ως ένα μαύρο κουτί.

Γενικά, η συνδρομή του μηχανικού στην επίλυση βιολογικών προβλημάτων είναι ακριβώς η δυνατότητά του να απομακρύνεται από την πραγματική φύση του προβλήματος και να το αντιμετωπίζει εφαρμόζοντας μεθοδολογίες από τα πεδία της φυσικής και των μαθηματικών. Αυτό μπορεί να φαίνεται αρχικά ως απλουστευση, αλλά στην πραγματικότητα είναι ένα κρίσιμο πλεονέκτημα. Είναι γεγονός ότι οι γνώσεις μας πάνω στα φυσικά φαινόμενα και οι δυνατότητές μας να αναπτύξουμε μαθηματικά μοντέλα για να τα περιγράψουμε βρίσκονται σε υψηλό επίπεδο. Αντίθετα, η ικανότητα μας να εξηγήσουμε και να μοντελοποιήσουμε φαινόμενα που αφορούν βιολογικά συστήματα και έμβιους οργανισμούς είναι ακόμα σε σχετικά πρωτόγονο στάδιο. Έτσι, η εφαρμογή του τρόπου σκέψης ενός μηχανικού στα προβλήματα αυτά, μπορεί να επιφέρει (και έχει επιφέρει τα τελευταία χρόνια που εφαρμόζεται) ταχύτατη ανάπτυξη ενός νέου πεδίου γνώσεων σε βιολογικές και ιατρικές εφαρμογές.

Σε μια προσπάθεια να απλουστεύσουμε την παρουσίαση των πεδίων που περικλείονται από την έννοια της βιοϊατρικής τεχνολογίας, μπορούμε να διακρίνουμε τους δύο παρακάτω επιστημονικούς κλάδους ([7]):

- **Βιοϊατρική Μηχανική (Biomedical Engineering)**
- **Βιολογική Μηχανική (Biological Engineering)**

Η διαφοροποίηση ανάμεσα στη βιοϊατρική μηχανική και τη βιολογική μηχανική δεν είναι ξεκάθαρη και οι δύο κλάδοι συχνά αλληλεπικαλύπτονται, ενώ πολλές δημοσιεύσεις χρησιμοποιούν τους δύο όρους ως ταυτόσημους. Γενικά, η βιοϊατρική μηχανική έχει ως επίκεντρο την ιατρική και εφαρμόζει γνώσεις από την βιολογία και άλλες επιστήμες στην επίλυση ιατρικών προβλημάτων. Από την άλλη πλευρά η βιολογική μηχανική εστιάζει κυρίως σε προβλήματα και ερωτήματα που ανακύπτουν μέσα από την ανάπτυξη της ίδιας της επιστήμης της βιολογίας και σκοπό έχει την εύρεση νέας γνώσης, είτε αυτή βρίσκει εφαρμογή στην ιατρική, είτε όχι. Έτσι αναγνωρίζουμε ότι κανένας από τους δύο κλάδους δεν εμπεριέχει τον άλλο πλήρως, αφού στη βιοϊατρική μηχανική περιλαμβάνονται και μη βιολογικές μέθοδοι που αναπτύσσονται για ιατρικούς σκοπούς, ενώ αντίθετα στη βιολογική μηχανική περιλαμβάνονται όλες οι βιολογικές μέθοδοι, ακόμα και αν δεν βρίσκουν εφαρμογή στην ιατρική.

Η ταξινόμηση που γίνεται στους κλάδους αυτούς έχει καθαρά φιλοσοφική έννοια και δεν περιορίζει τα πεδία δράσης του κάθε ερευνητή. Στη συνέχεια ο κάθε κλάδος διαιρείται και πάλι σε διάφορα ερευνητικά πεδία, ανάλογα με τον τομέα εφαρμογής αυτή τη φορά. Στον κλάδο της βιοϊατρικής μηχανικής

υπάρχουν πολλά πεδία που σκοπό έχουν να αντιμετωπίσουν τα διάφορα προβλήματα, όπως αυτά τίθενται στην ιατρική. Από την άλλη πλευρά, ο κλάδος της βιολογικής μηχανικής έχει λιγότερους τομείς, οι οποίοι με τη σειρά τους αλληλεπικαλύπτονται σε πολλά σημεία αφού σε κάθε περίπτωση σκοπός είναι η απόκτηση γνώσης πάνω στη λειτουργία βιολογικών συστημάτων, είτε αυτά είναι μικροσκοπικά οργάνδια των κυττάρων, είτε ολόκληροι έμβιοι οργανισμοί.

Παρακάτω και για λόγους πληρότητας θα πραγματοποιηθεί μία συνοπτική παρουσίαση των επιμέρους αυτών κατηγοριών.

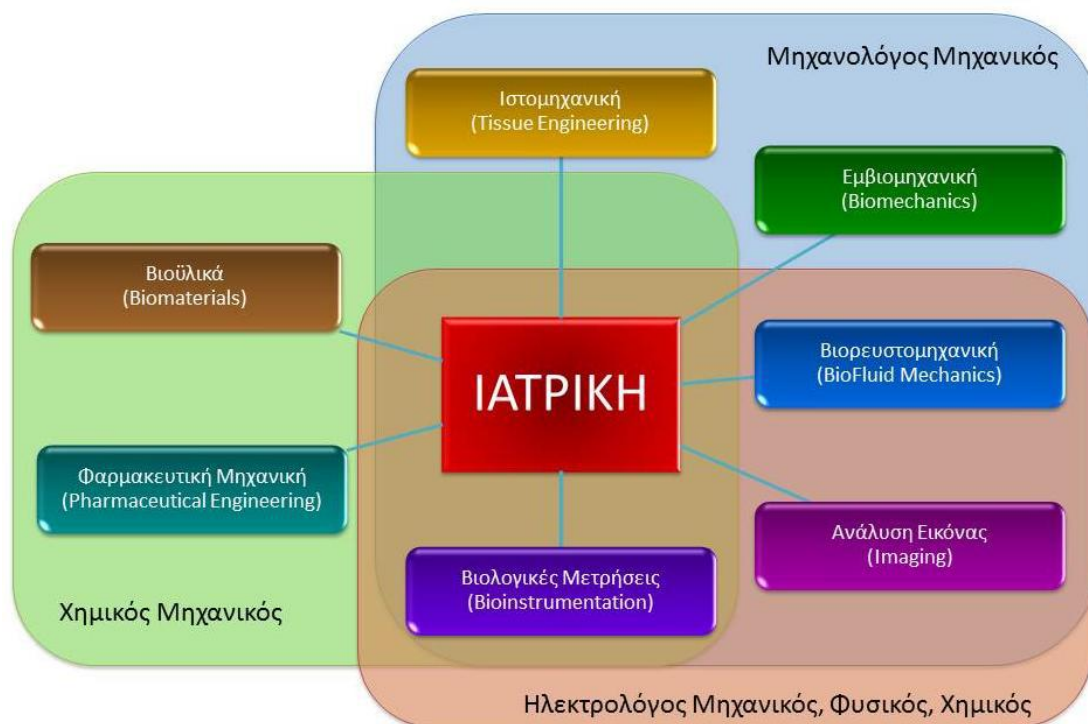
### **2.1.1 Βιοϊατρική Μηχανική (Biomedical Engineering)**

Ο κλάδος της βιοϊατρικής μηχανικής τοποθετεί στο κέντρο του την ιατρική. Αυτό σημαίνει πως ανεξάρτητα με τις μεθόδους και τις διαδικασίες που χρησιμοποιούνται σκοπός είναι πάντοτε να επιλυθούν προβλήματα της ιατρικής. Δηλαδή, από την πλευρά της θεώρησης αυτής, μας ενδιαφέρει να αντιμετωπίσουμε ιατρικά γεγονότα και να βελτιώσουμε την ποιότητα ζωής των ανθρώπων, χωρίς κατ' ανάγκη να αναλύουμε και να εξηγούμε τις αιτίες του προβλήματος που αντιμετωπίζουμε ή την ίδια τη φύση του προβλήματος ([4]).

Για παράδειγμα, από τη σκοπιά της βιοϊατρικής μηχανικής, ενδιαφερόμαστε να αναπτύξουμε τεχνητά όργανα τα οποία θα μπορούν να αντικαταστήσουν όργανα του οργανισμού που έχουν υποστεί βλάβη. Σε αυτή τη διαδικασία μπορεί να εμπλακεί και η προσπάθεια κατανόησης του τρόπου λειτουργίας του ίδιου του οργάνου και των παραγόντων που επέφεραν τη φθορά του, αλλά μπορεί και όχι, αφού σκοπός μας είναι πρωταρχικά να λύσουμε το πρόβλημα, ακόμα και αν δεν καταφέρουμε να κατανοήσουμε όλες τις πτυχές του ([4]).

Η βιοϊατρική μηχανική χωρίζεται περαιτέρω σε επιμέρους τομείς. Οι τομείς αυτοί εμπίπτουν στα ενδιαφέροντα διαφόρων επιστημόνων και μηχανικών. Φυσικά η διάκριση αυτή δεν είναι δεσμευτική και όπως ήδη ειπώθηκε δεν υπάρχουν στεγανά, αλλά γίνεται με βάση τους γενικότερους και συνηθέστερους τομείς δραστηριότητας κάθε επαγγέλματος. Επίσης είναι προφανές πως σε διάφορους τομείς υπάρχουν επικαλύψεις και έτσι δεν είναι καθόλου σπάνια η συνεργασία επιστημόνων από διάφορους κλάδους για την επίλυση ενός προβλήματος βιοϊατρικής ([4]).

Οι διάφοροι τομείς της βιοϊατρικής μηχανικής, καθώς και οι κλάδοι στους οποίους γενικά εμπίπτουν, παρουσιάζονται στο παρακάτω γράφημα.



**Εικόνα 1: Τομείς Βιοϊατρικής Μηχανικής ([7])**

Στη συνέχεια θα προχωρήσουμε σε συνοπτική παρουσίαση των επιμέρους τομέων της Βιοϊατρικής Μηχανικής. Σκοπός μας αυτή τη στιγμή είναι να πληροφορήσουμε συνοπτικά τον αναγνώστη για τις δυνατότητες και τις εφαρμογές που υπάρχουν αυτή τη στιγμή, αλλά και για τις τάσεις που εμφανίζονται για το μέλλον ([7]).

### **Ιστομηχανική (Tissue Engineering)**

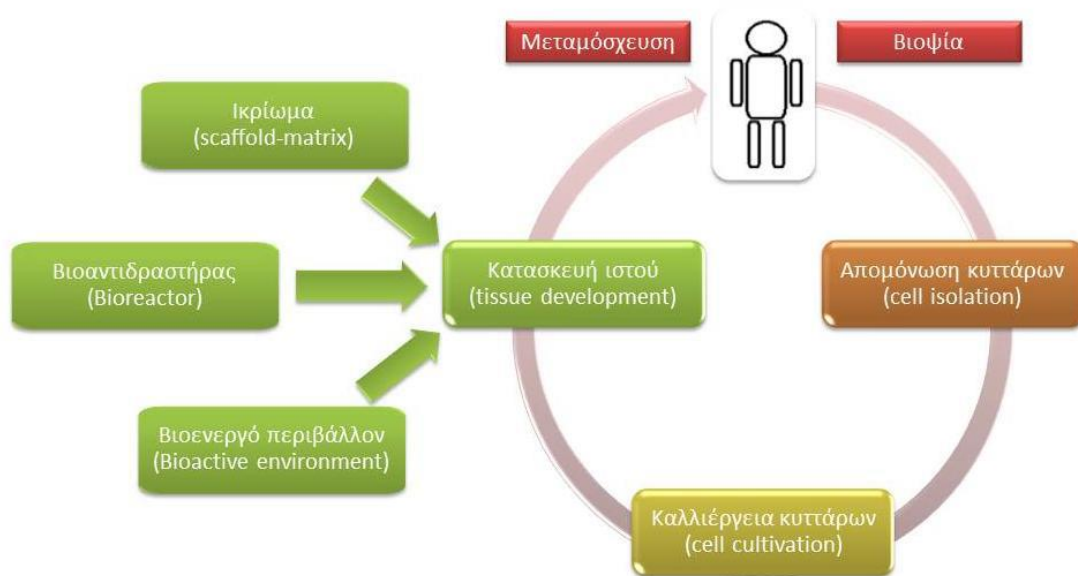
Ο τομέας αυτός αποτελούσε αρχικά μια υποκατηγορία των βιοϋλικών αλλά λόγω της μεγάλης του ανάπτυξης και της σημασίας του θεωρείται πλέον ανεξάρτητος τομέας. Η συμβολή του τομέα αυτού στην ιατρική είναι μεγάλη, αφού τελικός σκοπός της ιστομηχανικής είναι να κατασκευαστεί βιολογικός ιστός για την αντικατάσταση αρχικών κατεστραμμένων ιστών ([7]).

Σημασία έχει να παρατηρήσουμε ότι ο τομέας αυτός δεν σχετίζεται με την κατασκευή τεχνητών μελών ή γενικά υποκατάστατων του ανθρώπινου σώματος που δεν παράγονται από βιολογικά υλικά. Αντίθετα, αφορά την κατασκευή ιστών με τεχνικό τρόπο έξω από το σώμα (in vitro) και στη συνέχεια εμφύτευση αυτών στο σώμα (in vivo) του ασθενούς.

Για να είναι δυνατή η κατασκευή βιολογικού ιστού σε συνθήκες εργαστηρίου, απαιτείται η συνεργασία τεσσάρων βασικών παραγόντων ([7]):

- Κύτταρα (cells)
- Ικρίωμα (Scaffold – Matrix)
- Βιο-ενεργό περιβάλλον (Bioactive environment)
- Βιο-αντιδραστήρας (Bioreactor)

Οι παραπάνω βασικοί παράγοντες συνδυάζονται έτσι ώστε αφού πραγματοποιηθεί βιοψία και εντοπισουμε τον ιστό που έχει υποστεί βλάβη στον οργανισμό, να κατασκευαστεί ο νέος ιστός στο εργαστήριο και στη συνέχεια να μεταμοσχευθεί στον ασθενή.



**Εικόνα 2: Διαδικασία κατασκευής ιστού ([7])**

Αυτή τη στιγμή οι έρευνες επικεντρώνονται στην κατασκευή δέρματος, χόνδρου, μεσοσπονδύλιων δίσκων, αγγείων, καρδιακών βαλβίδων και ήπατος.

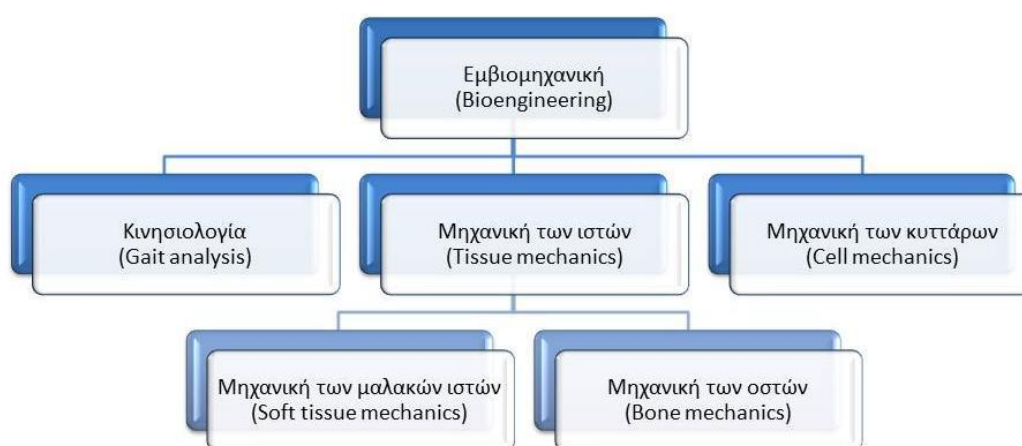
### **Εμβιομηχανική (Biomechanics)**

Ο τομέας της εμβιομηχανικής ασχολείται κατά κύριο λόγο με την ανάπτυξη εμφυτευμάτων και τεχνητών μελών, αλλά και με την ερμηνεία βιολογικών συστημάτων με βάση παραδοσιακά μοντέλα της μηχανικής. Ο τομέας της εμβιομηχανικής έχει τη μεγαλύτερη συνάφεια με τη μηχανολογία αφού πολλές φορές, απλά με χρήση της Νευτώνειας μηχανικής και της επιστήμης των υλικών, μπορούμε να λάβουμε προσεγγιστικά συμπεράσματα για διάφορα βιολογικά συστήματα ([7]).

Συχνά, σε προβλήματα της εμβιομηχανικής χρησιμοποιείται ανάλυση μηχανισμών, στατική, κινηματική ή δυναμική μηχανική. Βέβαια τα βιολογικά συστήματα είναι συνήθως πολύ πιο πολύπλοκα από αυτά που κατασκευάζει ο άνθρωπος και για το λόγο αυτό σχεδόν πάντα χρησιμοποιούνται

υπολογιστικές μέθοδοι. Αυτό σημαίνει πως ακόμα και για σχετικά απλά προβλήματα πρέπει να γίνουν επαναληπτικά πολλά μοντέλα, τα οποία θα επιλυθούν και στη συνέχεια θα συσχετιστούν με πειραματικά δεδομένα έως ότου η προσέγγιση να είναι ικανοποιητική.

Γενικά μπορούμε να πούμε ότι ο τομέας της εμβιομηχανικής είναι ένας τομέας όπου βρίσκουν άμεση εφαρμογή πολλές μεθοδολογίες της μηχανολογίας. Επίσης είναι ένα πεδίο ιδιαίτερα ευρύ και οι εφαρμογές του μπορεί να κυμαίνονται από την βελτίωση της ποιότητας ζωής ασθενών με κινητικά (κατά κύριο λόγο) προβλήματα, μέχρι τη βελτιστοποίηση των επιδόσεων αθλητών. Στη συνέχεια ακολουθεί ένα διάγραμμα με τις κατηγορίες που διακρίναμε παραπάνω.



Εικόνα 3:Κλάδοι της Εμβιομηχανικής ([7])

### **Βιορρευστομηχανική (BioFluid Mechanics)**

Ο τομέας της βιορρευστομηχανικής ασχολείται με την μελέτη της ροής βιολογικών ρευστών μέσα σε ζωντανούς οργανισμούς. Το πεδίο αυτό έχει ιδιαίτερα μεγάλη σημασία, γιατί η ροή των βιολογικών ρευστών σε ένα οργανισμό καθορίζει σε μεγάλο βαθμό και την κατάσταση των ιστών διαμέσου των οποίων πραγματοποιείται η ροή ([7]).

Στη βιορρευστομηχανική μελετώνται οι αρχές της ροής βιολογικών ρευστών σε ζωντανούς ιστούς και πώς οι αλλαγές στη ροή επηρεάζουν τόσο το ίδιο το ρευστό, όσο και τον ιστό τον οποίο αυτό διαρρέει. Τα ρευστά που μελετώνται είναι κυρίως το αίμα, το νερό και ο αέρας και στη συνέχεια τα λοιπά σωματικά υγρά των οργανισμών. Στην περίπτωση των φυτών μελετάται η ροή των χυμών τους. Αντικείμενο μελέτης αποτελεί η κίνηση και η ισορροπία των δυνάμεων, τόσο σε ακίνητα, όσα και ρέοντα ρευστά μέσα στο σώμα.

Τα ευρήματα της βιορρευστομηχανικής βρίσκουν εφαρμογή στην φαρμακευτική, στα βιοϋλικά, σε νέες μεθόδους διάγνωσης καθώς και στη χειρουργική. Βασικότερο αντικείμενο μελέτης είναι η ροή του αίματος και πώς

αυτή επηρεάζει και επηρεάζεται από στενώσεις, φράξεις αγγείων και δυσλειτουργίες των καρδιακών βαλβίδων. Χαρακτηριστικό δείγμα των δυσκολιών που εμφανίζονται στον τομέα αυτό είναι ότι μακροσκοπικά το αίμα θεωρείται ως συνεχές και μη συμπιεστό Νευτώνειο ρευστό, όμως όταν εξετάζουμε τη ροή του σε στενώσεις και τριχοειδή αγγεία η παραπάνω θεώρηση δεν ισχύει.

### **Ανάλυση εικόνας (Imaging)**

---

Η ανάλυση εικόνας ή ιατρική ανάλυση εικόνας (medical imaging) αναφέρεται στο πεδίο εκείνο που σκοπό έχει την εύρεση τεχνικών και διαδικασιών για τη δημιουργία απεικονίσεων του σώματος, ή ορισμένων οργάνων, ή μελών αυτού. Η απεικόνιση αυτή σχετίζεται άμεσα με ιατρικούς σκοπούς όπως η διάγνωση και αντιμετώπιση ασθενειών ή η ανάλυση της φυσιολογίας του ατόμου ([7]).

Σαν γενικό πεδίο, η ανάλυση εικόνας συμπεριλαμβάνει την ακτινολογία (radioscopy), την ενδοσκόπηση (endoscopy), την πυρηνική ιατρική (nuclear medicine), τη θερμογραφία (thermography), την υπερηχογραφία (ultra sound scan), την ιατρική φωτογραφία και τις εφαρμογές μικροσκοπίων (microscopy) γενικά για βιολογικούς και ιατρικούς σκοπούς.

Τα θέματα που κυρίως απασχολούν την ανάλυση εικόνας σήμερα είναι ([7]):

- ***Η αύξηση του όγκου των δεδομένων που χρησιμοποιούνται***
- ***Η δημιουργία τρισδιάστατων απεικονίσεων υψηλής ακρίβειας***
- ***Η μείωση του όγκου αποθήκευσης των δεδομένων***
- ***Χρήση μεθόδων απεικόνισης κατά τη διάρκεια κλινικών μελετών φαρμάκων***

### **Βιολογικές μετρήσεις (Bioinstrumentation)**

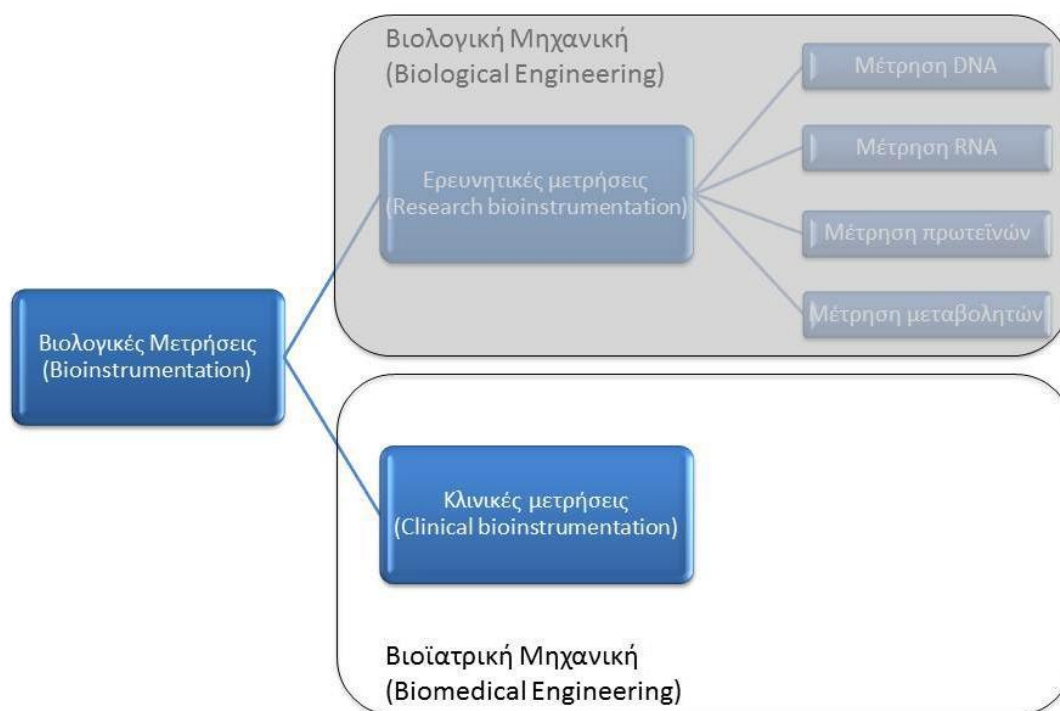
---

Οι βιολογικές μετρήσεις (Bioinstrumentation) περιλαμβάνουν τις συσκευές, τα συστήματα και τις διαδικασίες με χρήση των οποίων πραγματοποιούνται μετρήσεις βιολογικών σημάτων. Τα βιολογικά σήματα μπορεί να ποικίλουν από τους καρδιακούς παλμούς και το ρυθμό της αναπνοής μέχρι τη μέτρηση του DNA και των πρωτεϊνών σε ένα κύτταρο ([7]).

Ο τομέας αυτός αποτελεί ένα γιγαντιαίο πεδίο με πολλαπλές εφαρμογές που απαιτεί γνώσεις από πολλούς διαφορετικούς επιστημονικούς κλάδους. Για την κατασκευή λοιπόν ενός συστήματος βιολογικών μετρήσεων πιθανώς να απαιτούνται γνώσεις γύρω από τα ψηφιακά ηλεκτρονικά κυκλώματα, τον αυτόματο έλεγχο, τα συστήματα ανίχνευσης σημάτων, τη βιοσυμβατότητα υλικών, τη βιολογία, τη μηχανική, τη χημεία, τα μαθηματικά, την οπτική κ.α.

Γενικότερος σκοπός είναι να συνδυαστούν οι απαιτούμενες αρχές από τα διάφορα επιστημονικά πεδία, έτσι ώστε να γίνει δυνατή η μέτρηση με επιτυχία και ακρίβεια ενός βιολογικού σήματος.

Δεδομένου ότι σε αυτή την παράγραφο εξετάζουμε τη Βιοϊατρική Μηχανική, θα εστιάσουμε στην παρουσίαση των κλινικών μετρήσεων. Αντίστοιχα, οι ερευνητικές μετρήσεις θα παρατεθούν στην επόμενη παράγραφο, ως υποκατηγορία της Βιολογικής Μηχανικής.



Εικόνα 4:Κλάδοι Βιολογικών Μετρήσεων ([7])

Στην κατηγορία των κλινικών μετρήσεων ανήκουν όλα εκείνα τα όργανα και συστήματα που σκοπό έχουν τη διάγνωση ή περίθαλψη ασθενειών. Τα συστήματα αυτά μπορούμε να τα συναντήσουμε σε νοσοκομεία, κέντρα περίθαλψης, διαγνωστικά κέντρα ή ακόμα και στα σπίτια μας και τόσο οι γιατροί όσο και οι ασθενείς έρχονται σε τακτική επαφή μαζί τους.

### **Φαρμακευτική Μηχανική (Pharmaceutical Engineering)**

Ο τομέας αυτός είναι ιδιαίτερα καινούριος και κατά βάση αποτελεί κλάδο της φαρμακευτικής. Σκοπό έχει την ανάπτυξη και παραγωγή προϊόντων, διαδικασιών και δραστικών ουσιών χρήσιμων στη φαρμακευτική βιομηχανία. Όπως και άλλοι τομείς της Βιοϊατρικής Μηχανικής, απαιτεί γνώσεις και συνεργασία διάφορων επιστημονικών κλάδων ([7]).

Οι μηχανικοί, οι οποίοι απασχολούνται στη φαρμακευτική βιομηχανία, εμπλέκονται στη σύλληψη, το σχεδιασμό, την κατασκευή και τη λειτουργία τόσο των εγκαταστάσεων έρευνας, όσο και των εργοστασίων παραγωγής της

φαρμακευτικής βιομηχανίας. Απαραίτητη είναι η συνεργασία με επιστήμονες άλλων τομέων, όπως είναι οι γιατροί, οι χημικοί και οι φαρμακοποιοί.

Ένας μηχανικός μπορεί να εφαρμόσει τις γνώσεις του και να απασχοληθεί σε διάφορα στάδια της παραγωγής φαρμάκων και δραστικών ουσιών. Στην παρακάτω εικόνα παρουσιάζονται συνοπτικά αυτά τα στάδια.



Εικόνα 5:Στάδια παρασκευής φαρμάκου ([7])

## Βιοϋλικά (Biomaterials)

Ως βιοϋλικό ονομάζουμε κάθε ουσία, επιφάνεια ή κατασκευή που αλληλεπιδρά με βιολογικά συστήματα. Η επιστήμη που ασχολείται με τη μελέτη και την ανάπτυξη βιοϋλικών έχει ιστορία σχεδόν 50 χρόνων και έχουν γίνει μεγάλες επενδύσεις στον τομέα αυτόν. Στην ανάπτυξη των βιοϋλικών εμπλέκονται οι επιστήμες της ιατρικής, της βιολογίας, της χημείας καθώς και η επιστήμη των υλικών και η εμβιομηχανική ([7]).

Τα βιοϋλικά είναι, είτε συνθετικές ουσίες που παράγονται στο εργαστήριο, είτε φυσικές ουσίες. Συνηθέστερα χρησιμοποιούνται για ιατρικούς σκοπούς και παίζουν σημαντικό ρόλο σε εφαρμογές εμβιομηχανικής και ιστομηχανικής.

Σε ότι αναφορά τις εφαρμογές τους, τα βιοϋλικά μπορεί να συναντώνται σε απλές εφαρμογές τις οδοντιατρικής ή σε σύνθετα και επεξεργασμένα υλικά που μεταμοσχεύονται στο σώμα του ασθενούς. Λόγω της ιδιότητάς τους να αλληλεπιδρούν με βιολογικά συστήματα, το κριτήριο που εξετάζουμε πάντα για τα βιοϋλικά είναι η βιοσυμβατότητά τους.

Παραδείγματα βιοϋλικών και εφαρμογών βιοϋλικών είναι ([7]):

- **Αποκατάσταση αρθρώσεων.**
- **Πλάκες υποστήριξης οστών.**
- **Τεχνητοί τένοντες.**
- **Οδοντικά εμφυτεύματα.**
- **Καρδιακές βαλβίδες.**
- **Υποκατάστατα δέρματος.**
- **Φακοί επαφής.**
- **Προσθετική στήθους.**



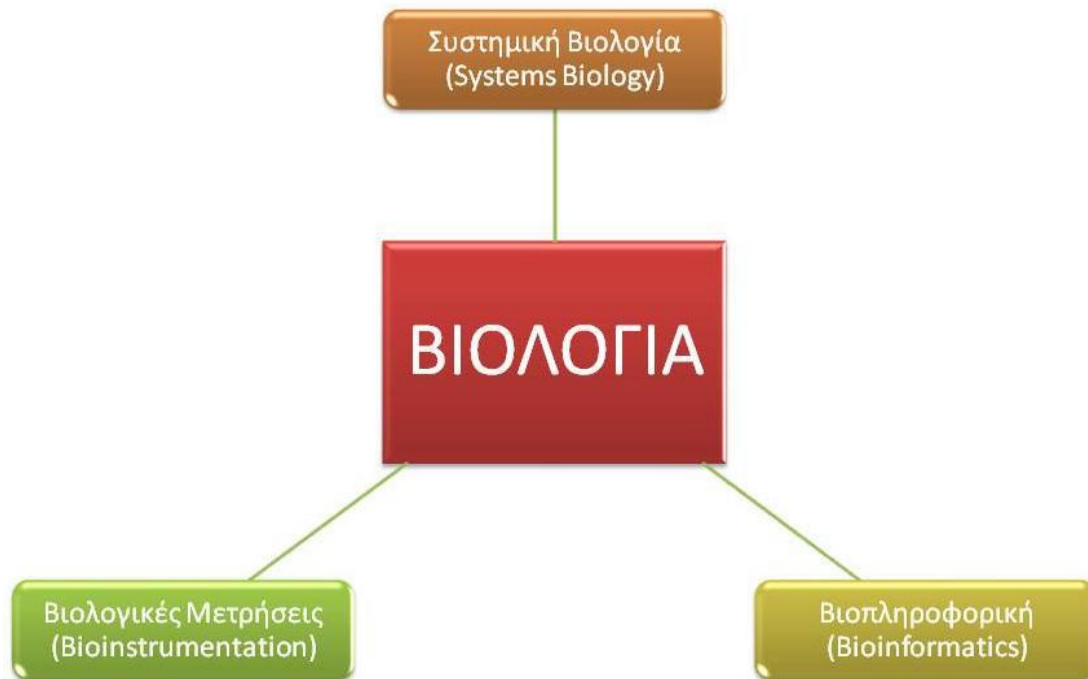
Μια άλλη σημαντική κατηγορία βιοϋλικών αποτελούν τα βιοπολυμερή. Αυτά είναι πολυμερή τα οποία συντίθενται από ζωντανούς οργανισμούς. Παραδείγματα αποτελούν η χλωροφύλλη, η γλυκόζη, οι πρωτεΐνες, το DNA κ.α. Ένα βασικό χαρακτηριστικό των βιοπολυμερών ([7]) είναι ότι τα περισσότερα είναι βιοδιασπώμενα, κάνοντας τα ιδανικά για εφαρμογές φιλικές προς το περιβάλλον ([7]).

### **2.1.2 Βιολογική Μηχανική (Biological Engineering)**

Ο κλάδος της βιολογικής μηχανικής τοποθετεί στο κέντρο του την βιολογία και γενικά την έρευνα σε βιολογικά πεδία. Σκοπός είναι να μελετηθούν και να αποκρυπτογραφηθούν όσο βαθύτερα γίνεται τα διάφορα βιολογικά συστήματα. Πρωτεύον αντικείμενο μελέτης αποτελεί ο άνθρωπος, ενώ ακόμα και μελέτες που γίνονται πάνω σε άλλους οργανισμούς στοχεύουν στην ερμηνεία των συστημάτων του ανθρώπινου οργανισμού ([7]).

Βέβαια και στην περίπτωση μελέτης των βιολογικών συστημάτων, απώτερος σκοπός είναι πάντα η αντιμετώπιση ασθενειών και κινδύνων από τους οποίους απειλείται ο άνθρωπος. Όμως επειδή στην έρευνα δεν τίθεται ένας συγκεκριμένος στόχος αντιμετώπισης ενός ιατρικού συμβάντος, αλλά περισσότερο ελπίζουμε ότι μέσα από την καλύτερη κατανόηση των βιολογικών συστημάτων θα καταφέρουμε να δώσουμε λύσεις σε υπάρχοντα και μελλοντικά προβλήματα, για το λόγο αυτό η βιολογική μηχανική παρουσιάζεται ως ξεχωριστός κλάδος από την βιοϊατρική μηχανική.

Ο κλάδος της βιολογικής μηχανικής εξετάζει γενικά όλη την έρευνα που αφορά την βιολογία και γίνεται με χρήση εργαλείων του μηχανικού. Για την καλύτερη παρουσίαση των εννοιών με τις οποίες θα ασχοληθούμε στη συνέχεια, κρίνεται σκόπιμο να διακρίνουμε τρεις τομείς. Δεδομένου ότι η βιολογική μηχανική αποτελεί έναν κλάδο ερευνητικό, ο διαχωρισμός της σε τομείς δεν γίνεται με βάση τον αντικειμενικό σκοπό ή το πεδίο ενδιαφέροντος. Αντίθετα διαχωρίζουμε και εξετάζουμε τους παρακάτω τομείς με βάση τη σκοπιά από την οποία προσεγγίζουν την ανάλυση των βιολογικών συστημάτων, καθώς και με βάση τα εργαλεία που χρησιμοποιούν.



Εικόνα 6: Τομείς Βιολογικής Μηχανικής ([7])

Στη συνέχεια θα προχωρήσουμε σε συνοπτική παρουσίαση των επιμέρους τομέων της βιολογικής μηχανικής. Σκοπός μας αυτή τη στιγμή είναι να πληροφορήσουμε συνοπτικά τον αναγνώστη για τις δυνατότητες και τις εφαρμογές που υπάρχουν, αλλά και για τις τάσεις που εμφανίζονται για το μέλλον.

Σημειώνουμε, πως αυτή η κατηγορία είναι που θα μας απασχολήσει ιδιαίτερα και στην παρούσα διπλωματική.

### **Συστημική Βιολογία (Systems Biology)**

Ο τομέας αυτός έχει σκοπό την προσέγγιση και ανάλυση των διάφορων βιολογικών λειτουργιών με βάση δομημένα συστήματα. Η σημασία του τομέα αυτού αυξάνεται ραγδαία τα τελευταία χρόνια, οπότε με την προσέγγιση που ακολουθείται αποκομίζουμε όλο και περισσότερες κρίσιμες πληροφορίες ([7]).

### **Βιοπληροφορική (Bioinformatics)**

Η βιοπληροφορική είναι ο τομέας στον οποίο εφαρμόζονται αρχές και γνώσεις από τον τομέα της πληροφορικής και της επιστήμης των υπολογιστών στη βιολογία. Ανάμεσα σε άλλα, η βιοπληροφορική πραγματεύεται αλγόριθμους, βάσεις δεδομένων, τεχνολογίες δικτύων, τεχνητή νοημοσύνη, εύρεση πληροφοριών (data mining), επεξεργασία εικόνας, μοντελοποίηση και προσομοίωση, επεξεργασία σήματος, διακριτοποίηση, θεωρία ελέγχου,

θεωρία κυκλωμάτων και στατιστική για την αποτελεσματική επίλυση των βιολογικών προβλημάτων ([7]).

Περισσότερα θα αναφερθούν σε παρακάτω ενότητα.

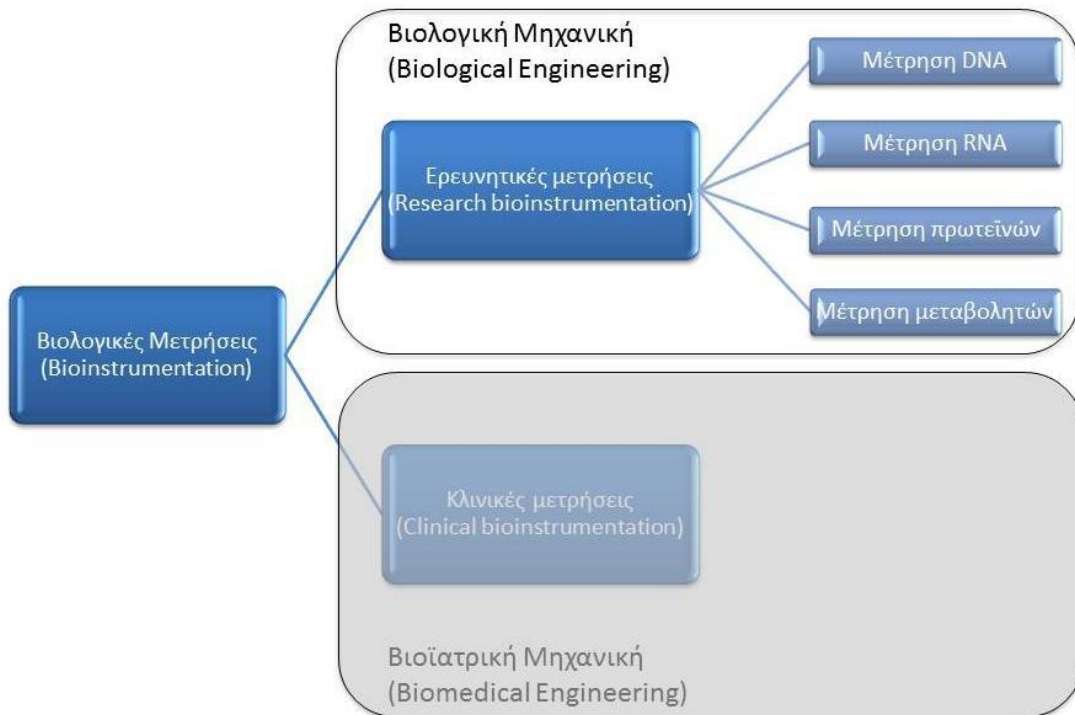
### **Βιολογικές Μετρήσεις (Bioinstrumentation)**

Όπως είδαμε παραπάνω, οι βιολογικές μετρήσεις (Bioinstrumentation) περιλαμβάνουν τις συσκευές, τα συστήματα και τις διαδικασίες με χρήση των οποίων πραγματοποιούνται μετρήσεις βιολογικών σημάτων. Τα βιολογικά σήματα μπορεί να ποικίλουν από τους καρδιακούς παλμούς και το ρυθμό της αναπνοής μέχρι τη μέτρηση του DNA και των πρωτεϊνών σε ένα κύτταρο ([9]).

Στο σημείο αυτό, που εστιάζουμε πλέον στη βιολογική μηχανική, θα εξετάσουμε περαιτέρω τον κλάδο των ερευνητικών μετρήσεων (Research bioinstrumentation), αφού τις κλινικές μετρήσεις τις έχουμε ήδη αναφέρει στη βιοϊατρική μηχανική.

Οι ερευνητικές μετρήσεις έχουν ως κυρίαρχο στόχο την αποκρυπτογράφηση των βασικών βιολογικών συστημάτων και την παροχή του απαραίτητου όγκου πληροφοριών στους ερευνητές προς επεξεργασία και μελέτη. Στην περίπτωση των ερευνητικών μετρήσεων βασικό αντικείμενο μελέτης δεν είναι πλέον ο ασθενής, αλλά η γενίκευση συμπερασμάτων με βάση πληροφορίες από ασθενή και υγιή άτομα ([7]).

Τις ερευνητικές μετρήσεις μπορούμε να τις αναλύσουμε περαιτέρω με βάση το αντικείμενο της μέτρησης. Δεδομένου ότι οι ερευνητικές μετρήσεις σχετίζονται στενά με την αποκρυπτογράφηση του ανθρώπινου DNA και των συμπεριφορών των βιολογικών συστημάτων που εξαρτώνται από την έκφραση των γονιδίων, έχουμε τις εξής κατηγορίες ([7]).



Εικόνα 7:Κλάδοι Βιολογικών Μετρήσεων ([7])

- Μέτρηση DNA (Genomics): Σκοπός είναι η μέτρηση και αποκρυπτογράφηση του DNA ως σύνολο (Genome sequencing) ή η εύρεση συγκεκριμένων αλυσίδων των βάσεων του DNA (arrays). Η μέτρηση του DNA είναι στατική, καθώς το DNA δεν μεταβάλλεται με το χρόνο και είναι σχεδόν το ίδιο σε κάθε κύτταρο ενός οργανισμού.
- Μέτρηση RNA (Transcriptomics): Η μέτρηση του RNA γενικά είναι δύσκολη καθώς αυτό διαλύεται πολύ γρήγορα. Για το λόγο αυτό, αντί του RNA μετράμε το cDNA (συμπληρωματικό DNA – complementary DNA), το οποίο δημιουργείται από RNA έτοιμο προς μετάφραση (mRNA – messenger RNA). Η μέτρηση του RNA είναι συνάρτηση του χρόνου και εξαρτάται από την εκάστοτε λειτουργία για την οποία δημιουργείται το RNA.
- Μέτρηση πρωτεϊνών (Proteomics): Η μέτρηση των πρωτεϊνών αποτελεί επίσης μέτρηση εξαρτημένη από το χρόνο. Παρουσιάζει μεγάλο ενδιαφέρον γιατί ουσιαστικά μας αποκαλύπτει άμεσα όλες τις λειτουργίες ενός κυττάρου, αλλά έχει και ορισμένες ιδιαιτερότητες ([7]):

- i. Ενώ το κύτταρο έχει γενικά ένα DNA, υπάρχουν πάρα πολλές πρωτεΐνες, αλλά και πρωτεΐνες που αλλάζουν μορφή ή συνενώνονται σχηματίζοντας νέες πρωτεΐνες με την πάροδο του χρόνου.
- ii. Μεταξύ πρωτεϊνών μέσα στο κύτταρο μπορεί να παρουσιάζονται και αναλογίες 1/1.000.000. Δηλαδή κάποιες πρωτεΐνες μπορεί να βρίσκονται σε πληθώρα και άλλες σε πολύ μικρές ποσότητες, δύσκολα μετρήσιμες.
- iii. Δεν γνωρίζουμε την πρωταρχική μορφή όλων των πρωτεϊνών στον οργανισμό, όπως δηλαδή γνωρίζουμε τα νουκλεοτίδια ως πρωταρχική πηγή του DNA. Πολλές πρωτεΐνες είναι γνωστές μόνο σε τελική ή ενδιάμεση μορφή, αφού αλλάξουν σχήμα ή συνενωθούν με άλλες πρωτεΐνες.
- iv. Δεν μπορούμε να αναπαράγουμε πρωτεΐνες, όπως δηλαδή «αντιγράφουμε» και αναπαράγουμε το DNA.

Γενικά, οι τεχνικές μέτρησης πρωτεϊνών ταξινομούνται ανάλογα με το αν γνωρίζουμε ή όχι τη δομή της εκάστοτε πρωτεΐνης που μετράμε. Μερικά παραδείγματα φαίνονται στο παρακάτω σχήμα.



**Εικόνα 8:Μέθοδοι μέτρησης πρωτεϊνών ([7])**

- Μέτρηση μεταβολητών (Metabolomics): Οι μεταβολήτες (metabolites) είναι ενδιάμεσα και τελικά προϊόντα του μεταβολισμού, ενώ ο όρος χρησιμοποιείται κυρίως για μικρά μόρια. Ένας πρωταρχικός μεταβολήτης συμμετέχει άμεσα στη διαδικασία της ανάπτυξης, της εξέλιξης και της αναπαραγωγής ενός βιολογικού συστήματος με βασικό παράδειγμα τις αλκοόλες. Άλλοι μεταβολήτες μπορεί να είναι αμινοξέα, νουκλεοτίδια, αντιοξειδωτικά, οργανικά οξέα, πολυόλες ή βιταμίνες.

Σε επόμενο κεφάλαιο θα εστιάσουμε περισσότερο στις μεθόδους μέτρησης πρωτεϊνών και θα τις αναλύσουμε περαιτέρω.

## 2.2 Συστημική Βιολογία (Systems Biology)

Η συστημική βιολογία είναι ένας τομέας που αναπτύχθηκε πολύ πρόσφατα, αλλά βρίσκει όλο και περισσότερους υποστηρικτές. Η γενική ιδέα ανάπτυξης σχετίζεται με την ερμηνεία διάφορων βιολογικών ζητημάτων, σε συμφωνία πάντοτε με τη θεωρία των συστημάτων. Η έννοια των συστημάτων χρησιμοποιείται εδώ και πολλά χρόνια από πολλές και διαφορετικές επιστήμες ισχυρό μαθηματικό υπόβαθρο. Έτσι, λοιπόν, η εισαγωγή μίας τέτοιας έννοιας στο επιστημονικό πεδίο της βιολογίας έδωσε μία νέα ώθηση στη σύλληψη και στην κατανόηση δομών και λειτουργιών ([1]).

Η αντίληψη και η επεξεργασία των πειραματικών διατάξεων, των πειραματικών μετρήσεων και γενικά ολόκληρων των πειραματικών διαδικασιών, είναι συχνά δύσκολη και αβέβαιη όταν δεν είναι πλήρως κατανοήσιμος ο μηχανισμός που παράγει τα αποτελέσματα αυτά. Σε ό,τι αφορά την περίπτωση της βιολογίας, οι περισσότερες λειτουργίες των έμβιων οργανισμών παραμένουν άγνωστες ακόμα και μετά από τόσα χρόνια παρατήρησης και μελέτης. Αντίθετα, προσπαθούμε να μαντέψουμε το «πώς» συνέβη κάτι με βάση διάσπαρτα ευρήματα, μετρήσεις μόνο συγκεκριμένων ποσοτήτων και εξωτερικά χαρακτηριστικά του φαινοτύπου του οργανισμού. Η προσπάθεια αυτή, εμπεριέχει, συχνά, την ανάδυση ανά περιόδους διάφορων θεωριών, οι οποίες βασίζονται σε ήδη γνωστά στοιχεία, αλλά περαιτέρω έρευνες διαψεύδουν ή περιορίζουν τη γενικότητα τους ([39]).

Η συστημική βιολογία, λοιπόν, έρχεται σαν νεοσύστατο επιστημονικό πεδίο δράσης, να συνδράμει στην εύρεση των αιτιών με βάση τα αποτελέσματα. Σήμερα, είναι δυνατόν να πραγματοποιηθούν με μεγάλη ταχύτητα μετρήσεις βιολογικών ποσοτήτων, παράγοντας μεγάλο όγκο δεδομένων και αποτελεσμάτων, τα οποία όμως συχνά μένουν ανεκμετάλλευτα λόγω των όσων αναφέρθηκαν στην παραπάνω παράγραφο. Στο αδιέξοδο αυτό παρέχει λύση η συστημική βιολογία, η οποία ενσωματώνει εύστοχα τα εξαγόμενα δεδομένα σε μαθηματικές θεωρίες και μοντέλα. Έτσι, η θεωρία της δομής και της λειτουργίας του συστήματος, επιβεβαιώνεται και ενισχύεται από τα δεδομένα, ή απορρίπτεται, οπότε και οδηγούμαστε σε νέες ανακαλύψεις και θεωρήσεις, οι οποίες με τη σειρά τους μένει να επαληθευτούν ([1]).

### 2.2.1 Γενική Περιγραφή

Όπως έχει ήδη αναφερθεί, η συστημική βιολογία έχει το χαρακτηριστικό της συνεργασίας επιστημόνων από διάφορα ερευνητικά και επιστημονικά πεδία, καθώς για τη δόμηση και επαλήθευση κάθε θεωρίας απαιτούνται γνώσεις από πεδία, όπως η βιολογία, τα μαθηματικά, η φυσική, η πληροφορική, όπως, επίσης, και άλλοι επιμέρους τομείς ([10]).

Η μεγάλη διαφοροποίηση στη διερεύνηση ενός βιολογικού συμβάντος από τη σκοπιά της συστημικής βιολογίας, έγκειται στο ότι δεν εξετάζουμε μία απλή συσχέτιση συμβάντων. Αντίθετα, με βάση τις πληροφορίες και τα δεδομένα, που παράγονται, δομείται ολόκληρο το βιολογικό σύστημα. Στο σύστημα αυτό ελέγχεται η επικοινωνία και η διαδραστικότητα μεταξύ των μερών, καθώς και η

αποτελεσματικότητα που έχει η ενεργοποίηση κάθε μέρους ξεχωριστά για το ευρύ σύνολο.

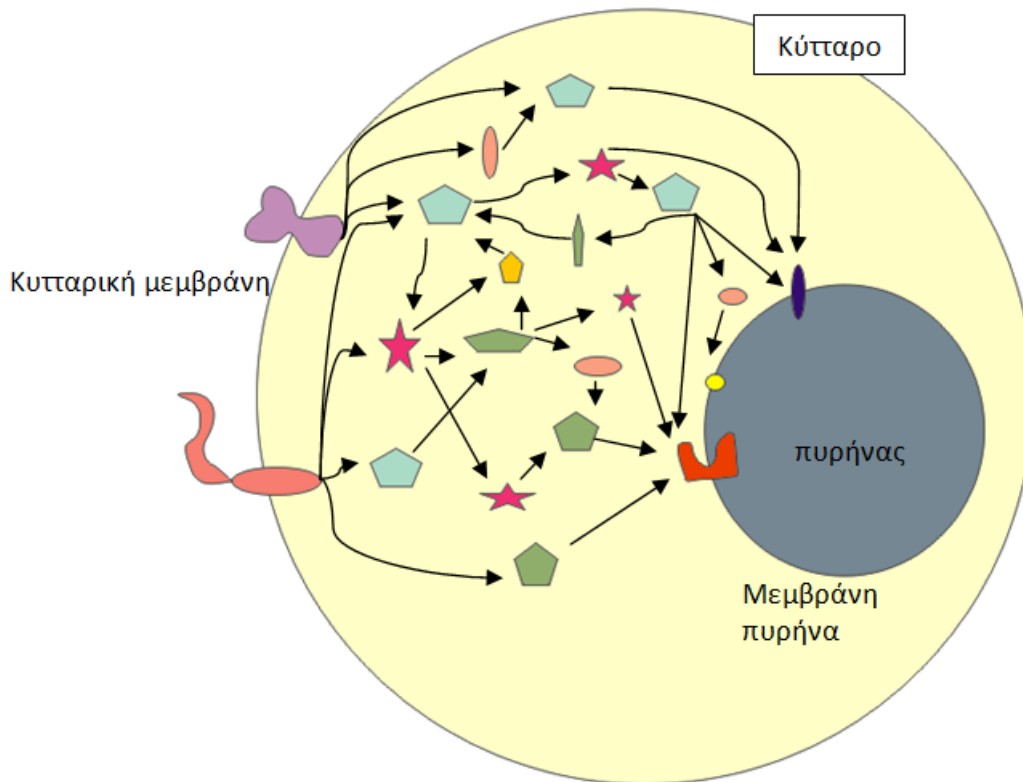
Για να γίνουν πιο «εύπεπτες» στον αναγνώστη οι έννοιες που μέχρι τώρα περιγράψαμε, θα προχωρήσουμε σε μία συνοπτική παρουσίαση κάποιων βασικών στοιχείων της επιστήμης της συστημικής βιολογίας. Φυσικά, είναι προφανές πως ο πλήρης και αδιαμφισβήτητος ορισμός του πεδίου αυτού παραμένει φιλοσοφικός και συμβολίζει περισσότερο έναν τρόπο σκέψης νοοτροπίας και προσέγγισης του επιστημονικού κλάδου της βιολογίας, παρά μία συγκεκριμένη εφαρμογή ([1]).

### **2.2.2 Βασικά Στοιχεία**

Το χαρακτηριστικό των βιολογικών συστημάτων είναι ότι απαρτίζονται από πάρα πολλά επιμέρους στοιχεία (ιστοί, κύτταρα, οργανίδια, πρωτεΐνες, αμινοξέα κ.λπ.) τα οποία αλληλεπιδρούν και επικοινωνούν σε διάφορες κλίμακες μεταξύ τους. Στο παρελθόν η δυνατότητα μας να πραγματοποιούμε βιολογικά πειράματα ήταν περιορισμένη και συνήθως δίναμε στο σύστημα ένα συγκεκριμένο ερέθισμα και μετρούσαμε την ύπαρξη ή όχι μιας συγκεκριμένης απόκρισης. Έτσι καταλήγαμε σε απευθείας συσχέτιση αιτίας και αποτελέσματος και θεωρούσαμε τη μετάδοση της πληροφορίας μέσα στο σύστημα άμεση και ευθύγραμμη ([39]).

Στη συνέχεια, καθώς οι δυνατότητές μας να πραγματοποιούμε βιολογικά πειράματα εξελίχθηκαν, μπορούσαμε να μετρήσαμε ταυτόχρονα πολλές αποκρίσεις, θέτοντας παράλληλα πολλά ερεθίσματα. Έτσι καταλήξαμε στη διαπίστωση ότι τα βιολογικά συστήματα, συμπεριλαμβανομένου του κυττάρου που αποτελεί βασικό αντικείμενο μελέτης, αποτελούνται από πολύπλοκα δίκτυα επικοινωνίας (networks) μεταξύ των επιμέρους στοιχείων. Αυτό σημαίνει ότι η ενεργοποίηση ενός στοιχείου στο σύστημα μπορεί να έχει ως αποτέλεσμα την ενεργοποίηση και πολλών άλλων, ενώ για να επιτύχουμε μια συγκεκριμένη απόκριση, ίσως χρειαστεί να δώσουμε έναν πολύπλοκο συνδυασμό ερεθισμάτων. Επίσης, όπως και στα περισσότερα δίκτυα, συχνά συναντάμε και βρόχους ανάδρασης που ελέγχουν τη λειτουργία του συστήματος. Συμπεραίνουμε λοιπόν ότι ένα στοιχείο του συστήματος σπάνια μπορεί να ελέγχει εξ ολοκλήρου μια βιολογική λειτουργία και αντίστροφα, κάθε στοιχείο του συστήματος μπορεί να συμμετέχει σε πολλές λειτουργίες ([10]).





Εικόνα 9: Μετάδοση σήματος στο κύτταρο ([10])

Οι νέες τεχνολογίες που αναπτύσσονται στο πεδίο των βιολογικών μετρήσεων προσφέρουν όλο και περισσότερες μεμονωμένες πληροφορίες. Ενώ λοιπόν σκοπός μας είναι να καταρτίσουμε ένα χάρτη, όπως η παραπάνω εικόνα, που να περιγράφει πλήρως τη λειτουργία ενός βιολογικού συστήματος, αυτό γίνεται πάρα πολύ δύσκολο εξαιτίας της πολυπλοκότητας του συστήματος. Επομένως, βρισκόμαστε στο στάδιο όπου η τεχνολογία επιτρέπει τη μαζική μέτρηση αποκρίσεων σε διάφορα ερεθίσματα και μπορεί να παραχθεί ένας μεγάλος όγκος δεδομένων ([10]).

Από την άλλη πλευρά η αποκρυπτογράφηση των πληροφοριών είναι δύσκολη και σίγουρα δεν μπορεί να γίνει με τις παραδοσιακές μεθόδους της βιολογίας. Εδώ λοιπόν εμφανίζεται η αξία του τομέα της συστημικής βιολογίας, όπου για τη μοντελοποίηση της λειτουργίας των βιολογικών συστημάτων χρησιμοποιούνται εργαλεία όχι μόνο της βιολογίας, αλλά και των μαθηματικών, της φυσικής, της πληροφορικής και της επιστήμης του μηχανικού ([10]).

Με τη χρήση λοιπόν των νέων τεχνολογιών μέτρησης και διεξάγοντας πολυάριθμα πειράματα αποσκοπούμε στην εύρεση του δικτύου, με βάση το οποίο πραγματοποιούνται οι λειτουργίες ενός βιολογικού συστήματος. Με τον τρόπο αυτό θα μπορούσαμε στη συνέχεια να αποφασίσουμε σε πιο σημείο του δικτύου μπορούμε να επέμβουμε, έτσι ώστε να προκαλέσουμε μια συγκεκριμένη απόκριση στο σύστημα ή και να εμποδίσουμε μία άλλη. Αν και η ολοκληρωτική επίτευξη αυτού του στόχου είναι ιδιαίτερα αμφίβολη και μακροπρόθεσμη, η προσπάθεια που γίνεται έχει αλλάξει ήδη ριζικά τον τρόπο που αντιμετωπίζονται ορισμένα θέματα και υπάρχει ελπίδα να συμβάλει στην

εξήγηση και την αντιμετώπιση κυρίως του καρκίνου, αλλά και άλλων ασθενειών ([10]).

Από την πλευρά του μηχανικού υπάρχει μεγάλο ενδιαφέρον και ευρύ πεδίο εφαρμογής, αφού οι γνώσεις πάνω στη λειτουργία των δικτύων, αλλά και η αναλυτική και συνθετική ικανότητα βοηθούν σημαντικά στον τομέα αυτό. Επίσης, σε μια μακροπρόθεσμη ιδεατή περίπτωση, σκοπός είναι να μοντελοποιηθεί πλήρως ως δίκτυο, όχι μόνο η λειτουργία απλών σχετικά βιολογικών συστημάτων, όπως ένα κύτταρο, αλλά και η λειτουργία ολόκληρου του ανθρώπινου οργανισμού ([1]).

## **2.3 Η επιστήμη της Βιοπληροφορικής (bioinformatics)**

Ο τομέας της βιοπληροφορικής σχετίζεται άμεσα με την ανάγκη δημιουργίας και συντήρησης διαδικτυακών βάσεων δεδομένων, χρήσιμων για την αποθήκευση πληροφοριών από την αρχή της γενετικής επανάστασης (genomic revolution) με σκοπό την πλήρη αποκρυπτογράφηση του ανθρώπινου DNA. Ακόμα, με σκοπό τη συσχέτιση των αλληλεπιδράσεων μεταξύ των διάφορων στοιχείων ενός βιολογικού συστήματος, είναι αναγκαία η διασύνδεση όλων των υπάρχοντων δεδομένων, ερεθισμάτων και αποκρίσεων ([42]).

### **2.3.1 Γενική Περιγραφή**

Το ερευνητικό πεδίο της Βιοπληροφορικής προϋπήρχε της τεράστιας έκρηξης στη συλλογή των γονιδιωματικών, κυρίως, αλλά και των πληθυσμιακών και οικολογικών πληροφοριών και είχε αρχίσει να αναπτύσσεται από τις αρχές της δεκαετίας του 1970. Αρκετές από τις σημερινές κατευθύνσεις στον τομέα αυτό βασίζονται σε θεμέλια που ήδη είχαν τεθεί από την περίοδο εκείνη. Ο πλούτος και η ποικιλομορφία των πληροφοριών που διατίθενται στις μέρες μας και χρήζουν ανάλυσης και επεξεργασίας, έδωσαν νέα ώθηση και προεκτάσεις στο πεδίο αυτό, το οποίο αποτελεί, σε αρκετές περιπτώσεις την αιχμή του δόρατος στη βασική αλλά και εφαρμοσμένη έρευνα των βιολογικών – βιοϊατρικών επιστημών ([53]).

Χαρακτηριστικό φαινόμενο καθ' όλη την ιστορία του κλάδου της βιοπληροφορικής έχει αποτελέσει η «στρατολόγηση» ειδικών επιστημόνων από διάφορα γνωστικά αντικείμενα (Βιολογία, Πληροφορική, Μαθηματικά, Φυσική, Χημεία κλπ.) με κοινό παρονομαστή τη χρήση μαθηματικών μεθόδων και υπολογιστικών τεχνικών για την περιγραφή και ανάλυση Βιολογικών Συστημάτων. Η ολοένα αυξανόμενη πολυπλοκότητα των προς ανάλυση δεδομένων και η ποικιλία τους καθιστά επιτακτική τη συνεισφορά και συνεργασία ειδικών από όλα τα παραπάνω πεδία, εκπαιδευμένων κατάλληλα με βάση τις γενικότερες αρχές και τη μεθοδολογία της σύγχρονης Βιοπληροφορικής, ώστε να είναι δυνατόν να ανταπεξέλθουν στις αυξημένες απαιτήσεις του πεδίου στην έρευνα και στην παραγωγή ([53]).

Οι σημαντικές αλλαγές που συντελέστηκαν τις τελευταίες δεκαετίες στο πεδίο της Μοριακής Βιολογίας (κλασικής και δομικής), σε συνδυασμό με την εξέλιξη της τεχνολογίας της γονιδιωματικής, αλλά και της μελέτη της βιοποικιλότητας και της διατήρησης της φύσης, οδήγησαν σε εκθετική αύξηση των πληροφοριών που παράγονται από τη βιολογική κοινότητα. Το γεγονός αυτό, κατέστησε απόλυτα αναγκαία τη διαχείριση, τον έλεγχο και την ανάλυση όλων αυτών των δεδομένων με τελικό σκοπό την αξιοποίησή τους για την εξαγωγή σημαντικών βιολογικών συμπερασμάτων. Άμεση συνέπεια των ανωτέρω ήταν η ανάγκη για την ανάπτυξη εξειδικευμένων υπολογιστικών εργαλείων (λογισμικού), αλλά και στην προσαρμογή, ήδη, υπάρχοντων δοκιμασμένων συστημάτων, για την αποθήκευση, οπτικοποίηση και ανάλυση δεδομένων, δίνοντας το έναυσμα για τη μεγάλη ανάπτυξη που παρατηρείται στις μέρες μας, στο επιστημονικό πεδίο της βιοπληροφορικής ([53]).

Συνεπώς, το πεδίο της βιοπληροφορικής έχει εξελιχθεί με τέτοιο τρόπο, ώστε ο κυριότερος στόχος να είναι η μελέτη και η σύνθεση των δεδομένων διαφόρου τύπου. Οι διαδικασίες αυτές αποτελούν τον κλάδο της υπολογιστικής βιολογίας (computational biology), ο οποίος αποτελεί με τη σειρά του έναν από τους σημαντικότερους κλάδους της βιοπληροφορικής ([53]).

Ένα παράδειγμα αποτελεί και η κατάρτιση ενός κατάλληλου πίνακα, ο οποίος συνδέει τις αποκρίσεις με τα ερεθίσματα μέσα στο κύτταρο. Ο πίνακας αυτός παρουσιάζει συνοπτικά τα αποτελέσματα πολλών πειραμάτων και τελικό σκοπό έχει τη διασύνδεση των αποτελεσμάτων, πράγμα που γίνεται με την εφαρμογή μαθηματικών μοντέλων, γενικών αλγορίθμων, αλγορίθμων βελτιστοποίησης, κ.α. ([7]).

Ένας τέτοιος πίνακας αποτελείται, συνήθως, από πολλές σειρές, όπου η κάθε σειρά αντιστοιχεί σε μία πειραματική διαδικασία. Αποτελείται, ακόμη, και από ένα σύνολο στηλών στις οποίες εμφανίζονται τα ερεθίσματα (cue). Στην απλούστερη περίπτωση, αν ένα ερέθισμα συμμετέχει στο πείραμα, τότε στην αντίστοιχη στήλη και γραμμή σημειώνουμε 1, αλλιώς σημειώνουμε 0. Τα ερεθίσματα μπορεί να περιλαμβάνουν την ενεργοποίηση συγκεκριμένης πρωτεΐνης (υποδοχέας) στην κυτταρική μεμβράνη ή την παρουσία φαρμάκων ([7]).

Για τη δημιουργία του συζυγούς μητρώου αναπαράστασης και τη χρησιμότητά του στη μοντελοποίηση των δικτύων μετάδοσης σημάτων, θα αναφερθούν περισσότερα σε παρακάτω κεφάλαια.

Από την άλλη πλευρά, οι αποκρίσεις χωρίζονται σε δύο διαφορετικές κατηγορίες. Από τη μία, έχουμε την ενεργοποίηση παραγόντων μέσα στο ίδιο το κύτταρο, όπου στην περίπτωση αυτή μετράμε ενδοκυτταρικά σήματα (signals). Από την άλλη, έχουμε την μέτρηση της απόκρισης του κυττάρου συνολικά (response), όπου στην περίπτωση αυτή μετράμε πρωτεΐνες που εκκρίνονται έξω από το κύτταρο, αλλά και τον συνολικό φαινότυπο του κυττάρου. Συνήθως, η μέτρηση της απόκρισης διαδραματίζεται συναρτήσεως του χρόνου μίας και είναι μία διαδικασία που απαιτεί χρόνο για να πραγματοποιηθεί ([7]).

Παρακάτω παρουσιάζεται η αναπαράσταση ενός τέτοιου πίνακα.

experiment	Ερέθισμα (CUE)						Ενδοκυτταρικά σήματα (SIGNALS)			Απόκριση (RESPONSE)					
	Υποδοχέας (Receptor)			Φάρμακα (Drugs)						Εκκρινόμενες πρωτεΐνες (Proteins)			Φαινότυπος (Phenotype)		
	R1	R2	...	D1	D2	...	S1	S2	...	Pr1	Pr2	...	Ph1	Ph2	...
1	0	0	...	0	0	...	0	0	...	0	0	...	0	1	...
2	0	1	...	1	0	...	1	1	...	1	0	...	1	1	...
3	1	0	...	0	0	...	0	0	...	0	1	...	1	0	...
...															

Εικόνα 10: Μητρώο αναπαράστασης απόκρισης κυτάρου ([7])

### 2.3.2 Στόχος

Στην προσπάθεια μελέτης της μετάλλαξης των φυσιολογικών κυτταρικών δραστηριοτήτων σε διαφορετικές νοσηρές καταστάσεις απαιτείται ο συνδυασμός βιολογικών δεδομένων για τη δημιουργία μίας κατανοήσιμης εικόνας αυτών των δραστηριοτήτων. Ως εκ τούτου, το πεδίο της βιοπληροφορικής εξελίχθηκε σε τέτοιο βαθμό, ώστε τα πιο επείγοντα «καθήκοντα» να περιλαμβάνουν την ανάλυση και την ερμηνεία διαφόρων τύπου δεδομένων, συμπεριλαμβανομένων νουκλεοτιδίων και ακολουθιών αμινοξέων, πρωτεϊνικών ακολουθιών και πρωτεϊνικών δομών. Όπως αναφέρθηκε και σε προηγούμενη παράγραφο, η ακριβής διαδικασία της ανάλυσης και ερμηνείας των δεδομένων αυτών αναφέρεται ως υπολογιστική βιολογία (computational biology). Ιδιαίτερα σημαντικοί επιμέρους κλάδοι του τομέα της βιοπληροφορικής και της υπολογιστικής βιολογίας αποτελούν ([53]):

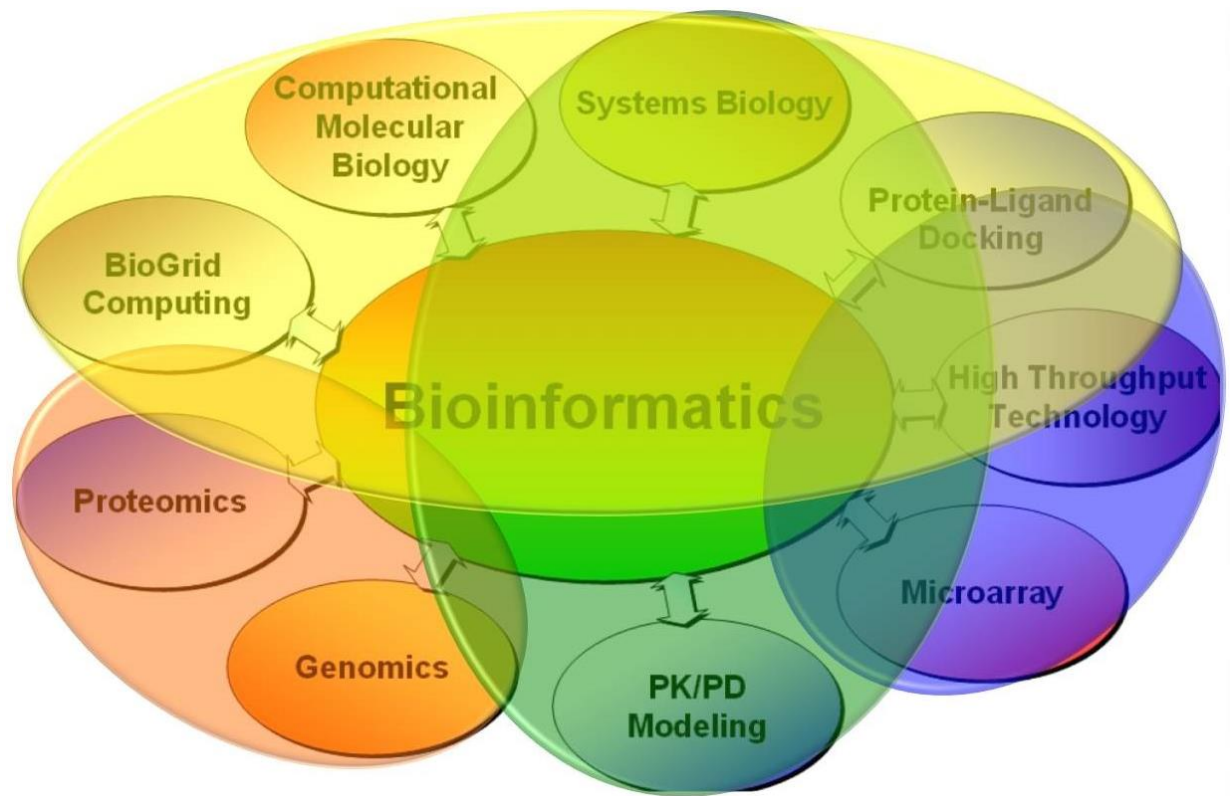
- η ανάπτυξη και η εκτέλεση κατάλληλων εργαλείων, τα οποία επιτρέπουν αποτελεσματική πρόσβαση, χρήση και διαχείριση διαφορετικών τύπου πληροφορίας
- η ανάπτυξη νέων αλγορίθμων (μαθηματικά, κυρίως, μοντέλα) και στατιστικών μοντέλων αναπαράστασης, με τα οποία θα διευκολύνεται η πρόσβαση στις σχέσεις μεταξύ μερών μεγάλων συνόλων δεδομένων. Για παράδειγμα, ανάπτυξη μεθόδων για τον εντοπισμό ενός γονίδιου μέσα σε μία ακολουθία, για την πρόβλεψη μίας πρωτεϊνικής δομής και

για την ταξινόμηση πρωτεϊνικών ακολουθιών σε οικογένειες συγγενών ακολουθιών.

Ο πρωταρχικός στόχος του τομέα της βιοπληροφορικής, είναι η αύξηση της κατανόησης των βιολογικών διαδικασιών. Αυτό που τον κάνει να ξεχωρίζει από άλλες προσεγγίσεις, όμως, είναι η εστίαση σε στην ανάπτυξη και εφαρμογή υπολογιστικά απαιτητικών μεθόδων για την επίτευξη του στόχου αυτού. Παραδείγματα: αναγνώριση προτύπων (pattern recognition), εύρεση δεδομένων (data mining), αλγόριθμοι γλώσσας μηχανής (machine learning algorithms) και οπτικοποίηση δεδομένων (data visualization). Βασικές ερευνητικές προσπάθειες στο πεδίο περιλαμβάνουν: εύρεση γονιδίων, «συναρμολόγηση» γονιδιώματος, εύρεση και ανάπτυξη φαρμάκων, πρόβλεψη πρωτεϊνικών δομών, πρόβλεψη γονιδιακής έκφρασης και πρωτεϊνικών αντιδράσεων και μοντελοποίηση της διαδικασίας της εξέλιξης ([42]).

Η βιοπληροφορική συνεπάγεται, σήμερα, τη δημιουργία και την προαγωγή βάσεων δεδομένων, αλγορίθμων, υπολογιστικών και στατιστικών μεθόδων και την ανάπτυξη θεωριών για την επίλυση τυπικών και πρακτικών εφαρμογών, οι οποίες αναδύονται από την διαχείριση και την ανάλυση βιολογικών δεδομένων ([42]).

Κατά τη διάρκεια των τελευταίων δεκαετιών η γρήγορη ανάπτυξη τεχνολογιών έρευνας γονιδιώματος και άλλων μορίων, αλλά και η ραγδαία ανάπτυξη των τεχνολογιών πληροφορικής, συνδυάστηκαν με σκοπό να παραχθεί ένα τεράστιο ποσό πληροφοριών σχετικές με τη μοριακή βιολογία. Συνοψίζοντας, λοιπόν, βιοπληροφορική είναι το όνομα που δόθηκε σε όλες αυτές τις μαθηματικές και υπολογιστικές προσεγγίσεις, οι οποίες χρησιμοποιούνται για την κατανόηση των βιολογικών διαδικασιών ([42]).



Εικόνα 11: Ο τομέας της βιοπληροφορικής

### 2.3.3 Ερευνητικές Προσεγγίσεις

Συνήθεις δραστηριότητες στον τομέα της βιοπληροφορικής περιλαμβάνουν τη σχεδίαση και την ανάλυση του γενετικού υλικού και των πρωτεϊνικών ακολουθιών, τη σύγκριση διαφορετικών γενετικών και πρωτεϊνικών ακολουθιών και τη δημιουργία και εποπτεία τρισδιάστατων μοντέλων πρωτεϊνικών δομών ([42]).

Υπάρχουν δύο βασικοί τρόποι μοντελοποίησης ενός βιολογικού συστήματος (ενός ζωντανού κυττάρου π.χ.) και οι δύο υπό την σκεπή των μεθόδων της βιοπληροφορικής ([42]):

- Στατικές Προσεγγίσεις
  - Ακολουθίες - πρωτεϊνών, νουκλεϊκών οξέων και πεπτιδίων
  - Αλληλεπίδραση δεδομένων μεταξύ των παραπάνω ποσοτήτων, συμπεριλαμβανομένων δεδομένων μικροσυστοιχιών και δικτύων πρωτεϊνών
- Δυναμικές Προσεγγίσεις
  - Δομές – πρωτεϊνών, νουκλεϊκών οξέων, συνδετικών ουσιών (συμπεριλαμβανομένων φαρμάκων και μεταβολιτών) και πεπτιδίων (μελετούμενες δομές με εργαλεία της βιοπληροφορικής δεν θεωρούνται

πια στατικές και η δυναμική τους αποτελεί συχνά τον πυρήνα των στατικών μελετών).

- Η συστημική βιολογία υπεισέρχεται σε αυτήν την κατηγορία, περιλαμβάνοντας αντιδράσεις και μεταβλητές συγκεντρώσεις μεταβολιτών.
- Πολύ – παραγοντικές μέθοδοι μοντελοποίησης, οι οποίες συλλαμβάνουν κυτταρικά συμβάντα, όπως η σηματοδότηση, η μεταγραφή και άλλες δυναμικές αντιδράσεις.

Μία ευρέως διαδεδομένη υποκατηγορία της βιοπληροφορικής είναι η δομική βιοπληροφορική.

Με τον όρο δομική βιοπληροφορική (structural bioinformatics), αναφερόμαστε σε μία πρόσφατα αναπτυσσόμενη προσέγγιση της βιοπληροφορικής, η οποία συνδέεται με την ανάλυση και την πρόβλεψη της τρισδιάστατης δομής των βιολογικών μακρομορίων, όπως είναι οι πρωτεϊνικές ακολουθίες, το RNA και το DNA. Ο όρος *δομική* έχει την ίδια σημασία όπως και στη δομική βιολογία και έτσι η δομική βιοπληροφορική μπορεί να θεωρηθεί σαν ένα μέρος της υπολογιστικής δομικής βιολογίας ([42]).

### **2.3.4 Βασικά Ερευνητικά Πεδία**

Παρακάτω αναφέρονται επιγραμματικά τα κύρια πεδία δραστηριοποίησης της επιστήμης της βιοπληροφορικής διεθνώς ([42]):

- Ανάλυση ακολουθιών (Sequence analysis)
- Σχολιασμός γονιδιωμάτων (Genome annotation)
- Εξελικτική υπολογιστική βιολογία (Computational evolutionary biology)
- Ανάλυση δημοσιεύσεων και έντυπων πόρων (Literature analysis)
- Ανάλυση γονιδιακής έκφρασης (Analysis of gene expression)
- Ανάλυση ρύθμισης (Analysis of regulation)
- Ανάλυση πρωτεϊνικής έκφρασης (Analysis of protein expression)
- Ανάλυση μεταλλάξεων σε καρκινικά κύτταρα (Analysis of mutations in cancer)
- Συγκριτική γονιδιωματική (Comparative genomics)
- Ανάλυση δικτύων και συστημική βιολογία (Network analysis and systems biology)
- Ανάλυση εικόνας υψηλής απόδοσης (High – throughput image analysis)



## 2.4 Μονοπάτια Μεταγωγής Σήματος

Στην ενότητα αυτή η θα αναλύσουμε αρχικά το τι είναι οι πρωτεΐνες, πως παράγονται στον ανθρώπινο οργανισμό και από τι δομούνται. Έπειτα, θα περιγραφεί ο τρόπος επικοινωνίας των κυττάρων μεταξύ των μέσω της επαφής των πρωτεϊνών και τέλος, θα ορισθούν τα μονοπάτια μεταγωγής σημάτων.

### 2.4.1 Εισαγωγικά

Το κύριο θέμα αυτής της διπλωματικής είναι η μοντελοποίηση μονοπατιών για τη μεταγωγή σημάτων μεταξύ κυττάρων, όπως έχει ήδη αναφερθεί πολλαπλές φορές και παραπάνω. Θα ήταν, όμως, βασικό να ξεκαθαρίσουμε σε αυτό το σημείο τι εννοούμε λέγοντας μεταγωγή σημάτων μεταξύ κυττάρων.

Ο όρος *κυτταρική μεταγωγή σήματος (Cellular Signal Transduction)* αναφέρεται ουσιαστικά στη διαδικασία «μετάφρασης» μίας περιβαλλοντικής διαταραχής σε μία κυτταρική απόκριση ([43]).

Όπως έχει αναφερθεί και παραπάνω τα κύτταρα ενυπάρχουν σε ένα πολύ σύνθετο βιοχημικό μικροπεριβάλλον. Ο τρόπος με τον οποίο το αντιλαμβάνονται οφείλεται σε ειδικευμένα πρωτεϊνικά μόρια, στην επιφάνειά τους, που ονομάζονται *υποδοχείς (receptors)* ([44]).

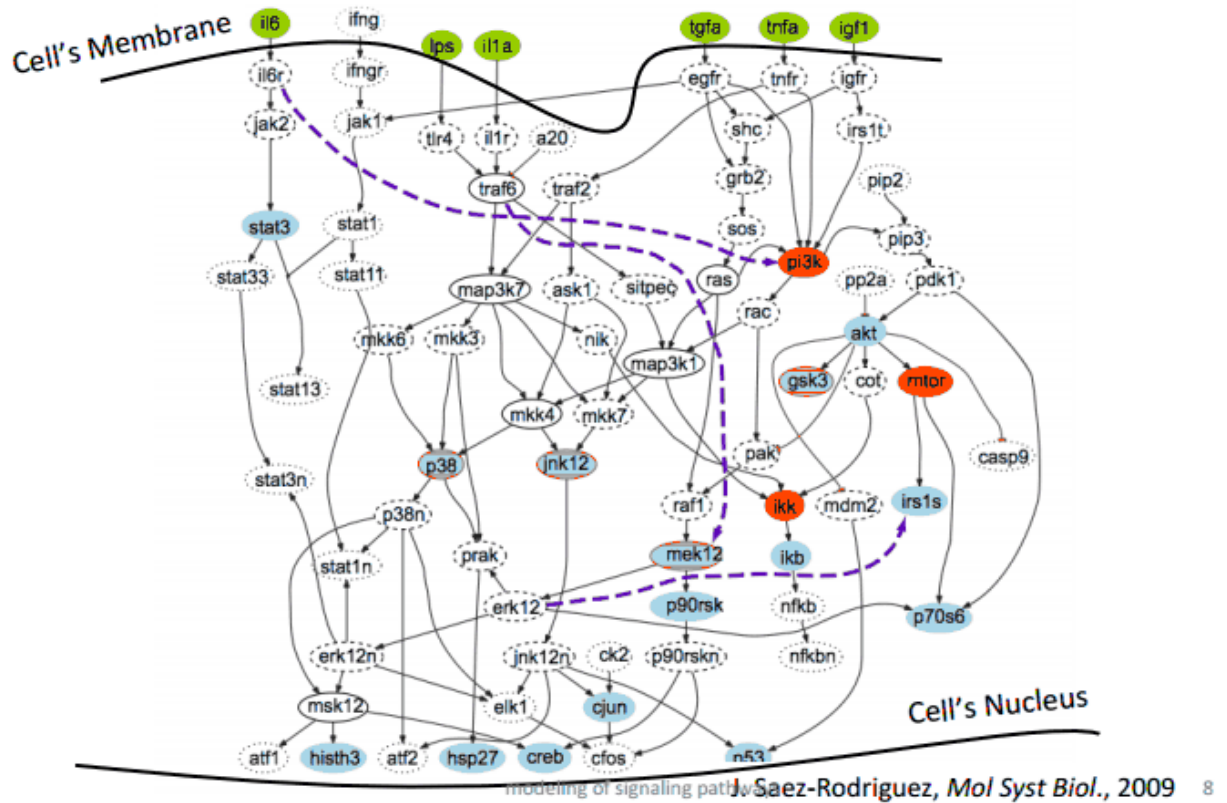
Ερεθίσματα από το περιβάλλον του κυττάρου προσκολλώνται στους υποδοχείς του, οι οποίοι κατά συνέπεια ενεργοποιούνται αρχίζοντας, έτσι, τη *διαδικασία σηματοδότησης*. Όταν ένας υποδοχέας ενεργοποιείται, ενεργοποιεί τις κατάντη πρωτεΐνες, οι οποίες με τη σειρά τους ενεργοποιούν τις κατάντη πρωτεΐνες, και η διαδικασία συνεχίζεται μέχρι η πληροφορία – σήμα να φτάσει μία οικογένεια πρωτεϊνών, η οποία ονομάζεται *παράγοντες μεταγραφής (transcription factors – TF)* ([44]).

Οι παραπάνω παράγοντες αφού ενεργοποιηθούν μπαίνουν μέσα στον πυρήνα του κυττάρου, προσκολλώνται στις προωθημένες περιοχές του γενετικού υλικού και εκφράζουν συγκεκριμένες κωδικοποιημένες περιοχές – γονίδια. Τα γονίδια μετά μεταφράζονται σε πρωτεΐνες και με αυτόν τον τρόπο το κύτταρο ενορχηστρώνει τη λειτουργία του και ρυθμίζονται συγκεκριμένες συμπεριφορές του, όπως ο πολλαπλασιασμός των κυττάρων, η μετανάστευση των κυττάρων, η κυτταρική απόπτωση, κτλ. ([44]).

Η μεταγωγή του σήματος είναι η διαδικασία της διάδοσής του από τη μία πρωτεΐνη στην επόμενη, ξεκινώντας από το επίπεδο των υποδοχέων φτάνοντας μέσα στον πυρήνα. Γενικά, η μελέτη της μεταγωγής σήματος μπορεί να δια φωτίσει τον τρόπο με τον οποίο τα κύτταρα αποκρίνονται στους παράγοντες του βιοχημικού τους μικροπεριβάλλοντος ([43]).

Τα γεγονότα της σηματοδότησης των κυττάρων αναπαρίστανται τυπικά σε δίκτυα, τα οποία αποκαλούνται δίκτυα σηματοδότησης, *μονοπάτια σηματοδότησης* ή *γράφοι αλληλεπίδρασης* ([12]).

# Signal transduction



Εικόνα 12: Δίκτυα μεταγωγής σημάτων ([12])

Στην παραπάνω εικόνα παρουσιάζεται ένα δίκτυο, στο οποίο απεικονίζονται τα συμβάντα σηματοδότησης κατάντη επτά υποδοχέων (***IL6R, IFNGR, TLR4, IL1R, EGFR, TNFR and IGF1R***). Οι υποδοχείς ενεργοποιούνται με τα προσκολλημένα σε αυτούς ερεθίσματα (πράσινοι κόμβοι). Οι άσπροι, οι μπλε και οι κόκκινοι κόμβοι είναι φωσφοπρωτεΐνες. Οι παράγοντες μεταγραφής (TFs) βρίσκονται στον πάτο του δικτύου ([12]).

Συνοψίζοντας, με τον όρο μονοπάτια μεταγωγής σημάτων μεταξύ κυττάρων (cell signaling transduction pathways), ουσιαστικά αναφερόμαστε σε μία ομάδα πρωτεϊνών, οι οποίες μεταφέρουν μεταγωγικά σήματα (το οποίο σημαίνει πως μετατρέπουν σήματα από το εξωτερικό περιβάλλον του κυττάρου σε διαφορετικά σήματα στο εσωτερικό του κυττάρου). Ως ένα τέτοιο εύστοχο παράδειγμα, μπορεί να αναφερθεί όταν μία ορμόνη δεσμεύεται στον υποδοχέα της κυτταρικής μεμβράνης (έξω από το κύτταρο). Ο υποδοχέας, ο οποίος μόλις ενεργοποιήθηκε μπορεί, τώρα, να αλληλεπιδράσει με ενδοκυτταρικές πρωτεϊνικές ουσίες, οι οποίες παράγουν νέα σήματα μέσα στο κύτταρο ([43]).

## 2.4.2 Πρωτεΐνες

Ο λόγος για τον οποίο μελετάμε τις πρωτεΐνες, είναι ότι αυτές αποτελούν τις δομικές μονάδες του κυττάρου και κατ' επέκταση ολόκληρου του οργανισμού. Χάρη στις πρωτεΐνες είναι δυνατή η εκτέλεση της πλειοψηφίας των λειτουργιών ενός κυττάρου, αλλά και η επικοινωνία των κυττάρων με άλλα κύτταρα και του οργανισμού ολόκληρου με το περιβάλλον του ([15]).

Όλες οι πρωτεΐνες, παρ' ότι ανήκουν στην ίδια μεγάλη οικογένεια, έχουν την δυνατότητα να επιτελούν πληθώρα εντελώς διαφορετικών λειτουργιών. Η ικανότητα τους αυτή σχετίζεται άμεσα με τον τεράστιο αριθμό τρισδιάστατων δομών που παρατηρούνται. Έτσι προκύπτει άμεσα ότι η λειτουργία μιας πρωτεΐνης εξαρτάται άμεσα από την τρισδιάστατη δομή της ([8]).

Στην παρακάτω εικόνα βλέπουμε μερικά μόνο παραδείγματα λειτουργιών που μπορεί να επιτελούν οι διάφορες πρωτεΐνες στον οργανισμό μας.



Εικόνα 13: Παραδείγματα λειτουργίας πρωτεϊνών ([8])

## Δομή πρωτεϊνών

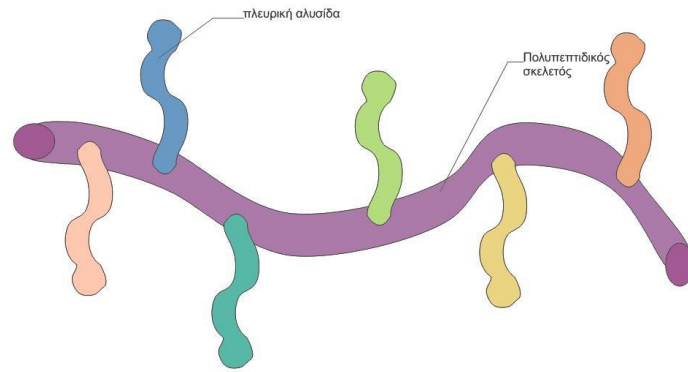
Γνωρίζουμε ότι οι πρωτεΐνες «κατασκευάζονται» από μικρότερα δομικά μόρια, τα οποία ονομάζονται αμινοξέα. Στον ανθρώπινο οργανισμό υπάρχουν είκοσι διαφορετικά αμινοξέα, τα οποία φαίνονται στον παρακάτω πίνακα. Από την αλληλουχία των αμινοξέων καθορίζεται η τρισδιάστατη δομή της κάθε πρωτεΐνης και η δομή της πρωτεΐνης με τη σειρά της καθορίζει τη λειτουργία της ([8]).

Αμινοξύ	Σύμβολο
Ασπαρταμικό οξύ	Asp
Γλουταμικό οξύ	Glu
Αργινίνη	Arg
Λυσίνη	Lys
Ιστιδίνη	His
Ασπαραγίνη	Asn
Γλουταμίνη	Gln
Σερίνη	Ser
Θρεονίνη	Thr
Τυροσίνη	Tyr
Αλανίνη	Ala
Γλυκίνη	Gly
Βαλίνη	Val
Λευκίνη	Leu
Ισολευκίνη	Ile
Προλίνη	Pro
Φαινυλαλανίνη	Phe
Μεθειονίνη	Met
Τρυποφάνη	Trp
Κυστεΐνη	Cys

Εικόνα 14: Αμινοξέα στον ανθρώπινο οργανισμό ([8])

Κάθε ένα από τα είκοσι παραπάνω αμινοξέα έχει τις δικές του διαφορετικές χημικές ιδιότητες. Για το σχηματισμό μιας πρωτεΐνης τα αμινοξέα συνδέονται μεταξύ τους με ομοιοπολικούς δεσμούς, σχηματίζοντας μακριές αλυσίδες. Έτσι, η κάθε πρωτεΐνη προκύπτει από μια μοναδική αλληλουχία αμινοξέων, η οποία είναι η ίδια για κάθε μόριο αυτής της πρωτεΐνης ([8]).

Καθώς τα διάφορα αμινοξέα συνδέονται αναμεταξύ τους, προκύπτει μια συνεχής αλληλουχία μορίων, που ονομάζεται πολυπεπτιδικός σκελετός (polypeptide backbone), και κάποιες πλευρικές ομάδες (side chains) που συνδέονται σε αυτόν. Οι πλευρικές ομάδες δεν συμμετέχουν στην κατασκευή του πολυπεπτιδικού σκελετού, αλλά είναι αυτές που δίνουν σε κάθε αμινοξύ μοναδικές ιδιότητες. Κάποιες πλευρικές ομάδες είναι μη πολικές και υδρόφοβες, κάποιες είναι πολικές με αρνητικό φορτίο και κάποιες με θετικό φορτίο. Στην παρακάτω εικόνα βλέπουμε μια σχηματική αναπαράσταση μιας πρωτεΐνης με τον πολυπεπτιδικό σκελετό και τις πλευρικές ομάδες ([8]).



Εικόνα 15: Τμήμα πρωτεΐνης ([8])

Ένα βασικό χαρακτηριστικό των πρωτεϊνών είναι ότι ο πολυπεπτιδικός σκελετός είναι ιδιαίτερα εύκαμπτος και μπορεί να αναδιπλώνεται. Έτσι λοιπόν άτομα του πολυπεπτιδικού σκελετού ή και άτομα των πλευρικών ομάδων μπορούν να συνδεθούν μεταξύ τους με ασθενείς, μη ομοιοπολικούς δεσμούς, όπως δεσμούς υδρογόνου, ιοντικούς δεσμούς, δυνάμεις Van der Waals ή εξαιτίας της διάταξης των υδρόφοβων και υδρόφιλων μονάδων. Αυτή η ιδιότητα των πρωτεϊνών, που καθορίζει και πολλές από τις λειτουργίες τους, ονομάζεται πτύχωση (folding) ([8]).

Η τελική τρισδιάστατη δομή που θα έχει μια πρωτεΐνη εξαρτάται λοιπόν αποκλειστικά από την αλληλουχία των αμινοξέων της και ονομάζεται πτυχωμένη δομή ή διαμόρφωση (conformation). Υπό την επίδραση κάποιων διαλυτών οι μη ομοιοπολικοί δεσμοί μπορεί να εξασθενίσουν και να σπάσουν και τότε η πρωτεΐνη ξεδιπλώνεται στην αρχική πολυπεπτιδική αλυσίδα. Η διαδικασία αυτή ονομάζεται μετουσίωση (denaturation) της πρωτεΐνης ([8]).

Η σημασία της τρισδιάστατης δομής μιας πρωτεΐνης είναι αναμφισβήτητη, αφού από αυτή εξαρτάται η σωστή λειτουργία της στον οργανισμό. Από την άλλη πλευρά αν κάποιες πρωτεΐνες για κάποιο λόγο δεν διπλωθούν σωστά, τότε μπορεί να σχηματίσουν αθροίσματα βλαπτικά για το κύτταρο ή και για ολόκληρους ιστούς. Ένα τέτοιο παράδειγμα αποτελούν κάποιες νευροεκφυλιστικές διαταραχές, όπως το Alzheimer και η νόσος του Huntington. Άλλες ασθένειες που οφείλονται σε συσσώρευση λάθος «διπλωμένων» πρωτεϊνών είναι η σπογγώδης εγκεφαλοπάθεια των βοοειδών (νόσος των τρελών αγελάδων) και η νόσος Creutzfeld-Jacob στους ανθρώπους. Οι ασθένειες αυτές οφείλονται στην άθροιση της πρωτεΐνης PrP με λάθος αναδίπλωση στον εγκέφαλο. Μάλιστα οι πρωτεΐνες PrP με λάθος αναδίπλωση μπορεί να αναγκάσουν φυσιολογικές πρωτεΐνες PrP να υιοθετήσουν την λάθος αναδίπλωση και έτσι η ασθένεια εξαπλώνεται γρήγορα στα εγκεφαλικά κύτταρα και εν τέλει προκαλεί το θάνατο ([8]).

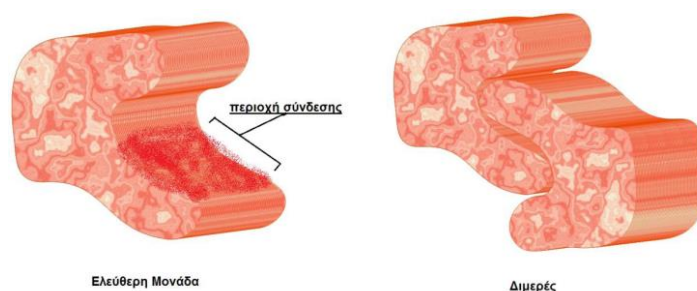
Για να καταλάβουμε την πολυπλοκότητα και την ποικιλομορφία των πρωτεϊνών, αρκεί να αναφέρουμε ότι μια πρωτεΐνη μπορεί να έχει από 30 μέχρι και πάνω από 10.000 αμινοξέα. Στην πλειοψηφία τους βέβαια οι πρωτεΐνες αποτελούνται από 50 έως 2.000 αμινοξέα, αλλά και πάλι οι δυνατότητες για διαφορετικούς συνδυασμούς είναι τεράστιες ([8]).

Πολύ βιολόγοι ασχολούνται σήμερα με τον εντοπισμό της αλληλουχίας των αμινοξέων διαφόρων πρωτεϊνών (Proteomics), χρησιμοποιώντας διάφορες άμεσες και έμμεσες μεθόδους που θα δούμε παρακάτω.

Πολλές πρωτεΐνες παρουσιάζουν παρεμφερείς ιδιότητες και σχετικά κοντινή δομή. Οι πρωτεΐνες αυτές λέμε ότι μπορούν να καταταγούν σε συγκεκριμένες πρωτεϊνικές οικογένειες (protein families). Τα μέλη μιας πρωτεϊνικής οικογένειας μπορεί να έχουν παρεμφερή αλληλουχία αμινοξέων και τρισδιάστατη διαμόρφωση, αλλά σε κάθε περίπτωση η κάθε πρωτεΐνη διεκπεραιώνει μια διαφορετική λειτουργία στον οργανισμό ([8]).

Όπως είδαμε, οι πρωτεΐνες αποκτούν τρισδιάστατη δομή με τους μη ομοιοπολικούς δεσμούς που σχηματίζονται ανάμεσα σε διάφορα σημεία του πολυπεπτιδικού σκελετού ή/και των πλευρικών ομάδων. Παρομοίως, διάφορες πρωτεΐνες μπορούν να συνδεθούν μεταξύ τους σχηματίζοντας μεγαλύτερες δομές. Επίσης, οι πρωτεΐνες μπορεί να συνδεθούν και με διάφορα μόρια του οργανισμού, μικρά ή μεγάλα. Οι περιοχές των πρωτεϊνών που μπορούν να συνδεθούν με άλλα μόρια ή πρωτεΐνες μέσω πολλών μη ομοιοπολικών δεσμών ονομάζονται περιοχές σύνδεσης ή πρόσδεσης (binding sites) ([8]).

Γενικά αποτελεί ιδιαίτερο γνώρισμα των πρωτεϊνών η δυνατότητά τους να σχηματίζουν δομές. Αρχικά, στην πιο απλή περίπτωση, δύο ίδιες πρωτεΐνες μπορεί να συνδέονται μεταξύ τους και να σχηματίζουν μια συμμετρική τελική δομή. Η τελική δομή που προκύπτει στην περίπτωση αυτή ονομάζεται διμερές (dimer) και η σταθερότητά του είναι αποτέλεσμα της αλληλεπίδρασης των δύο ταυτιζόμενων περιοχών σύνδεσης. Επίσης, νέες πρωτεΐνες μπορεί να σχηματίζονται και από δύο ή περισσότερα διαφορετικά είδη πολυπεπτιδικών αλυσίδων ([8]).



**Εικόνα 16: Σχηματισμός διμερούς πρωτεΐνης ([8])**

Μια άλλη σημαντική ιδιότητα των πρωτεϊνών είναι η ύπαρξη συμπληρωματικών θέσεων σύνδεσης πάνω στην ίδια πρωτεΐνη. Έτσι μια περιοχή σύνδεσης μπορεί να ενωθεί με μια άλλη περιοχή σύνδεσης πάνω στην ίδια πρωτεΐνη και με τον τρόπο αυτό είναι δυνατό να σχηματισθούν μεγάλα συμπλέγματα πρωτεϊνών ([8]).

Όταν τα μόρια διατάσσονται σε μορφή έλικας, τότε σχηματίζεται ένα εκτεταμένο νήμα που ονομάζεται ινίδιο. Παράδειγμα πρωτεΐνης που

σχηματίζει ινίδια αποτελεί η ακτίνη, η οποία συμμετέχει στην διαδικασία κίνησης των μυών. Περισσότερες λεπτομέρειες για την συμβολή των πρωτεϊνών στους μηχανισμούς κίνησης θα δούμε στην επόμενη παράγραφο, η οποία περιγράφει τη λειτουργία των πρωτεϊνών. Επίσης, οι πρωτεΐνες σχηματίζουν δακτυλίους, φύλλα, σφαίρες ή σωλήνες. Η πληροφορία για τον τρόπο σύνδεσης στις πρωτεΐνες δεν χρειάζεται να δοθεί εξωτερικά, αφού οφείλεται στην αλληλουχία των αμινοξέων ([8]).

Η τρισδιάστατη δομή των διαφόρων πρωτεϊνών είναι συνήθως ιδιαίτερα πολύπλοκη. Παρ' όλα αυτά αν δούμε τις διάφορες πρωτεΐνες από απόσταση μπορούμε να ξεχωρίσουμε χονδρικά δύο μεγάλες υποκατηγορίες. Υπάρχουν οι πρωτεΐνες εκείνες στις οποίες η πολυπεπτιδική αλυσίδα διπλώνεται αλληπάλληλα σε μια συμπαγή δομή. Οι πρωτεΐνες αυτές ονομάζονται σφαιρικές (globular proteins), και στις σφαιρικές πρωτεΐνες ανήκουν κατά το πλείστον τα ένζυμα. Στη δεύτερη κατηγορία ανήκουν εκείνες οι πρωτεΐνες που έχουν μεγάλο μήκος σε σχέση με τις άλλες διαστάσεις τους. Οι πρωτεΐνες αυτές ονομάζονται ινώδεις (fibrous proteins), και στις ινώδεις πρωτεΐνες ανήκουν σε μεγάλο ποσοστό οι εξωκυττάρειες πρωτεΐνες όπως το κολλαγόνο ([8]).

Τέλος σε ότι αναφορά τις εξωκυττάρειες πρωτεΐνες, παρατηρείται η ιδιαιτερότητα της σταθεροποίησης της τρισδιάστατης δομής τους με ομοιοπολικούς δεσμούς. Οι πρωτεΐνες αυτές βρίσκονται εκτεθειμένες στις συνθήκες του εξωτερικού περιβάλλοντος, έξω από την προστατευτική κυτταρική μεμβράνη. Έτσι, στην περίπτωση αυτή απαιτείται μια πιο σταθερή διασύνδεση, όπως αυτή που προσφέρει ο ομοιοπολικός δεσμός ([8]).

Χαρακτηριστικό παράδειγμα αποτελούν οι ομοιοπολικοί δεσμοί ανάμεσα σε άτομα θείου, ή αλλιώς δισουλφιδρυλικοί δεσμοί (disulfide bonds), ή αλλιώς δεσμός S-S. Γενικά οι δισουλφιδρυλικοί δεσμοί δεν μεταβάλουν τη διαμόρφωση μιας πρωτεΐνης αλλά συνδέουν διαφορετικά τμήματα μιας ή περισσότερων πρωτεϊνών. Οι δεσμοί αυτοί χρησιμοποιούνται για παράδειγμα στο σχηματισμό της τρίχας, όπου ξεχωριστές ίνες κερατίνης συνδέονται ομοιοπολικά μεταξύ τους ([8]).

### **Λειτουργίες πρωτεϊνών**

Οι λειτουργίες που επιτελούν οι διάφορες πρωτεΐνες εξαρτώνται σε μεγάλο βαθμό από την τρισδιάστατη δομή τους, αφού τα διαφορετικά αμινοξέα που συμμετέχουν στη σύνθεση μιας πρωτεΐνης της προσδίδουν και διαφορετικές χημικές ιδιότητες. Αυτός λοιπόν ο συνδυασμός εξωτερικής δομής και χημείας δίνει στις διάφορες πρωτεΐνες τη δυνατότητα να συμμετέχουν και να καθορίζουν δυναμικά φαινόμενα στα ζωντανά κύτταρα ([8]).

Για να επιτελέσουν τις λειτουργίες του οργανισμού όλες οι πρωτεΐνες απαιτείται να έχουν κάποια φυσική αλληλεπίδραση με άλλα μόρια. Όλες οι πρωτεΐνες λοιπόν συνδέονται με άλλα μόρια με μια ποικιλία διαφορετικών δεσμών που μπορεί να είναι από πολύ ισχυροί και μόνιμοι, μέχρι ασθενείς και ιδιαίτερα βραχύβιοι. Βέβαια, οι συνδέσεις αυτές δεν είναι ούτε άπειρες, ούτε

τυχαίες. Αντίθετα κάθε πρωτεΐνη παρουσιάζει μεγάλη εξειδίκευση (specificity) και μπορεί να συνδεθεί μόνο με ένα ή μερικά από τα χιλιάδες μόρια που υπάρχουν στον οργανισμό μας.

Όπως αναφέραμε ήδη παραπάνω, η ικανότητα μιας πρωτεΐνης να συνδέεται με μια άλλη πρωτεΐνη ή ένα άλλο μόριο οφείλεται στο σχηματισμό πολλών ασθενών, μη ομοιοπολικών δεσμών σε όλη την περιοχή σύνδεσης. Επίσης, για να είναι αρκετά ισχυρός ο δεσμός αυτός και να μπορεί να επιτελέσει την απαιτούμενη λειτουργία, θα πρέπει οι διάφορες περιοχές να ταιριάζουν αρκετά καλά γεωμετρικά, όπως τα κομμάτια σε ένα πάζλ. Έτσι οι αποστάσεις μεταξύ των διαφόρων περιοχών των αμινοξέων που έχουν τη δυνατότητα να σχηματίσουν δεσμούς μικραίνουν και από την πληθώρα των ασθενών δεσμών που σχηματίζονται, έχουμε τελικά μια αρκετά ισχυρή σύνδεση μεταξύ πρωτεΐνης και μορίου.

Η ουσία που συνδέεται με μια πρωτεΐνη ονομάζεται *συνδέτης ή προσδέτης (ligand)*. Ο συνδέτης μπορεί να είναι ένα ιόν, ένα μικρό μόριο, μια άλλη πρωτεΐνη ή και ένα ογκώδες μακρομόριο ([8]).

Από όλες τις πρωτεΐνες του οργανισμού μας, αυτές που παρουσιάζουν μεγαλύτερη αποτελεσματικότητα στη σύνδεσή τους με διάφορους συνδετές, είναι τα αντισώματα. Τα αντισώματα, ή αλλιώς ανοσοσφαιρίνες, είναι πρωτεΐνες που παράγονται από το ανοσοποιητικό σύστημα ενός οργανισμού ως απόκριση σε μια εξωτερική εισβολή. Τα αντισώματα λοιπόν παράγονται με σκοπό να συνδεθούν ισχυρά με το μόριο στόχο και είτε να τον αδρανοποιούν άμεσα, είτε του δίνουν την απαραίτητη σήμανση ώστε στη συνέχεια να τον καταστρέψουν τα πιο ισχυρά κύτταρα του ανοσοποιητικού συστήματος. Ο στόχος, τον οποίο αναγνωρίζει με μεγάλη εξειδίκευση ένα αντίσωμα, ονομάζεται κατ' αντιστοιχία αντιγόνο.

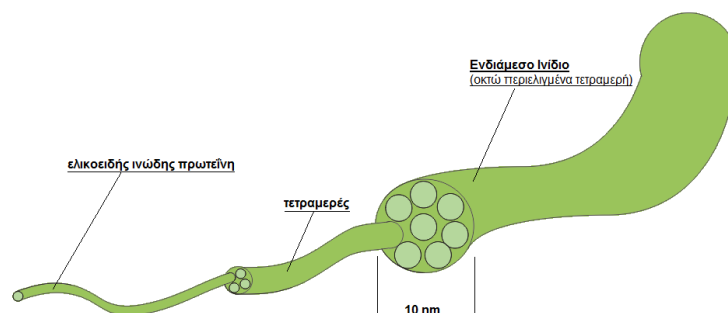
Εκτός όμως από τη δυνατότητα τους να συνδέονται με άλλες πρωτεΐνες ή παθογόνους μικροοργανισμούς, τα πρωτεϊνικά μόρια πολλές φορές, για να εκτελέσουν τις λειτουργίες τους, συνδέονται με άλλα μικρά μόρια, ιόντα ή ιχνοστοιχεία. Για παράδειγμα η αιμοσφαιρίνη του αίματος περιέχει ένα άτομο σιδήρου απαραίτητο για τη σύνδεση με το οξυγόνο. Άλλες πρωτεΐνες μπορεί να συνδέονται σε μόνιμες δομές, όπως οι πρωτεΐνες της κυτταρικής μεμβράνης που φροντίζουν για την επικοινωνία του κυττάρου με το εξωτερικό περιβάλλον.

Μια άλλη ιδιότητα των πρωτεϊνών που παρουσιάζει ξεχωριστό ενδιαφέρον είναι η ικανότητά τους να δομούν και να κινούν ολόκληρους ιστούς του οργανισμού. Το περίπλοκο δίκτυο των πρωτεϊνικών ινιδίων, στο οποίο οφείλει το κύτταρο την ικανότητα του να οργανώνει τα διάφορα συστατικά του, να λαμβάνει διάφορα σχήματα και να πραγματοποιεί κινήσεις, ονομάζεται κυτταροσκελετός (cytoskeleton). Ο κυτταροσκελετός είναι πολύ σημαντικός για την επιβίωση των κυττάρων ζωικών οργανισμών. Από την άλλη πλευρά, τα φυτικά κύτταρα που είναι πιο στατικά και δεν επιτελούν κινήσεις διαθέτουν το πιο άκαμπτο κυτταρικό τοίχωμα.



Ο κυτταροσκελετός κατασκευάζεται από ένα πλέγμα που αποτελείται από τρία είδη πρωτεϊνικών ινιδίων: τα ενδιάμεσα ινίδια (intermediate filaments), τους μικροσωληνίσκους (microtubules) και τα νημάτια ή ινίδια ακτίνης (actin filaments). Στην συνέχεια θα εξετάσουμε συνοπτικά καθένα από αυτά τα τρία είδη ([8]).

➤ Ενδιάμεσα Ινίδια (intermediate filaments):



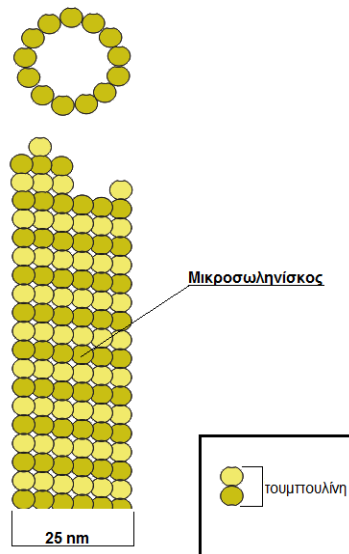
Εικόνα 17: Δομή ενδιάμεσου ινιδίου ([8])

Σχηματίζονται από μια οικογένεια ινώδων πρωτεϊνών. Η ονομασία «ενδιάμεσα» προκύπτει από το γεγονός ότι η διάμετρός τους είναι μεν μικρότερη από τη διάμετρο των μικροσωληνίσκων, αλλά είναι μεγαλύτερη από τη διάμετρο των ινιδίων ακτίνης. Αποτελούν τα σκληρότερα και ανθεκτικότερα από τα τρία είδη ινιδίων και βρίσκονται στο κυτταρόπλασμα των περισσότερων ζωικών κυττάρων. Επίσης αποτελούν σημαντικό ενισχυτικό και υποστηρικτικό υλικό του πυρηνικού περιβλήματος ([8]).

Στη δομή, τα ενδιάμεσα ινίδια μοιάζουν πολύ με σχοινιά. Οι μικρότεροι κλώνοι είναι μακριές ινώδεις πρωτεΐνες. Αυτές αποτελούν τα μονομερή των ινιδίων, ενώ κάθε μια έχει ελικοειδή δομή στο κυρίως σώμα της. Λόγω της ελικοειδούς δομής τους οι πρωτεΐνες αυτές μπορούν να περιελίσσονται η μια γύρω από την άλλη και να συνδέονται ανά δύο σε σταθερά διμερή. Στη συνέχεια, δύο διμερή συνδέονται μεταξύ τους με μη ομοιοπολικούς δεσμούς και σχηματίζουν ένα τετραμερές. Τέλος, οκτώ τετραμερή συνδέονται και πάλι μεταξύ τους με μη ομοιοπολικούς δεσμούς και σχηματίζουν το τελικό ενδιάμεσο ινίδιο. Η διάμετρος ενός ενδιάμεσου ινιδίου είναι περίπου 10nm ([8]).

Το βασικό χαρακτηριστικό των ενδιάμεσων ινιδίων είναι ότι προσδίδουν στα κύτταρα την απαραίτητη μηχανική ισχύ και για αυτό βρίσκονται σε αφθονία στο κυτταρόπλασμα κυττάρων που υπόκεινται σε μηχανική καταπόνηση. Τέτοια κύτταρα είναι τα νευρικά, τα μυϊκά και τα επιθηλιακά κύτταρα. Ειδικά στα κύτταρα της επιδερμίδας η προστασία είναι απαραίτητη, αφού συχνά υφίστανται ισχυρή μηχανική καταπόνηση. Τέλος, τα ενδιάμεσα ινίδια καλύπτουν και υποστηρίζουν την εσωτερική επιφάνεια της πυρηνικής μεμβράνης, οργανωμένα σε μια δομή που λέγεται πυρηνικός υμένας (nuclear lamina) ([8]).

➤ Μικροσωληνίσκοι (microtubules):



**Εικόνα 18: Δομή μικροσωληνίσκου ([8])**

Σχηματίζονται από υπομονάδες τουμπουλίνης (tubulin). Κάθε μονάδα τουμπουλίνης είναι ένα διμερές από δύο ισχυρά συνδεδεμένες μεταξύ του παρόμοιες σφαιρικές πρωτεΐνες. Οι μικροσωληνίσκοι έχουν μεγάλο μήκος και είναι σχετικά άκαμπτοι κοίλοι σωλήνες. Το βασικό χαρακτηριστικό τους είναι ότι συναρμολογούνται και αποσυναρμολογούνται πολύ γρήγορα. Στα κύτταρα σχηματίζουν ένα δίκτυο μέσω του οποίου μετακινούνται διάφορα μόρια και συστατικά του κυττάρου ([8]).

Τα δομικά συστατικά των μικροσωληνίσκων είναι οι υπομονάδες της τουμπουλίνης. Οι υπομονάδες αυτές στοιβάζονται η μια πάνω στην άλλη, πάντα με τον ίδιο προσανατολισμό, και σχηματίζουν τελικά ένα κοίλο κυλινδρικό μικροσωληνίσκο. Η σύνδεση μεταξύ των υπομονάδων γίνεται με μη ομοιοπολικούς δεσμούς. Οι μικροσωληνίσκοι που παράγονται με αυτή τη διαδικασία έχουν πάντα 13 υπομονάδες τουμπουλίνης στην περιφέρειά τους, ενώ κατά μήκος οι υπομονάδες έχουν πάντα τον ίδιο προσανατολισμό. Η εξωτερική διάμετρος ενός μικροσωληνίσκου είναι περίπου 25nm ([8]).

Ένα βασικό χαρακτηριστικό των μικροσωληνίσκων είναι ότι μέσα στο κύτταρο η δομή τους παρουσιάζει αστάθεια. Ανάλογα με την διαδικασία που επιτελείται κάθε φορά δημιουργούνται νέοι μικροσωληνίσκοι, κάποιοι άλλοι διασπώνται πλήρως, ενώ άλλοι μπορεί να αυξομειώνουν το μήκος τους ([8]).

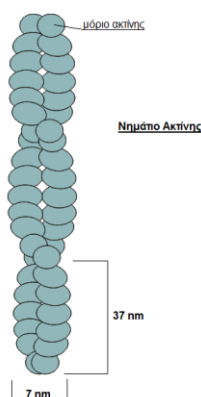
Με βάση λοιπόν το εκάστοτε δίκτυο των μικροσωληνίσκων πραγματοποιείται η οργάνωση του εσωτερικού του κυττάρου. Οι μικροσωληνίσκοι χρησιμοποιούνται ως τροχιές πάνω στις οποίες

ταξιδεύουν και διακινούνται τα διάφορα υλικά του κυττάρου, όπως κυστίδια, οργανίδια κ.α. με τη βοήθεια κάποιων άλλων επικουρικών πρωτεϊνών. Οι πρωτεΐνες που προκαλούν την μετακίνηση των ουσιών πάνω στους μικροσωληνίσκους ονομάζονται κινητήριες πρωτεΐνες (motor proteins) και χωρίζονται γενικά στις κινεσίνες (kinesins) και στις δυνεΐνες (dyneins). Το κριτήριο με βάση το οποίο γίνεται η διάκριση είναι η κατεύθυνση της κίνησης των πρωτεϊνών αυτών πάνω στους μικροσωληνίσκους, αφού όπως είδαμε, λόγω της δομής τους, οι μικροσωληνίσκοι έχουν συγκεκριμένο και διακριτό προσανατολισμό. Γενικά, οι κινητήριες πρωτεΐνες προσδένονται με το ένα άκρο στον μικροσωληνίσκο και με το άλλο άκρο τους παραλαμβάνουν μια ουσία προς μεταφορά. Στη συνέχεια κινούνται προς την καθορισμένη τους διεύθυνση καταναλώνοντας ενέργεια από το κύτταρο ([8]).

Μία άλλη πολύ βασική λειτουργία των μικροσωληνίσκων είναι ο διαχωρισμός των χρωμοσωμάτων κατά τη μίτωση ενός κυττάρου. Τότε οι μικροσωληνίσκοι που βρίσκονται στο κυτταρόπλασμα μεταβάλουν τη δομή τους και σχηματίζουν μια δομή που ονομάζεται μιτωτική άτρακτος (mitotic spindle). Η μιτωτική άτρακτος είναι απολύτως απαραίτητη ώστε να διαχωριστούν εξίσου τα χρωμοσώματα στα θυγατρικά κύτταρα ([8]).

Τέλος, οι μικροσωληνίσκοι μπορεί να σχηματίσουν και μόνιμες δομές οι οποίες να προεκβάλλουν από το κύτταρο. Οι δομές αυτές μπορεί να είναι κροσσοί (cilia) ή μαστίγια (flagella) και χρησιμοποιούνται για να σαρώνουν το χώρο γύρω από την κυτταρική επιφάνεια ή για να προωθούν το κύτταρο. Ο πυρήνας αυτών των δομών είναι μια διασυνδεδεμένη ομάδα μικροσωληνίσκων, η οποία αποκτά σχετική κίνηση χάρη σε έναν μηχανισμό παρόμοιο με αυτόν που περιγράφηκε για τη διακίνηση ουσιών στο κύτταρο μέσω του δικτύου των μικροσωληνίσκων. Χαρακτηριστικό παράδειγμα μαστίγιου σε ευκαρυωτικό κύτταρο αποτελεί η «ουρά» του σπερματοζωαρίου. Τα μαστίγια των ευκαρυωτικών κυττάρων δεν πρέπει να συγχέονται με τα μαστίγια των βακτηρίων που έχουν εντελώς διαφορετική δομή ([8]).

➤ Νημάτια ακτίνης (actin filaments):



Εικόνα 19: Δομή νηματίου ακτίνης ([8])

Τα νημάτια ή ινίδια της ακτίνης έχουν ως δομική μονάδα τη μικρή σφαιρική πρωτεΐνη ακτίνη. Τα νημάτια που σχηματίζονται από την ακτίνη είναι γενικά πιο λεπτά και ευλύγιστα από τους μικροσωληνίσκους και συνήθως έχουν μικρότερο μήκος. Οι ινώδεις αυτές δομές είναι άφθονες σχεδόν σε όλα τα είδη κυττάρων, αφού είναι απαραίτητες για πολλές κινήσεις του κυττάρου ως σύνολο, με τη συμμετοχή της κυτταρικής επιφάνειας. Επίσης παρουσιάζουν αρκετές δομικές και λειτουργικές ομοιότητες με τους μικροσωληνίσκους ([8]).

Τα νημάτια της ακτίνης ομοιάζουν στην όψη με αυτό που δηλώνει και το όνομά τους, δηλαδή με λεπτά νήματα. Τα μόρια της ακτίνης συνδέονται μεταξύ τους και σχηματίζουν ελικοειδείς δομές. Οι έλικες αυτές συνδέονται στην συνέχεια ανά δύο και σχηματίζουν μια περιελιγμένη αλυσίδα που αποτελεί το νημάτιο. Τα μόρια της ακτίνης είναι πολωμένα και όταν συνδέονται μεταξύ τους διατάσσονται όλα προς την ίδια κατεύθυνση, έτσι και τα νημάτια της ακτίνης, όπως και οι μικροσωληνίσκοι, είναι προσανατολισμένα. Η διάμετρος κάθε νημάτιου είναι περίπου 7nm και η στροφές στη δίκλωνη έλικά του απέχουν μεταξύ τους 37nm ([8]).

Όπως και οι μικροσωληνίσκοι, έτσι και τα νημάτια της ακτίνης έχουν τη δυνατότητα να αυξομειώνουν συνεχώς το μήκος τους, να αποδομούνται ή να αναδομούνται με την βοήθεια άλλων πρωτεϊνών. Άλλωστε, η ιδιότητα αυτή δίνει και στα νημάτια την ικανότητα να μεταφέρουν ουσίες του κυττάρου, όπως και οι μικροσωληνίσκοι ([8]).

Μια δεύτερη ομοιότητα των νηματίων της ακτίνης με τους μικροσωληνίσκους είναι η δυνατότητά τους να συνδέονται με μεγάλο αριθμό πρωτεϊνών, οι οποίες στην περίπτωση αυτή ονομάζονται γενικά πρωτεΐνες που προσδέονται ακτίνη (actin-binding proteins). Οι πρωτεΐνες αυτές επιτρέπουν στα νημάτια να εκτελούν μια πληθώρα λειτουργιών ([8]).

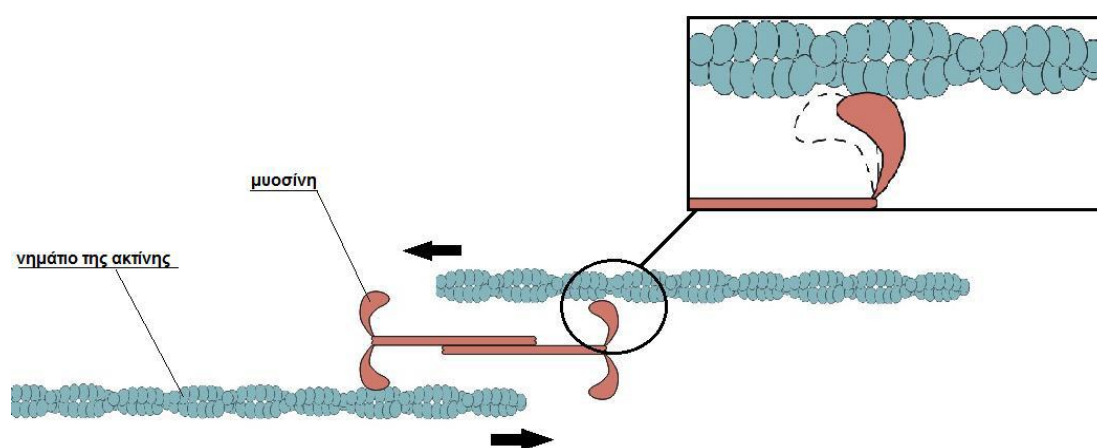
Ανάλογα με το είδος των πρωτεϊνών που προσδέονται πάνω τους τα νημάτια της ακτίνης μπορούν να σχηματίσουν διάφορες παροδικές ή μόνιμες δομές. Χαρακτηριστικό είναι το γεγονός ότι από τα ινίδια της ακτίνης σχηματίζονται οι μικρολάχνες (microvilli), που είναι δομές του εντέρου, οι δέσμες συστολής (contractile bundles) στο κυτταρόπλασμα, οι οποίες λειτουργούν σαν μύες του κυττάρου κ.α. ([8]).

Ακόμα, διάφορες άλλες πρωτεΐνες συγκρατούν τα νήματα της ακτίνης και τα συγκεντρώνουν σε ένα στρώμα ακριβώς κάτω από την κυτταρική μεμβράνη. Η περιοχή αυτή ονομάζεται κυτταρικός φλοιός (cell cortex) και έχει τη μεγαλύτερη συγκέντρωση νηματίων της ακτίνης από όλο το κυτταρόπλασμα. Επίσης, υπάρχουν άλλες πρωτεΐνες που κόβουν τα νημάτια της ακτίνης και απομονώνουν τα μονομερή τους. Όπως είδαμε και παραπάνω, τα νημάτια της ακτίνης, όπως και οι μικροσωληνίσκοι μεταφέρουν ουσίες των κυττάρων. Η λειτουργία αυτή ακολουθεί τον ίδιο μηχανισμό που περιγράφηκε παραπάνω με τη βοήθεια διάφορων συνδετικών πρωτεϊνών ([8]).

Τέλος, το χαρακτηριστικό των νηματίων ακτίνης στο οποίο θα σταθούμε λίγο περισσότερο εδώ, είναι η ικανότητά τους να σχηματίζουν δέσμες συστολής

στα μυϊκά κύτταρα για την κίνηση του οργανισμού. Αυτό επιτυγχάνεται με την σύνδεση των νηματιών ακτίνης με διάφορες κινητήριες πρωτεΐνες που ανήκουν στην οικογένεια της μυοσΐνης. Έτσι με έναν μηχανισμό που καταναλώνει ενέργεια από το κύτταρο επιτυγχάνεται η σχετική κίνηση των νηματιών της ακτίνης και σε μεγαλύτερη κλίμακα η συστολή και χαλάρωση των μυών ([8]).

Στην παρακάτω εικόνα φαίνεται σχηματικά μια αναπαράσταση της σύνδεσης και αλληλεπίδρασης των ινιδίων της ακτίνης με τη μυοσΐνη.



Εικόνα 20: Μηχανισμός μυϊκής συστολής ([8])

Όσα ειπώθηκαν παραπάνω προφανώς δεν καλύπτουν το φάσμα και την ποικιλομορφία των πρωτεϊνών και των λειτουργιών που αυτές επιτελούν σε έναν οργανισμό. Σκοπός μας είναι περισσότερο να κατανοήσει ο αναγνώστης τη σημασία των πρωτεϊνών αλλά και το πόσο πολύπλοκες και πολλές φορές ανεξερεύνητες είναι οι διαδικασίες των κυττάρων. Επίσης, σε μια προσπάθεια να επεξηγηθεί ο μηχανισμός κίνησης ενός ολόκληρου οργανισμού, αλλά και ρίχνοντας μια πρώτη ματιά στην επικοινωνία των κυττάρων, εξηγήσαμε παραπάνω τις βασικές νηματοειδείς πρωτεΐνες του ανθρώπινου οργανισμού ([8]).

## Παραγωγή πρωτεϊνών

Παραπάνω παρουσιάσαμε πως η αλληλουχία των αμινοξέων καθορίζει την τρισδιάστατη δομή και τις δυνατές λειτουργίες μιας πρωτεΐνης. Οι πρωτεΐνες με τη σειρά τους εκτελούν το σύνολο των λειτουργιών ενός οργανισμού. Επομένως το στοιχείο αυτό που «ελέγχει» τις πρωτεΐνες, «ελέγχει» και ολόκληρο τον οργανισμό. Είδαμε σε παραπάνω παράγραφο ότι στις αλληλουχίες των νουκλεοτιδίων του DNA είναι αποθηκευμένες όλες οι πληροφορίες που αφορούν έναν οργανισμό και τα γονίδια φέρουν τις απαραίτητες πληροφορίες για την παραγωγή των πρωτεϊνών. Στο σημείο

αυτό θα εξετάσουμε σε γενικές γραμμές πώς πραγματοποιείται αυτή η μετάβαση από το DNA στις πρωτεΐνες ([8]).

Στην περίπτωση που μια πρωτεΐνη είναι αναγκαία για το κύτταρο, η δίκλωνη αλυσίδα του DNA διαχωρίζεται στο αντίστοιχο σημείο όπου βρίσκεται το γονίδιο που περιέχει τις πληροφορίες για την παραγωγή αυτής της πρωτεΐνης. Στη συνέχεια το «ανοιχτό» τμήμα του DNA αντιγράφεται σε ένα άλλο είδος νουκλεϊκού οξέως, το ριβονουκλεϊκό οξύ (RiboNucleic Acid, RNA). Αυτά τα RNA αντίγραφα χρησιμοποιούνται με τη σειρά τους για την παραγωγή των αντίστοιχων πρωτεϊνών ([8]).

Αν και εκ πρώτης όψεως ίσως φαίνεται άσκοπη η παρεμβολή του RNA στην όλη διαδικασία, στην πραγματικότητα είναι ζωτικής σημασίας για τη διαφύλαξη της ακεραιότητας του εκάστοτε γονιδίου που είναι συνήθως μοναδικό σε κάθε κύτταρο. Επίσης, με παράλληλη παραγωγή πολλών μορίων RNA αυξάνεται η παραγωγικότητα των γονιδίων και μπορούν να παραχθούν τεράστιες ποσότητες πρωτεϊνών μέσα σε λίγα δευτερόλεπτα ([8]).

Η διαδικασία, όπου η αλληλουχία των νουκλεοτιδίων του DNA οδηγεί στην δημιουργία μια κατάλληλης αλληλουχίας RNA, ονομάζεται *μεταγραφή* (*transcription*). Στη δομή του RNA δεν θα σταθούμε ιδιαίτερα καθώς είναι παρόμοια με αυτή του DNA. Οι κύριες διαφορές μεταξύ DNA και RNA είναι ([8]):

- Το RNA είναι μονόκλωνο, ενώ το DNA είναι δίκλωνο.
- Το RNA περιέχει στο νουκλεοτίδιό του το σάκχαρο ριβόζη, αντί της δεοξυριβόζης που περιέχουν τα νουκλεοτίδια του DNA.
- Το RNA περιέχει τη βάση ουρακίλη (U), αντί της θυμίνης (T) του DNA. Κατ' αντιστοιχία με τη θυμίνη (T), η ουρακίλη (U) συνδέεται σε ζεύγος με την αδενίνη (A).

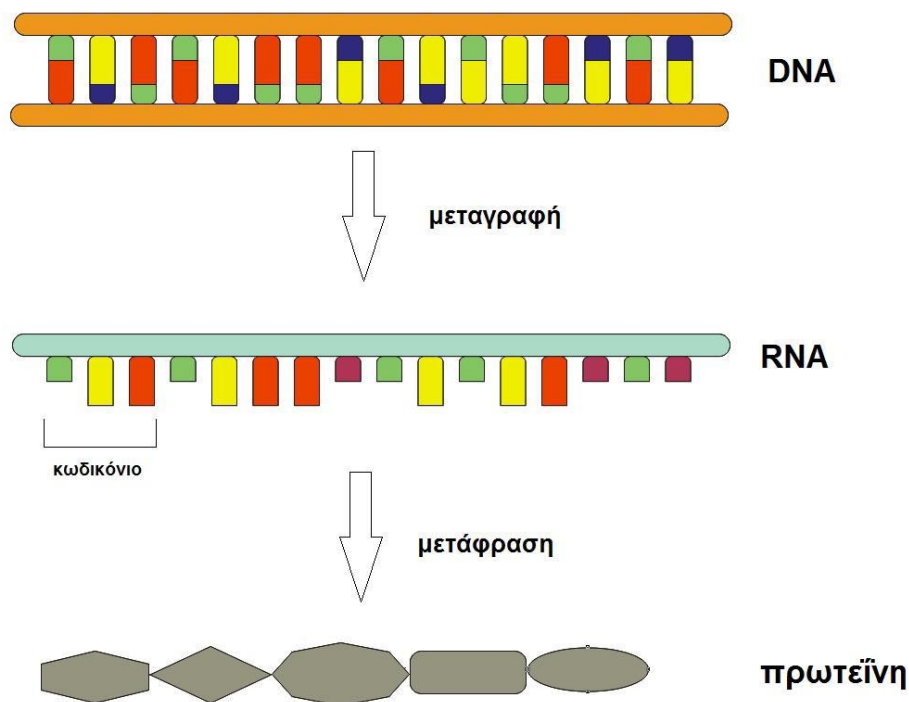
Παρατηρούμε με βάση τα παραπάνω ότι το DNA δεν διαφέρει ιδιαίτερα από το RNA. Όμως μια σημαντική δομική διαφορά είναι ότι το RNA είναι μονόκλωνο και έχει τη δυνατότητα να αναδιπλώνεται όπως οι πολυπεπτιδικές αλυσίδες. Έτσι το RNA εκτός από τον ενδιάμεσο σταθμό κατά την παραγωγή πρωτεϊνών μπορεί να επιτελέσει και δομικές, πληροφοριακές ή καταλυτικές λειτουργίες για το ίδιο το κύτταρο ([8]).

Το ερώτημα στο οποίο μένει να απαντήσουμε στη φάση αυτή είναι πώς από την αλληλουχία των τεσσάρων διαφορετικών ριβονουκλεοτιδίων του RNA μπορεί να κωδικοποιηθεί η αλληλουχία των 20 διαφορετικών αμινοξέων που συνθέτουν κάθε πρωτεΐνη. Η διαδικασία αυτή της μετάβασης από τα νουκλεοτίδια του RNA στα αμινοξέα των βάσεων ονομάζεται *μετάφραση* (*translation*) ([8]).

Προφανώς, δεδομένου του ότι έχουμε 20 αμινοξέα και μόνο 4 βάσεις η αντιστοιχία αμινοξέων και βάσεων αποκλείεται να είναι ένα προς ένα, αλλά στη μετάφραση παρεμβάλλεται κάποια κωδικοποίηση. Το σύνολο των κανόνων που εφαρμόζονται κατά τη μετάφραση ονομάζεται γενετικός κώδικας (*genetic code*) ([8]).

Οι επιστήμονες, μετά από έρευνες, ανακάλυψαν ότι για την παραγωγή των πρωτεϊνών το RNA διαβάζεται σειριακά και κάθε αμινοξύ αντιστοιχεί σε μια σειρά από τρία νουκλεοτίδια. Βέβαια δεδομένου ότι το RNA αποτελείται από τέσσερα διαφορετικά νουκλεοτίδια, οι συνδυασμοί από τρία νουκλεοτίδια που μπορεί να εμφανιστούν είναι  $4^3=64$ , δηλαδή [AAA, AUA, AAU, AGC...] ([8]).

Δεδομένου ότι οι δυνατοί συνδυασμοί είναι περισσότεροι από τα κωδικοποιούμενα αμινοξέα, οι επιστήμονες απέδειξαν ότι ο κώδικας αυτός υπερκαλύπτει τις ανάγκες και κάποια αμινοξέα παράγονται με περισσότερους από έναν συνδυασμούς. Κάθε συνδυασμός τριών νουκλεοτιδίων πάντως κωδικοποιεί τουλάχιστον ένα αμινοξύ και αποκαλείται από τους ερευνητές κωδικόνιο (codon). Σήμερα έχει αποκρυπτογραφηθεί το σύνολο των αντιστοιχίσεων μεταξύ κωδικονίων και αμινοξέων ([8]).



Εικόνα 21: Μεταγραφή / μετάφραση ([8])

### 2.4.3 Σηματοδότηση κυττάρων

#### **Βασικοί Ορισμοί**

---

Η μετάδοση σημάτων μεταξύ κυττάρων (Cell Signaling Transduction) είναι μέρος ενός πιο σύνθετου συστήματος επικοινωνίας, το οποίο διέπει βασικές κυτταρικές δραστηριότητες και προσανατολίζει κυτταρικές πράξεις. Η δυνατότητα των κυττάρων να αντιλαμβάνονται και αντιδρούν ανάλογα στο μικρο – περιβάλλον τους, είναι θεμελιώδης αρχή της ανάπτυξης, της ανάπτυξης των ιστών και της ανοσίας και ομοιοσύστασης των ιστών. Τυχόντα σφάλματα στη διεργασία μετάδοσης κυτταρικών πληροφοριών ευθύνονται για ασθένειες, όπως ο καρκίνος, η αυτοανοσία και η υπεργλυκαιμία (σάκχαρο). Η πλήρης και σωστή κατανόηση της μεταγωγής σημάτων μεταξύ κυττάρων, μπορεί να οδηγήσει σε αποτελεσματική αντιμετώπιση ασθενειών και, θεωρητικά, σε δημιουργία τεχνητών ιστών και οργάνων ([44]).

Παραδοσιακά, η επιστήμη της βιολογίας έχει εστιάσει στη μελέτη των επιμέρους τμημάτων των μονοπατιών μεταγωγής σημάτων μεταξύ κυττάρων. Η έρευνα στον τομέα της συστημικής βιολογίας μας βοηθάει να κατανοήσουμε καλύτερα, την υποκείμενη δομή των δικτύων σηματοδότησης κυττάρων και πως οι αλλαγές σε αυτά τα δίκτυα μπορεί να επηρεάσουν τη μετάδοση και τη ροή της πληροφορίας. Τέτοιου τύπου δίκτυα είναι πολύπλοκα συστήματα σε ό,τι αφορά την οργάνωσή τους και μπορούν να επιδείξουν ένα σημαντικό αριθμό αναδυόμενων ιδιοτήτων, όπως η βιοσταθερότητα και η υπερευαίσθητοποίηση. Η ανάλυση αυτών των δικτύων απαιτεί ένα συνδυασμό πειραματικών και θεωρητικών προσεγγίσεων, συμπεριλαμβανομένων της ανάπτυξης και της ανάλυσης της προσομοιώσεως και της μοντελοποίησης των αντίστοιχων βιολογικών συστημάτων ([12]).

#### **Κυτταρική επικοινωνία**

---

Για να επιβιώσει ένα κύτταρο μέσα στο πολύπλοκο περιβάλλον του απαιτείται να έχει τη δυνατότητα να ερμηνεύει ποικίλα σήματα που εκπέμπονται από άλλα κύτταρα και να προσαρμόσει ανάλογα τη συμπεριφορά του. Η ιδιότητα αυτή των κυττάρων είναι κρίσιμη, από τη σωστή ανάπτυξη του εμβρυϊκού οργανισμού μέχρι και την καθημερινή λειτουργία ενός πλήρως ανεπτυγμένου ατόμου ([12]).

Στο σημείο αυτό θα εξετάσουμε συνοπτικά τους τρόπους με τους οποίους τα κύτταρα επικοινωνούν μεταξύ τους καθώς και με ποιους βασικούς τρόπους διαδίδεται και αξιοποιείται η πληροφορία από το ίδιο το κύτταρο. Ξεκινώντας από κάποιους βασικούς ορισμούς, το κύτταρο το οποίο εκπέμπει την πληροφορία ονομάζεται σηματοδοτικό κύτταρο (signaling cell), ενώ το κύτταρο το οποίο λαμβάνει την πληροφορία ονομάζεται κύτταρο στόχος (target cell). Τα σήματα που μεταδίδονται ανάμεσα στα κύτταρα είναι συνήθως συγκεκριμένα χημικά μόρια. Η διαδικασία της ανταλλαγής

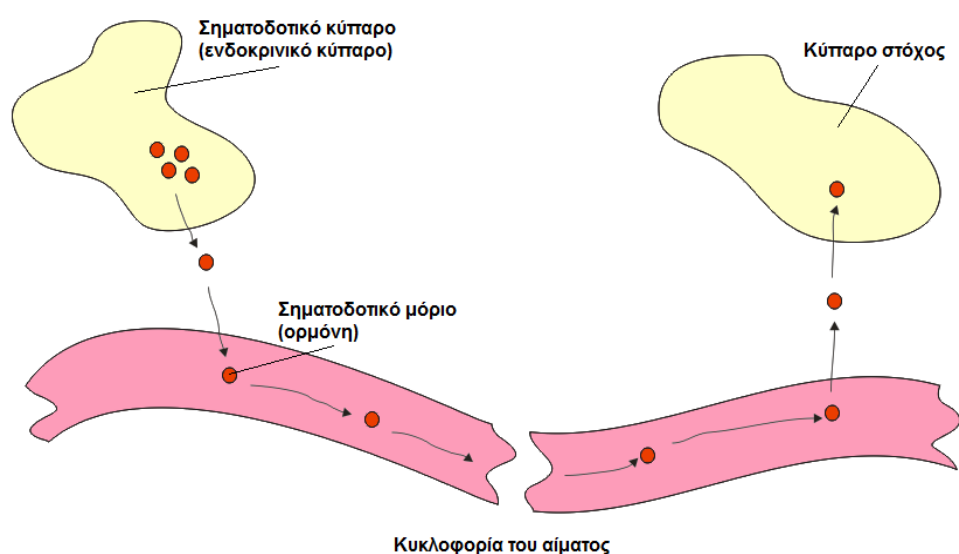


πληροφοριών μεταξύ των κυττάρων αλλά και η διάδοση της πληροφορίας μέσα στο κύτταρο δέκτη, ονομάζεται κυτταρική σηματοδότηση (cell signaling) ([12]).

Τα σήματα που αποστέλλονται ανάμεσα στα κύτταρα ενός πολυκύτταρου οργανισμού μπορεί να είναι πρωτεΐνες, πεπτιδία, αμινοξέα, νουκλεοτίδια, στεροειδή, παράγωγα λιπαρών οξέων ή ακόμα και αέριες ουσίες σε διάλυμα. Ανάλογα με την ταχύτητα και την επιλεκτικότητα με την οποία πρέπει να παραδοθούν τα σήματα στους παραλήπτες τους, διακρίνουμε τους παρακάτω τέσσερις τρόπους με τους οποίους επιτυγχάνεται η επικοινωνία μεταξύ των κυττάρων ([8]).

### ➤ Ενδοκρινής επικοινωνία

Αυτός ο τρόπος επικοινωνίας είναι ο πιο κοινός και περιλαμβάνει τη μετάδοση ενός σήματος σε όλον τον οργανισμό μέσω της κυκλοφορίας του αίματος. Τα μόρια που χρησιμοποιούνται στην περίπτωση αυτή ως σήματα ονομάζονται ορμόνες (hormones). Κάποια κύτταρα λοιπόν, που ονομάζονται ενδοκρινή, παράγουν τα σήματα, που ονομάζονται ορμόνες, και τα εκκρίνουν στο αίμα το οποίο στη συνέχεια τα μεταφέρει σε ολόκληρο τον οργανισμό ([8]).

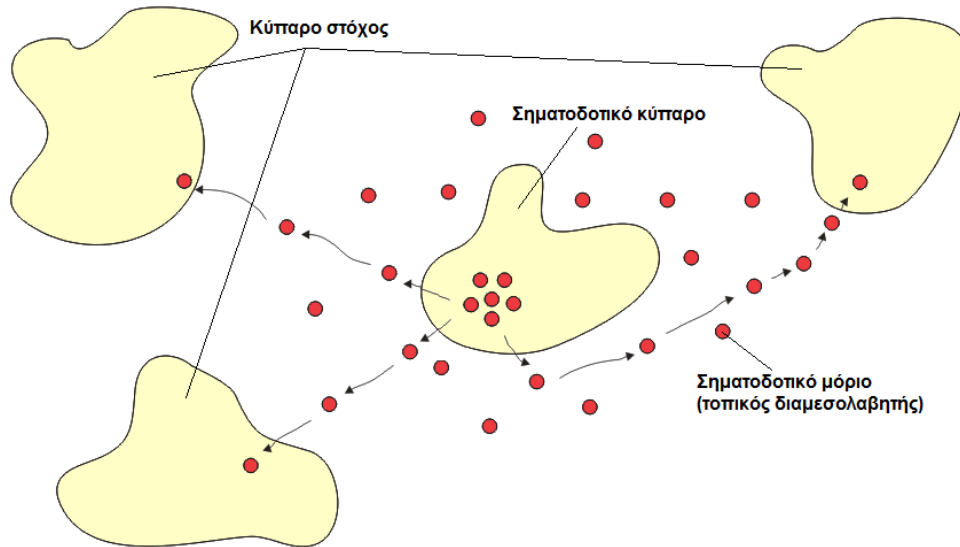


Εικόνα 22: Ενδοκρινής επικοινωνία ([8])

### ➤ Παρακρινής επικοινωνία

Αυτός ο τρόπος επικοινωνίας μεταξύ των κυττάρων δεν είναι τόσο κοινός και αφορά την μετάδοση ενός σήματος τοπικά, όπως στην περίπτωση του ελέγχου μιας φλεγμονής ή στην επούλωση κάποιου τραύματος. Στην περίπτωση αυτή τα σηματοδοτικά μόρια διαχέονται τοπικά στον εξωκυτταρικό χώρο, χωρίς να εισέρχονται στην κυκλοφορία του αίματος.

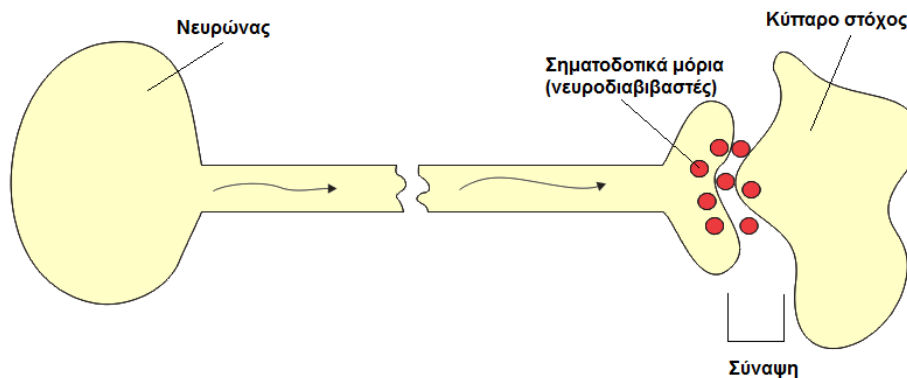
Έτσι τα μόρια αυτά παραμένουν κοντά στο κύτταρο που τα εκκρίνει και δρουν ως τοπικοί διαμεσολαβητές (local mediators) επηρεάζοντας σταδιακά τα γειτονικά κύτταρα ([8]).



Εικόνα 23: Παρακρινής επικοινωνία ([8])

➤ Νευρωνική επικοινωνία

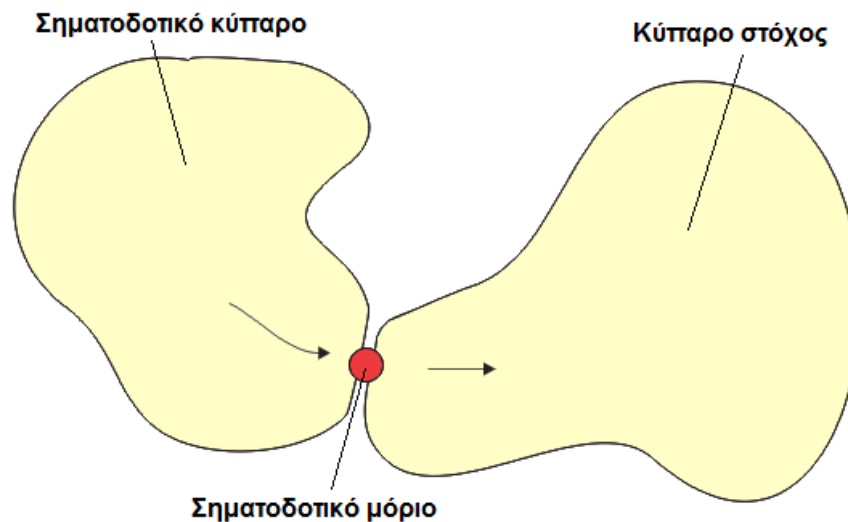
Αυτός ο τρόπος επικοινωνίας μοιάζει με την ενδοκρινή επικοινωνία στο γεγονός ότι τα σήματα διανύουν σχετικά μεγάλες αποστάσεις μέσα στο σώμα, αλλά διαφέρει ριζικά στο γεγονός ότι οι ταχύτητες μετάδοσης είναι κατά πολύ μεγαλύτερες. Στην περίπτωση αυτή τα σήματα ανταλλάσσονται μεταξύ συγκεκριμένων κυττάρων που έχουν δημιουργήσει εξειδικευμένες συνάψεις τα νευρικά κύτταρα. Όταν ένας νευρώνας στέλνει σήματα στο κύτταρο στόχο αυτά έχουν τη μορφή ηλεκτρικών παλμών και ταξιδεύουν με ταχύτητες της τάξης των 100m/s. Μόλις φτάσουν στο στόχο τους οι ηλεκτρικοί παλμοί διεγείρουν την απόληξη, η οποία με τη σειρά της εκκρίνει σηματοδοτικά μόρια που ονομάζονται νευροδιαβιβαστές (neurotransmitters). Οι νευροδιαβιβαστές διαχέονται στο στενό χώρο της σύναψης και εισέρχονται ταχύτατα στο κύτταρο στόχο ([8]).



Εικόνα 24: Νευρωνική επικοινωνία ([8])

➤ Επικοινωνία εξαρτώμενη από επαφή

Ο τρόπος αυτός επικοινωνίας δεν χρησιμοποιείται παρά μόνο σε ειδικές περιπτώσεις και έχει ιδιαίτερα μικρή διάρκεια. Το χαρακτηριστικό του είναι ότι τα σηματοδοτικά μόρια δεν απελευθερώνονται καθόλου στον εξωκυτταρικό χώρο αλλά μεταφέρονται με άμεση επαφή του ενός κυττάρου με το άλλο. Αυτή η σηματοδότηση παίζει κυρίαρχο ρόλο κατά την εμβρυική ανάπτυξη, όπου κύτταρα πολύ κοντινά μεταξύ τους πρέπει να διαφοροποιηθούν μεταξύ τους και αυτή η πληροφορία πρέπει να μεταδοθεί από το ένα στο άλλο ([8]).



Εικόνα 25: Επικοινωνία εξαρτώμενη από επαφή ([8])

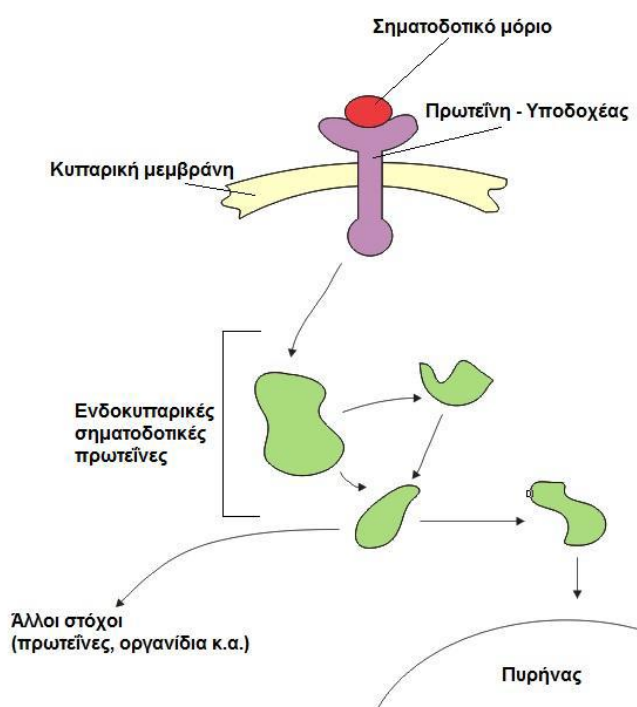
Παραπάνω εξετάσαμε πώς τα σήματα διαδίδονται ανάμεσα στα σηματοδοτικά κύτταρα και στα κύτταρα στόχους. Στο σημείο αυτό θα παρουσιάσουμε πως τα κύτταρα διακρίνουν τα σήματα που απευθύνονται προς αυτά και τα προσλαμβάνουν ([8]).

Για να ανιχνεύσει ένα κύτταρο ένα σηματοδοτικό μόριο θα πρέπει στην επιφάνεια του κυττάρου να υπάρχει μια πρωτεΐνη η οποία θα δράσει ως υποδοχέας (receptor protein). Αν ένα κύτταρο δεν διαθέτει τον συγκεκριμένο υποδοχέα δεν μπορεί να ανταποκριθεί στο αντίστοιχο σήμα. Με αυτό τον τρόπο περιορίζεται ο αριθμός των σημάτων στα οποία μπορεί να αντιδράσει ένα κύτταρο, αφού γύρω του μπορεί να υπάρχουν εκατοντάδες σήματα, αλλά κάθε κύτταρο διαθέτει μόνο μερικές δεκάδες διαφορετικούς υποδοχείς ([8]).

Ακόμα και στην περίπτωση που ένα κύτταρο διαθέτει το συγκεκριμένο υποδοχέα, η διαδικασία μετάδοσης του σήματος στο εσωτερικό του κυττάρου για την επίτευξη του αντίστοιχου σκοπού είναι αρκετά πολύπλοκη. Σε αυτή τη διαδικασία συμμετέχουν, εκτός από την πρωτεΐνη υποδοχέα της κυτταρικής μεμβράνης και μια ομάδα από άλλες ενδοκυτταρικές σηματοδοτικές πρωτεΐνες.

Ανάλογα με το είδος του σήματος, το είδος του κυττάρου στόχου και το ενδοκυτταρικό σύστημα μετάδοσης της πληροφορίας, τα αποτελέσματα καθώς και ο χρόνος εμφάνισής τους, μπορεί να ποικίλουν ([8]).

Επίσης, όπως είδαμε παραπάνω, ένα κύτταρο μπορεί να έχει μερικές δεκάδες υποδοχείς και να δέχεται διαφορετικά σήματα ταυτόχρονα. Αυτά τα σήματα μπορεί να αλληλεπιδρούν μεταξύ καθώς διακινούνται μέσα από το ενδοκυτταρικό σηματοδοτικό δίκτυο. Έτσι η παρουσία ενός σήματος μπορεί να μπλοκάρει πλήρως ή μερικά τη διάδοση ενός άλλου. Αντίστροφα, για την διάδοση ενός σήματος μπορεί να απαιτείται και η παρουσία ενός άλλου. Ο ακριβής τρόπος πάντως με τον οποίο ένα κύτταρο ανταποκρίνεται στα εξωτερικά ερεθίσματα δεν έχει διευκρινιστεί ακόμα και αποτελεί αντικείμενο μελέτης, μεταξύ άλλων, της συστημικής βιολογίας ([8]).



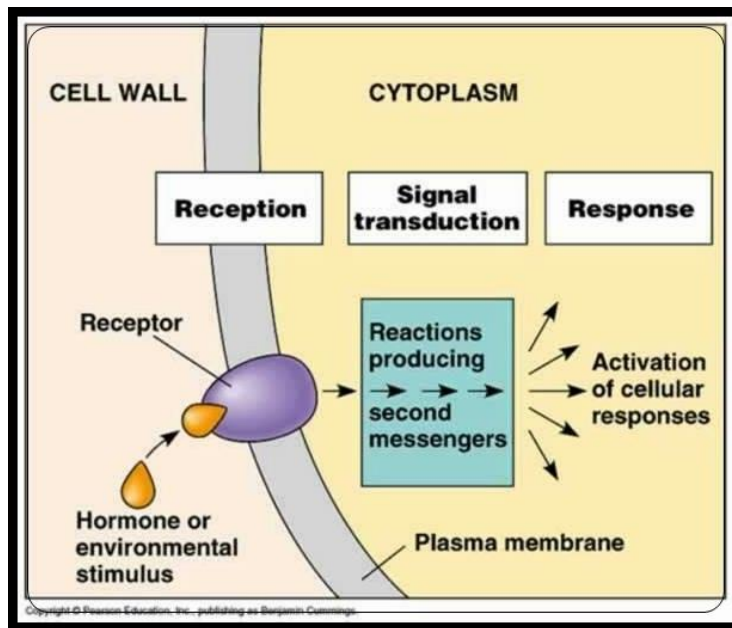
Εικόνα 26: Διάδοση σήματος στο κύτταρο ([8])

Επομένως είδαμε ότι για να ανιχνεύσει ένα κύτταρο ένα σήμα από το περιβάλλον του θα πρέπει κατ' αρχήν να διαθέτει την αντίστοιχη πρωτεΐνη υποδοχέα στην κυτταρική του μεμβράνη. Στον κανόνα αυτό υπάρχουν ορισμένες εξαιρέσεις, αφού ορισμένα σηματοδοτικά μόρια μπορούν να διαπερνούν την κυτταρική μεμβράνη και να επιδρούν απευθείας μέσα στο κύτταρο ([8]).

Ένα παράδειγμα τέτοιων μορίων αποτελούν οι στεροειδείς ορμόνες και οι θυρεοειδείς ορμόνες. Αυτές διαπερνούν ελεύθερα την κυτταρική μεμβράνη και προσδένονται απευθείας σε υποδοχείς που βρίσκονται στο κυτταροδιάλυμα ή

στον πυρήνα. Αυτές οι ορμόνες έχουν συνήθως πολύ σημαντική επίδραση στην λειτουργία ολόκληρου του οργανισμού ([8]).

Όλα όσα αναφέρθηκαν παραπάνω παρουσιάζονται πιο συνοπτικά και εύστοχα στην παρακάτω εικόνα.



Εικόνα 27: Μηχανισμός μετάδοσης σήματος στο κύτταρο ([12])

## 2.5 Μοντελοποίηση μονοπατιών μεταγωγής σημάτων

Το κεφάλαιο αυτό αποτελεί ουσιαστικά ένα από τα πιο βασικά της παρούσης διπλωματικής, μιας και πραγματεύεται το κυρίως θέμα της. Παρόλα αυτά θα αναφερθούμε σε πολλαπλούς τρόπους μοντελοποίησης δικτύων, αλλά και σε θεμελιώδεις αλγορίθμους μοντελοποίησης δικτύων.

### 2.5.1 Εισαγωγή

Τα δίκτυα σηματοδότησης, όπως λαμβάνονται από τη διεθνή βιβλιογραφία είναι γραφικές αναπαραστάσεις των συμβάντων σηματοδότησης, τα οποία διεξάγονται στα κύτταρα. Ο μηχανισμός με τον οποίο η μεταγωγή των σημάτων πραγματοποιείται δεν απεικονίζεται σε αυτά τα δίκτυα. Η μοντελοποίηση των δικτύων σηματοδότησης αναφέρεται στη χρήση του μαθηματικού φορμαλισμού στο να περιγραφεί το πώς το σήμα διαδίδεται από τον ένα κόμβο στον άλλο, καταλήγοντας σε εκτελέσιμα μοντέλα ([24]).

### Εφαρμόσιμες Μεθοδολογίες

- **Συνήθεις διαφορικές Εξισώσεις:** Χρήση συνήθων διαφορικών εξισώσεων για τον υπολογισμό της πορείας του χρόνου των επιπέδων ενεργοποίησης των πρωτεϊνών στα δίκτυα, κατά τη διαταραχή με κάποιο ερέθισμα, φάρμακο ή άλλο παράγοντα. Ουσιαστικά οι διαφορικές εξισώσεις αυτές εκφράζουν τους νόμους της κινηματικής που *διέπουν τις αντιδράσεις της μεταγωγής των σημάτων*. Ο ρυθμός της φωσφορυλίωσης (ή της αποφωσφορυλίωσης) ενός ορισμένου είδους είναι ανάλογος προς τη συγκέντρωση των ανάντη ειδών και αντιστρόφως ανάλογος με τη συγκέντρωση του ([24]).
- **Λογική Boole:** Χρήση των κανόνων της λογικής του Boole για την περιγραφή της διάδοσης ενός σήματος από τον ένα κόμβο στον επόμενο. *Η συνεκτικότητα των πρωτεϊνών μοντελοποιείται μέσω των λογικών πυλών του Boole*. Μία εξωτερική διαταραχή επιβάλλεται σε ένα υποσύνολο κόμβων (αρχικές συνθήκες), έπειτα η λογική του Boole χρησιμοποιείται για τη διάδοση των αρχικών συνθηκών κατάντη ([24]).
- **Θεωρία Γράφων:** Θα αναπτυχθεί εκτενώς παρακάτω

Σαφέστατα υπάρχουν πολλές και διαφορετικές μέθοδοι, οι οποίες χρησιμοποιούνται για τη μοντελοποίηση των δικτύων σηματοδότησης από την κοινότητα της βιοπληροφορικής. Παρόλα αυτά, στην ενότητα αυτή θα αναλυθούν μόνο οι χρησιμοποιηθέντες σε αυτή τη διπλωματική εργασία μέθοδοι.

### 2.5.2 Θεμελιώδης θεωρία δικτύων

Ένας σημαντικός αριθμός προβλημάτων συντίθεται συχνά σε όρους αντικειμένων και σύνδεσης μεταξύ τους. Τα ηλεκτρικά κυκλώματα αποτελούν ένα τέτοιο προφανές παράδειγμα, στο οποίο οι διασυνδέσεις μεταξύ αντικειμένων παίζει ένα σημαντικό ρόλο. Τα στοιχεία των κυκλωμάτων, όπως τα τρανζίστορ, οι αντιστάσεις και οι πυκνωτές είναι περίπλοκα συνδεδεμένα μαζί τους. Τέτοιου τύπου κυκλώματα μπορεί να αντιπροσωπευτούν και να επεξεργασθούν μέσα σε έναν ηλεκτρονικό υπολογιστή, με σκοπό να απαντηθούν απλές ερωτήσεις, όπως “Είναι όλα τα στοιχεία σωστά συνδεδεμένα μεταξύ τους;” ή περίπλοκες ερωτήσεις, όπως “Εάν αυτό το κύκλωμα φτιαχτεί, θα δουλέψει;”. Η απάντηση στην πρώτη ερώτηση εξαρτάται από τις ιδιότητες των διασυνδέσεων, ενώ η απάντηση στη δεύτερη απαιτεί λεπτομερείς πληροφορίες και για τις διασυνδέσεις και για τα αντικείμενα τα οποία συνδέουν ([2]).

Ένα ακόμα τέτοιο παράδειγμα αποτελεί και ο “προγραμματισμός εργασίας”, όπου τα αντικείμενα αποτελούν καθήκοντα να εκτελεστούν (π.χ. κατασκευαστικές διαδικασίες) και οι διασυνδέσεις υποδεικνύουν ποιες εργασίες πρέπει να γίνουν πρώτα και ποιες μετά ([2]).

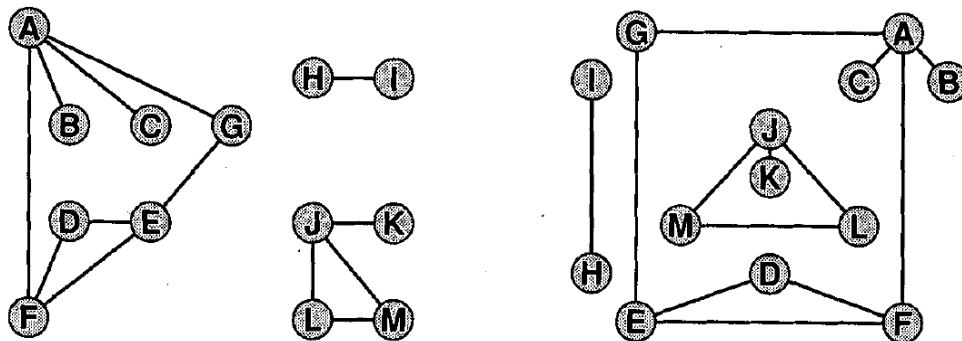
Ένας γράφος είναι ένα μαθηματικό αντικείμενο, το οποίο μοντελοποιεί με ακρίβεια τέτοιου είδους καταστάσεις. Αρχικά, θα μελετήσουμε κάποιες βασικές ιδιότητες των γράφων και έπειτα θα μελετήσουμε μία ποικιλία αλγορίθμων που μας βοηθάνε να αντιμετωπίσουμε προβλήματα σαν αυτά τα οποία αναφέρθηκαν παραπάνω ([2]).

Η *θεωρία των γράφων* αποτελεί ένα βασικό κλάδο των συνδυαστικών μαθηματικών και έχει μελετηθεί εντατικά για πολύ καιρό τώρα. Πολλές σημαντικές και χρήσιμες ιδιότητες των γράφων έχουν αποδειχτεί, αλλά πολλά δύσκολα προβλήματα μένουν να ακόμα να επιλυθούν ([2]).

Η θεωρία των αλγορίθμων μόλις πρόσφατα ξεκίνησε να εξετάζεται από μία αλγοριθμική άποψη. Αν και κάποιιοι από τους θεμελιώδεις αλγορίθμους είναι αρκετά παλιοί, πολλοί από τους πιο ενδιαφέροντες ανακαλύφθηκαν τα τελευταία δέκα χρόνια. Ακόμα και οι ασήμαντοι γραφικοί αλγόριθμοι οδηγούν σε ενδιαφέροντα υπολογιστικά προγράμματα και οι σημαντικοί αλγόριθμοι που θα εξεταστούν στην παρούσα και στις παρακάτω ενότητες, ανήκουν στους πιο «κομψούς» και ενδιαφέροντες (αν και δύσκολο να κατανοηθούν) γνωστών αλγορίθμων ([2]).

## Βασική ονοματολογία δικτύων

Ένας γράφος αποτελείται από κορυφές και ακμές. Οι κορυφές είναι απλά αντικείμενα, τα οποία μπορούν να έχουν ονόματα και άλλες ιδιότητες; μία ακμή αποτελεί ουσιαστικά μία σύνδεση μεταξύ δύο ακμών. Ουσιαστικά κάποιος θα μπορούσε να σχεδιάσει ένα γράφο με σημεία για κορυφές και με γραμμές που θα τα ενώνουν για ακμές, αλλά θα πρέπει να αναφερθεί πως ο γράφος είναι ορισμένος ανεξαρτήτως αναπαράστασης ([2]).



Εικόνα 28: Δύο αναπαράστασεις του ίδιου γράφου ([2])

Τα δύο σχήματα τα οποία παρουσιάζονται στις παραπάνω εικόνες αναπαριστούν ουσιαστικά τον ίδιο γράφο. Ορίζουμε αυτό το γράφο, θεωρώντας ότι συνίσταται από το σύνολο των κορυφών A B C D E F G H I J K L M και το σύνολο των ακμών μεταξύ των κορυφών αυτών AG AB AC LM JM JL JK ED FD HI FE AF GE.

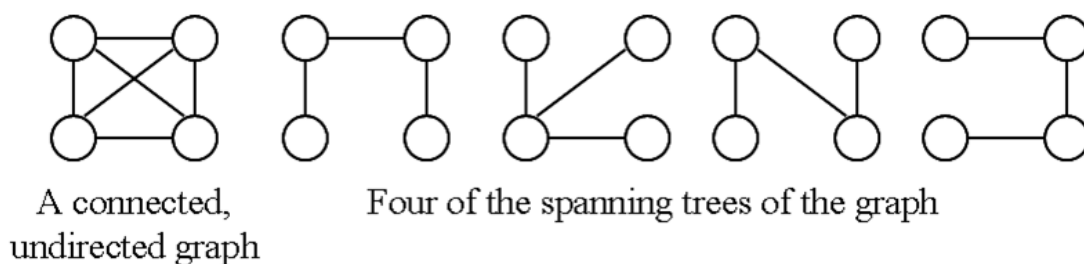
Ένα μονοπάτι από την κορυφή x στην κορυφή y σε ένα γράφο αποτελεί μία λίστα κορυφών στην οποία διαδοχικές κορυφές συνδέονται με ακμές στο γράφο. Για παράδειγμα, BAFEG είναι ένα μονοπάτι από την κορυφή B στην κορυφή G, όπως φαίνεται και στην παραπάνω εικόνα.

Ένας γράφος ονομάζεται *συνεκτικός*, εάν υπάρχει ένα μονοπάτι από κάθε κόμβο (κορυφή) σε όλους τους άλλους κόμβους στο γράφο. Διαισθητικά, εάν οι κορυφές ήταν φυσικά αντικείμενα και οι ακμές ήταν χορδές, οι οποίες τις ένωναν, ένας συνδεδεμένος γράφος θα έμενε σε ένα κομμάτι εάν «πιανόταν» από οποιαδήποτε κορυφή. Ένας γράφος, ο οποίος δεν είναι συνδεδεμένος αποτελείται από *συνεκτικά στοιχεία* (*connected components*): Για παράδειγμα στην παραπάνω εικόνα ο γράφος έχει τρία συνδεδεμένα στοιχεία. Ένα *απλό μονοπάτι* είναι ένα μονοπάτι στο οποίο κανένας κόμβος (κορυφή) δεν επαναλαμβάνεται (για παράδειγμα BAFEGAC δεν είναι απλό μονοπάτι). *Κύκλος* ονομάζεται ένα μονοπάτι, το οποίο είναι απλό αλλά ο τελευταίος κόμβος είναι ο ίδιος με τον τελευταίο (ουσιαστικά ένα απλό μονοπάτι από έναν κόμβο μέχρι πίσω στον εαυτό του): το μονοπάτι AFEGA είναι κύκλος ([2]).



Ένα μονοπάτι χωρίς κύκλους αποκαλείται *δέντρο*. Ένα σύνολο μη – συνεκτικών δέντρων ονομάζεται *δάσος*. Ένα ζευγνύον δέντρο ή όπως αλλιώς αναφέρεται στην ελληνική βιβλιογραφία *δέντρο κάλυψης (spanning tree)* ενός γράφου είναι ουσιαστικά ένας υπό - γράφος, ο οποίος περιλαμβάνει όλες τις κορυφές, αλλά μόνο αρκετές ακμές για το σχηματισμό ενός δέντρου. Για παράδειγμα, οι ακμές AB AD AF DE EG σχηματίζουν ένα ζευγνύον δέντρο για το μεγάλο συστατικό του γράφου της εικόνας 21 ([22]).

Στην παρακάτω εικόνα 22 φαίνεται ένας γράφος και τα όλα τα πιθανά εκτεταμένα δέντρα του. Τονίζεται πως η παρούσα εικόνα αποτελεί ένα ιδιαίτερα απλουστευμένο παράδειγμα.



Εικόνα 29: Ζευγνύον δέντρο γράφου ([51])

Σημειώνουμε πως αν προσθέσουμε οποιαδήποτε ακμή σε ένα δέντρο, πρέπει να σχηματίζεται ένας κύκλος (επειδή υπάρχει ήδη ένα μονοπάτι μεταξύ των δύο κορυφών που ενώνει). **Κάτι πολύ σημαντικό είναι πως ένα δέντρο  $V$  κορυφών έχει ακριβώς  $V-1$  ακμές.** Εάν ένας γράφος με  $V$  κορυφές έχει λιγότερες από  $V-1$  ακμές, τότε δεν μπορεί να είναι συνδεδεμένο. Εάν έχει περισσότερες από  $V-1$  ακμές, τότε πρέπει να έχει οπωσδήποτε κύκλο (αλλά εάν έχει ακριβώς  $V-1$  ακμές, δεν χρειάζεται να είναι δέντρο) ([2]).

Από δω και πέρα, για ένα δεδομένο γράφο θα αναφερόμαστε στον **αριθμό των κορυφών ως  $V$**  και στον **αριθμό των ακμών ως  $E$** . Να σημειώσουμε πως το  $E$  μπορεί να κυμαίνεται από 0 μέχρι την τιμή  $\frac{1}{2} \cdot V \cdot (V - 1)$ . Γράφοι με όλες τις ακμές παρούσες, ονομάζονται **ολοκληρωμένοι** γράφοι; γράφοι με σχετικά μικρό αριθμό ακμών (ας πούμε μικρότερους από  $V \cdot \log V$ ), ονομάζονται **αραιοί**. Γράφοι με μικρό αριθμό πιθανών ακμών να λείπουν από το γράφο ονομάζονται **πυκνοί**. Αλγόριθμοι οι οποίοι χρειάζονται  $V^2$  βήματα προτιμούνται για τους πυκνούς αλγόριθμους, ενώ αλγόριθμοι που χρειάζονται  $(E + V) \cdot \log E$  βήματα, προτιμούνται για αραιούς γράφους ([2]).

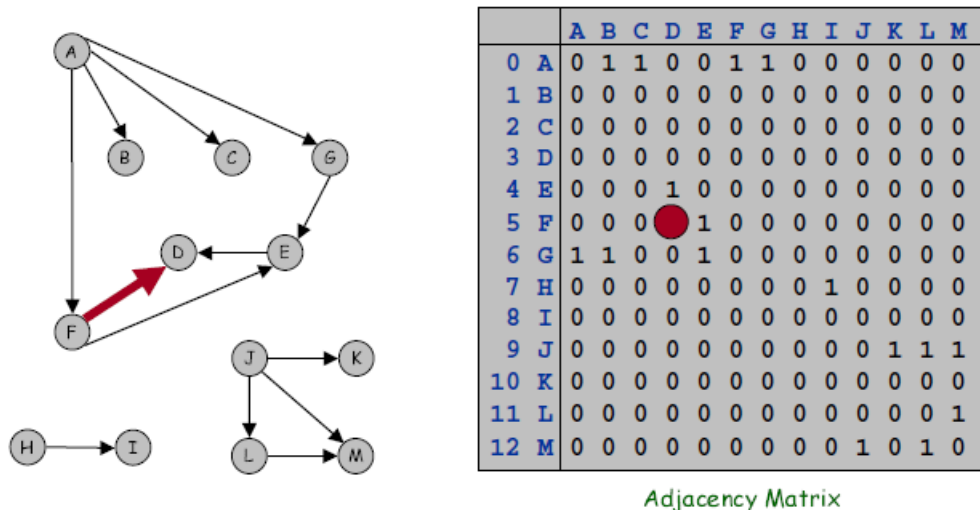
Οι γράφοι όπως έχουν ορισθεί μέχρι αυτού του σημείου, ονομάζονται **μη-προσανατολισμένοι (undirected) γράφοι**, ο πιο απλός τύπος γράφου. Υπάρχουν και άλλοι τύποι γράφων πιο πολύπλοκοι, στους οποίους περισσότερες πληροφορίες σχετίζονται με τους κόμβους και τις ακμές. Στους γράφους με βάρος (**weighted graphs**) ακέραιοι αριθμοί μπορεί να αντιστοιχίζονται σε κάθε ακμή για να αναπαραστήσουν, ως πούιμε, αποστάσεις ή κόστη. Στους **προσανατολισμένους γράφους (directed graphs)**, οι ακμές είναι μίας κατεύθυνσης: μία ακμή μπορεί να πάει από το  $x$  μέχρι το  $y$ , αλλά ποτέ από το  $y$  μέχρι το  $x$ . Οι προσανατολισμένοι γράφοι με βάρος (**directed weighted graphs**) ονομάζονται συχνά **δίκτυα (networks)**. Με αυτήν την κατηγορία γράφων ασχολείται και η παρούσα διπλωματική ([2]).

### Αναπαράσταση γράφων

Με σκοπό να προσπελάσουμε ένα γράφο με ένα υπολογιστικό κώδικα, πρέπει πρώτα να τον αναπαραστήσουμε μέσα στον υπολογιστή. Θα κοιτάξουμε δύο συνήθειες χρησιμοποιούμενες αναπαραστάσεις: η επιλογή μεταξύ των εξαρτάται πρωτίστως από το εάν ο γράφος είναι αραιός ή πυκνός, αν και συνήθως, η φύση των λειτουργιών που πρέπει να εκτελεστούν παίζει ιδιαίτερα σημαντικό ρόλο ([2]).

Το πρώτο βήμα στην αναπαράσταση ενός γράφου είναι να αντιστοιχήσουμε τα ονόματα κάθε κορυφής με ακέραιους μεταξύ 1 έως  $V$ . Ο κύριος λόγος για να το κάνουμε αυτό, είναι να έχουμε γρήγορη πρόσβαση σε πληροφορίες, οι οποίες αντιστοιχούν σε κάθε κορυφή, χρησιμοποιώντας ευρετήριο πίνακα. Οποιαδήποτε στάνταρ μέθοδος αναζήτησης μπορεί να χρησιμοποιηθεί για αυτό το σκοπό: για παράδειγμα, μπορούμε να μεταφράσουμε τα ονόματα των κορυφών με ακέραιους μεταξύ 1 έως  $V$  διατηρώντας έναν πίνακα, ο οποίος μπορεί να προσπελαστεί ώστε να ευρεθεί ο ακέραιος που αντιστοιχεί σε οποιαδήποτε δεδομένο όνομα κορυφής. Έστω πως υπάρχει μία δεδομένη συνάρτηση *περιεχόμενο (index)* για τη μετατροπή από τα ονόματα κορυφών σε ακέραιες θέσεις μεταξύ 1 έως  $V$  και μία συνάρτηση *όνομα (name)* για τη μετατροπή των ακεραίων αυτών σε ονόματα κορυφών ([2]).

Η πιο απλή αναπαράσταση για γράφους είναι η αποκαλούμενη αναπαράσταση με το «**μητρώο γεινίασης**» (**adjacency matrix**). Πιο αναλυτικά, δημιουργείται ένα μητρώο διαστάσεων  $V \times V$  στον οποίο διατηρούνται τιμές boolean (0 ή 1), με  $a[x][y]$  να ισούται με 1, εάν υπάρχει μία ακμή από το κόμβο  $x$  στον κόμβο  $y$  και 0 στην αντίθετη περίπτωση. Παρακάτω φαίνεται ένα καθαρό παράδειγμα ενός τέτοιου πίνακα για το γράφο – παράδειγμα της εικόνας 21 ([2]).



Εικόνα 30: Πίνακας γειτνίασης (adjacency matrix) ([22])

Όπως προαναφέρθηκε και παραπάνω, ένας γράφος ορίζεται ως ένα σύνολο κόμβων και ένα σύνολο ακμών να τους συνδέει. Για να πάρουμε ένα γράφο σαν είσοδο, πρέπει να διαλέξουμε μία μορφή για να για να διαβάσουμε αυτά τα δεδομένα. Μία πρώτη επιλογή είναι να χρησιμοποιήσουμε το μητρώο γειτνίασης σαν είσοδος, αλλά όπως έχει εξετασθεί αυτή η επιλογή είναι ακατάλληλη για αραιούς γράφους. Αντί για αυτού, μπορούμε να χρησιμοποιήσουμε μία πιο άμεση μορφή: αρχικά διαβάζουμε τα ονόματα των κορυφών, μετά ζευγαρώνουμε τα ονόματα των κορυφών (έτσι ώστε να ορίζονται ακμές). Όπως αναφέρθηκε και παραπάνω, ένας εύκολος τρόπος να προχωρήσουμε είναι να διαβάσουμε τα ονόματα των κορυφών και να τα καταχωρήσουμε σε έναν πίνακα και να αντιστοιχήσουμε σε κάθε όνομα κορυφής έναν ακέραιο για να χρησιμοποιηθεί για την πρόσβαση του πίνακα κορυφών – περιεχομένων, όπως είναι και ο πίνακας γειτνίασης. Η  $i$ -οστή κορυφή μπορεί να αντιστοιχηθεί στον ακέραιο  $i$ . Για απλότητα στα προγράμματά μας, στην αρχή διαβάζουμε τους ακέραιους  $V$  και  $E$ , μετά τα ονόματα των κορυφών και των ακμών ([2]).

### 2.5.3 Δίκτυα στη βιολογία

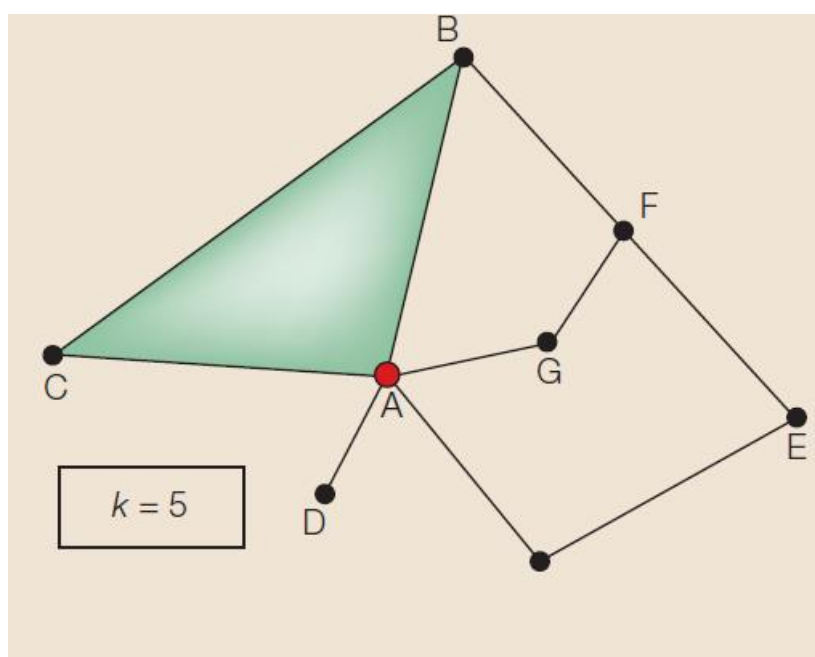
Η συμπεριφορά των πιο πολύπλοκων συστημάτων, από το κύτταρο μέχρι το διαδίκτυο, αναδύεται από την ενορχηστρωμένη ενεργοποίηση πολλών μερών, τα οποία αντιδρούν το ένα με το άλλο μέσω ζευγών αλληλεπίδρασης. Σε ένα εξαιρετικά αφηρημένο επίπεδο, τα μέρη μπορεί να μειωθούν σε μία σειρά κόμβων οι οποίοι συνδέονται ο ένας με τον άλλον με γραμμές, με την κάθε γραμμή να αντιπροσωπεύει τις αντιδράσεις μεταξύ των δύο μερών. Οι κόμβοι αυτοί και οι γραμμές αυτές σχηματίζουν ένα **δίκτυο (network)** ή σε μία πιο επίσημη μαθηματική γλώσσα, ένα **γράφο (graph)** ([24]).

Η θεωρία των δικτύων στη βιολογία προσφέρει μία ποσοτική περιγραφή των δικτύων, η οποία χαρακτηρίζει ποικίλα βιολογικά συστήματα. Παρακάτω ορίζουμε τα πιο βασικά κριτήρια που μας επιτρέπουν να συγκρίνουμε και χαρακτηρίζουμε διαφορετικά και πολύπλοκα δίκτυα.

## Βαθμός

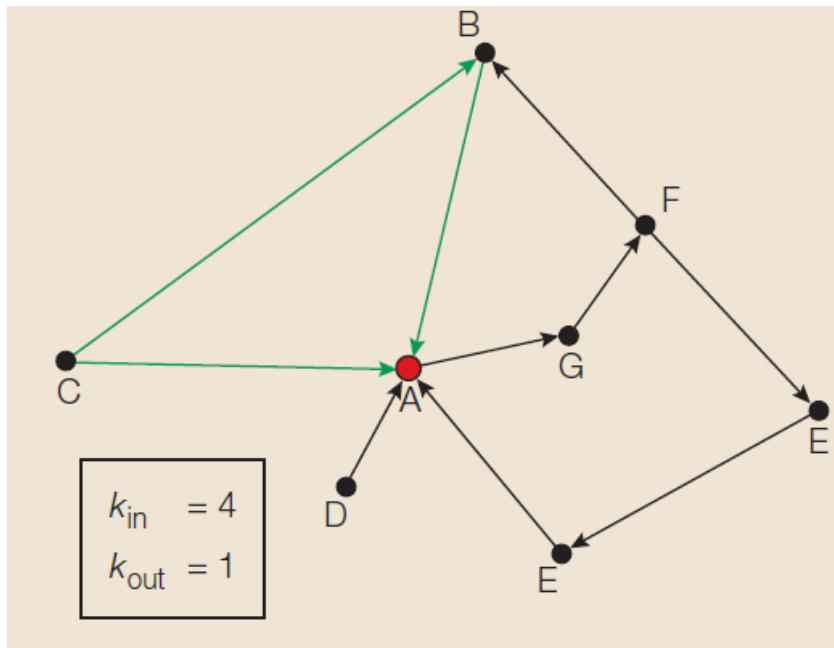
---

Το πιο θεμελιώδες χαρακτηριστικό ενός κόμβου είναι ο βαθμός του (ή συνεκτικότητα),  $k$ , η οποία δηλώνει πόσες γραμμές έχει ένας κόμβος σε άλλους κόμβους. Για παράδειγμα, στο μη – προσανατολισμένο δίκτυο το οποίο παρουσιάζεται στην παρακάτω εικόνα, ο κόμβος A έχει βαθμό  $k=5$ .



Εικόνα 31: Μη προσανατολισμένο δίκτυο ([24])

Στα δίκτυα στα οποία κάθε γραμμή έχει μία συγκεκριμένη κατεύθυνση (όπως φαίνεται και στην παρακάτω εικόνα), υπάρχει ένας εισερχόμενος βαθμός,  $k_{in}$ , ο οποίος υποδηλώνει τον αριθμό των γραμμών που υποδεικνύουν σε ένα κόμβο, και ένας εξερχόμενος βαθμός,  $k_{out}$ , ο οποίος υποδηλώνει τον αριθμό των γραμμών που ξεκινούν από αυτόν. Για παράδειγμα, ο κόμβος A στην παρακάτω εικόνα έχει  $k_{in}=4$  και  $k_{out}=1$ . Ένα μη – προσανατολισμένο δίκτυο με  $N$  κόμβους και  $L$  γραμμές χαρακτηρίζεται από ένα μέσο βαθμό  $\langle k \rangle = 2L/N$  (όπου το σύμβολο  $\langle \rangle$  υποδηλώνει το μέσο όρο) ([24]).

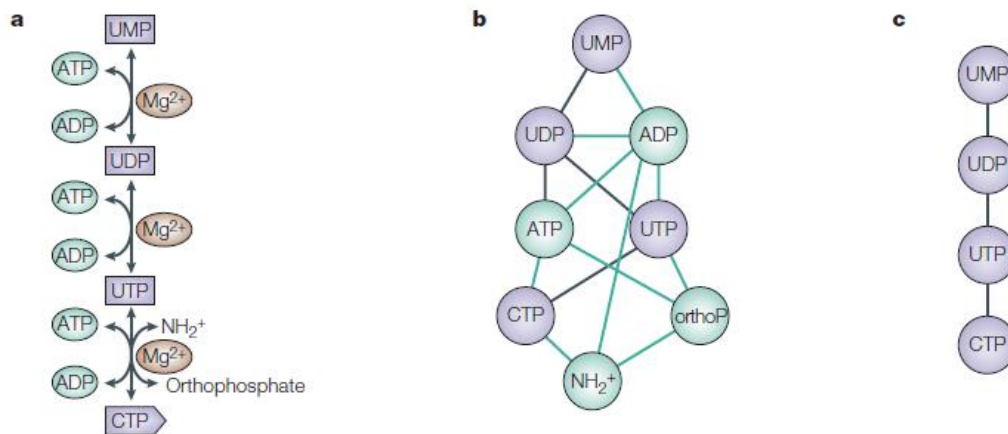


Εικόνα 32: Προσανατολισμένο δίκτυο ([24])

## Κατανομή βαθμού

Η κατανομή βαθμού,  $P(k)$ , υποδηλώνει την πιθανότητα ένας επιλεγμένος κόμβος να έχει ακριβώς  $k$  γραμμές. Η πιθανότητα  $P(k)$  λαμβάνεται μετρώντας τον αριθμό των κόμβων  $N(k)$  με  $k=1,2,\dots$  γραμμές και διαιρώντας τον με το συνολικό αριθμό κόμβων  $N$ . Η κατανομή βαθμού μας επιτρέπει να διακρίνουμε μεταξύ των διαφορετικών κατηγοριών των δικτύων ([24]).

Η καθιέρωση της ταυτότητας διάφορων κυτταρικών δικτύων δεν είναι ασήμαντη. Φυσικές αντιδράσεις μεταξύ μορίων, αντιδράσεις όπως πρωτεΐνη - πρωτεΐνη, πρωτεΐνη - νουκλεϊκό οξύ και πρωτεΐνη - μεταβολίτης, μπορεί εύκολα να αντιληφθεί εννοιολογικά χρησιμοποιώντας την ονοματολογία κόμβοι - γραμμές. Παρόλα αυτά, πιο πολύπλοκες λειτουργικές αντιδράσεις μπορούν επίσης να συμπεριληφθούν σε αυτού του τύπου την αναπαράσταση. Για παράδειγμα, μικρό -μοριακά υποστρώματα μπορούν να θεωρηθούν σαν τους κόμβους ενός μεταβολικού δικτύου και οι ένζυμο - καταλυτικές αντιδράσεις που μετατρέπουν τον ένα μεταβολίτη σε έναν άλλον, σαν τις γραμμές του, όπως φαίνεται και στην παρακάτω εικόνα ([24]).



Εικόνα 33: Χαρακτηριστικά μεταβολικά δίκτυα ([24])

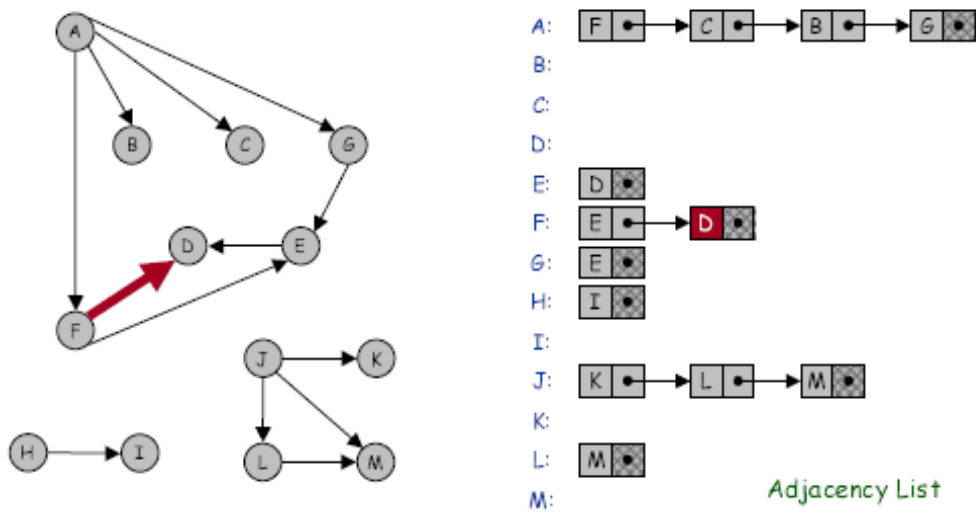
### 2.5.4 Προσανατολισμένα δίκτυα

Όπως έχει ήδη αναφερθεί και παραπάνω, τα **προσανατολισμένα δίκτυα (directed graphs)** είναι δίκτυα των οποίων οι ακμές που ενώνουν τους κόμβους είναι μόνο μίας κατεύθυνσης. Αυτό το χαρακτηριστικό κάνει πολλές φορές ιδιαίτερα δύσκολο τον προσδιορισμό κάποιες συγκεκριμένες ιδιότητες των δικτύων. Τα δίκτυα τα οποία πραγματευόμαστε στην παρούσα διπλωματική ανήκουν σε αυτή την κατηγορία δικτύων : της μίας κατεύθυνσης. Για το λόγο αυτό, στην υπό-ενότητα αυτή αναλύεται πλήρως αυτός ο τύπος δικτύων ([2]).

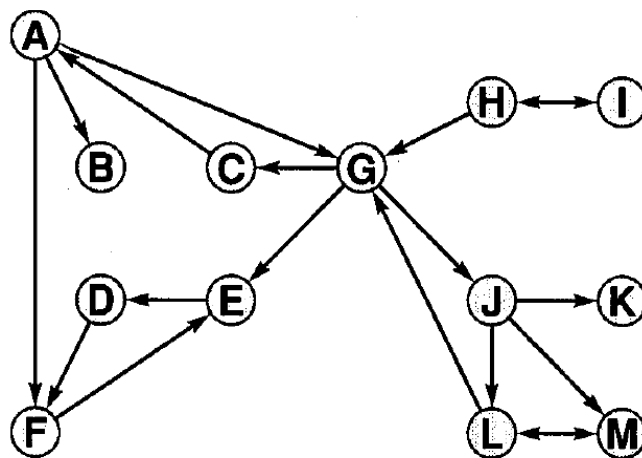
Συχνά η κατεύθυνση της ακμής αντιπροσωπεύει κάποια σχέση προτεραιότητας στην εκάστοτε εφαρμογή που είναι να μοντελοποιηθεί. Για παράδειγμα, ένα προσανατολισμένο δίκτυο μπορεί να χρησιμοποιηθεί για να μοντελοποιηθεί μια γραμμή παραγωγής: οι κόμβοι αντιστοιχούν σε εργασίες που πρέπει να πραγματοποιηθούν και μία γραμμή που υπάρχει από τον κόμβο  $x$  στον κόμβο  $y$ , εάν η εργασία που αντιστοιχεί στον κόμβο  $x$  πρέπει να πραγματοποιηθεί πριν από αυτή που αντιστοιχεί στον κόμβο  $y$  ([2]).

Οι αναπαραστάσεις για τα προσανατολισμένα δίκτυα είναι απλές προεκτάσεις (για την ακρίβεια περιορισμοί), των αναπαραστάσεων των μη-προσανατολισμένων δικτύων. Κατά την αναπαράσταση με λίστα γειτνίασης, κάθε ακμή εμφανίζεται μόνο μία φορά: η ακμή από τον κόμβο  $x$  στον κόμβο  $y$  αναπαρίσταται σαν έναν κόμβο λίστα που περιλαμβάνει το  $y$  στη συνδεδεμένη λίστα που αντιστοιχεί στο  $x$ . Κατά την αναπαράσταση με μητρώο γειτνίασης, πρέπει να δημιουργηθεί ένα μητρώο διαστάσεων  $V \times V$ , με 1 στη θέση της γραμμής  $x$  και στήλης  $y$ , εάν υπάρχει μία ακμή από τον κόμβο  $x$  στον κόμβο  $y$  (το αντίθετο σε αυτήν την περίπτωση δεν ισχύει) ([2]).

Παρακάτω παρουσιάζεται για τα προσανατολισμένα δίκτυα η αναπαράσταση με λίστα γειτνίασης, ενώ ήδη παραπάνω στην εικόνα 24 φαίνεται η αναπαράσταση ενός προσανατολισμένου δικτύου με μητρώο γειτνίασης ([2]).



Εικόνα 34: Αναπαράσταση με λίστα γειννίαςης ([22])



Εικόνα 35: Ένα προσανατολισμένο δίκτυο ([2])

Το παραπάνω δίκτυο αποτελείται από τις ακμές AG AB CA LM JM JL JK ED DF HI FE AF GE GC HG GJ LG IH ML. Η σειρά με την οποία εμφανίζονται οι κορυφές στις συγκεκριμένες ακμές είναι ιδιαίτερα σημαντική: η αναφορά AG περιγράφει μία ακμή που δείχνει από τον κόμβο A προς τον κόμβο G, και όχι το αντίθετο, το οποίο τώρα δεν είναι επιτρεπτό. Παρ' όλα αυτά είναι επιτρεπτό, να υπάρχει μία διπλή ακμή μεταξύ δύο κόμβων, μία προς κάθε

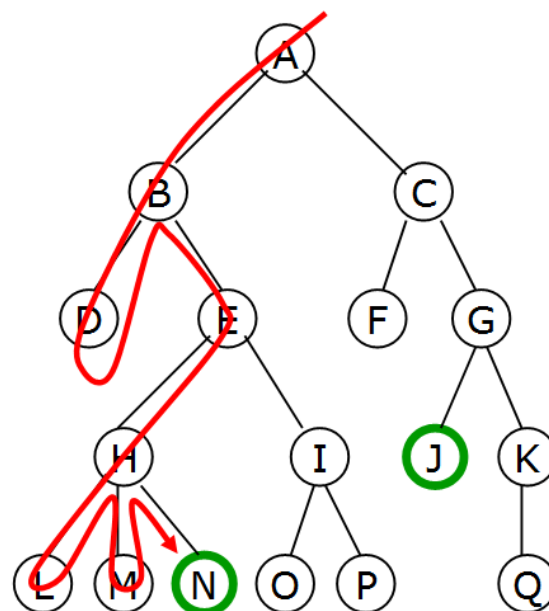
κατεύθυνση. Έτσι, στην παραπάνω εικόνα, φαίνεται καθαρά πως το δίκτυο έχει και HI και IH και LM και ML ακμές).

Παρακάτω παρουσιάζονται βασικοί αλγόριθμοι δικτύων, ειδικά για την περίπτωση των προσανατολισμένων δικτύων.

### Βασικοί Αλγόριθμοι Προσανατολισμένων Δικτύων

#### ➤ Αναζήτηση κατά Βάθος (*Depth-First Search*)

Ο αλγόριθμος κατά βάθος αναζήτησης ταιριάζει πάρα πολύ καλά στην περίπτωση των προσανατολισμένων δικτύων. Στην πραγματικότητα, η λειτουργία του είναι λίγο πιο απλή από την αντίστοιχη των μη-προσανατολισμένων δικτύων, επειδή εξ' ορισμού δεν χρειάζεται να ανησυχούμε με ακμές δύο κατευθύνσεων, εκτός αν υπάρχουν ήδη στην τοπολογία του δικτύου. Παρόλα αυτά, η αναζήτηση των δικτύων έχει κάπως πιο πολύπλοκη δομή. Στην παρακάτω εικόνα παρουσιάζεται μία αναπαράσταση του αλγορίθμου αναζήτησης κατά βάθος ([21]).



Εικόνα 36: Αναζήτηση κατά βάθος ([21])

Όπως φαίνεται και στην παραπάνω εικόνα, η αναζήτηση του δικτύου ξεκινάει από τον κόμβο A και συνεχίζει προς τα κάτω, όπως δείχνει η κόκκινη γραμμή, η οποία επιδεικνύει την πορεία αναζήτησης. Έστω, λοιπόν, πως οι κόμβοι – στόχος είναι οι κόμβοι που είναι κυκλωμένοι με πράσινο χρώμα. Η σειρά



επίσκεψης των κόμβων είναι η παρακάτω: **A B D E H L M N I O P C F G J K Q**. Ο κόμβος **N** επισκέπτεται πριν από τον κόμβο **J**. Ψάχνοντας πρώτα τον **A**, μετά τον **B**, μετά τον **D**, η αναζήτηση υπαναχωρεί και προσπαθεί να βρει μία άλλη διαδρομή από τον κόμβο **B**.

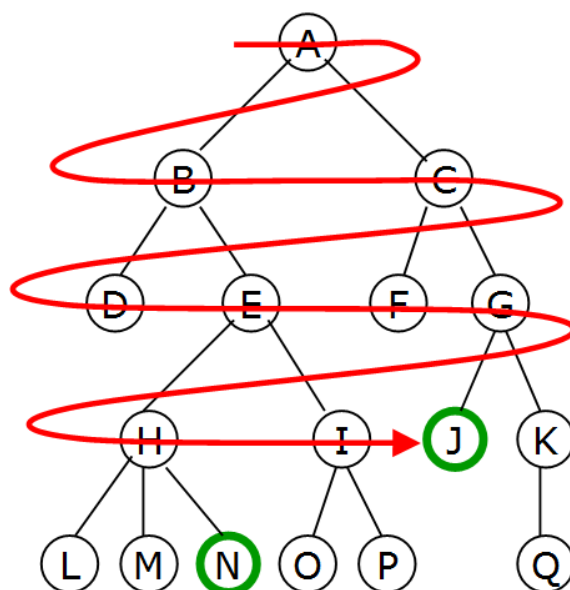
Ας σημειώσουμε σε αυτό το σημείο πως οι φορές των ακμών κάνουν τον αλγόριθμο αναζήτησης κατά βάθος αρκετά διαφορετικό από αυτόν των μη-προσανατολισμένων δικτύων. Για παράδειγμα, ακόμα και αν το πρωτότυπο δίκτυο ήταν συνδεδεμένο, η δομή της κατά βάθος αναζήτησης που ορίζεται από τις κόκκινες γραμμές δεν είναι συνδεδεμένη: πρόκειται για δάσος, όχι για δέντρο ([21]).

Σε ό,τι αφορά τα μη-προσανατολισμένα δίκτυα, μας ενδιαφέρουν κυρίως οι ιδιότητες συνδεσιμότητας των προσανατολισμένων δικτύων. Θα μας ενδιέφερε να είμαστε σε θέση να απαντήσουμε σε ερωτήματα όπως: “Υπάρχει ένα προσανατολισμένο μονοπάτι από την κορυφή  $x$  στην κορυφή  $y$  (ένα μονοπάτι που ακολουθεί ακμές μόνο σε υποδεικνυόμενη κατεύθυνση);” και “Ποιες κορυφές μπορούμε να επισκεφτούμε από την κορυφή  $x$  μέσω ενός προσανατολισμένου μονοπατιού;” και “Υπάρχει μήπως κάποιο προσανατολισμένο μονοπάτι από την κορυφή  $x$  στην κορυφή  $y$  και κάποιο προσανατολισμένο μονοπάτι από την κορυφή  $y$  στην κορυφή  $x$ ;” ([2]).

### ➤ **Αναζήτηση κατά Πλάτος (Breadth-First Search)**

Υπάρχουν δύο αλγοριθμικές μέθοδοι *εξερεύνησης δικτύων* η αναζήτηση κατά πλάτος (Depth-First Search), η οποία αναλύθηκε και παραπάνω. Εδώ αναλύουμε τη δεύτερη, η οποία είναι η **Αναζήτηση κατά Πλάτος (Breadth-First Search)**. Όπως είδαμε και παραπάνω, η μέθοδος αναζήτησης κατά βάθος, μπορούμε να ορίσουμε ένα δάσος από τις ακμές, οι οποίες μας οδηγούν για πρώτη φορά σε κάθε άλλο κόμβο. Η μέθοδος αναζήτησης κατά πλάτος, αντιστοιχεί στην εξερεύνηση των δέντρων αυτού του δάσους ανά επίπεδο ([2]).

Και στους δύο αλγορίθμους, μπορούμε να αντιληφθούμε τις κορυφές σαν δέντρα ταξινομημένα σε τρεις κατηγορίες: *δέντρα-(ή επισκεπτόμενες)* κορυφές, αυτές οι οποίες έχουν παρθεί από τη δομή των δεδομένων; *περιφερειακές* κορυφές, αυτές οι οποίες γειτονεύουν με τις δέντρα-κορυφές, αλλά δεν έχουν ακόμα επισκεφθεί; και οι *αφανείς* κορυφές, αυτές οι οποίες δεν έχουν ανακύψει καθόλου ακόμα. Στην παρακάτω εικόνα παρουσιάζεται μία αναπαράσταση του αλγορίθμου αναζήτησης κατά πλάτος ([2]).



Εικόνα 37: Αναζήτηση κατά πλάτος ([21])

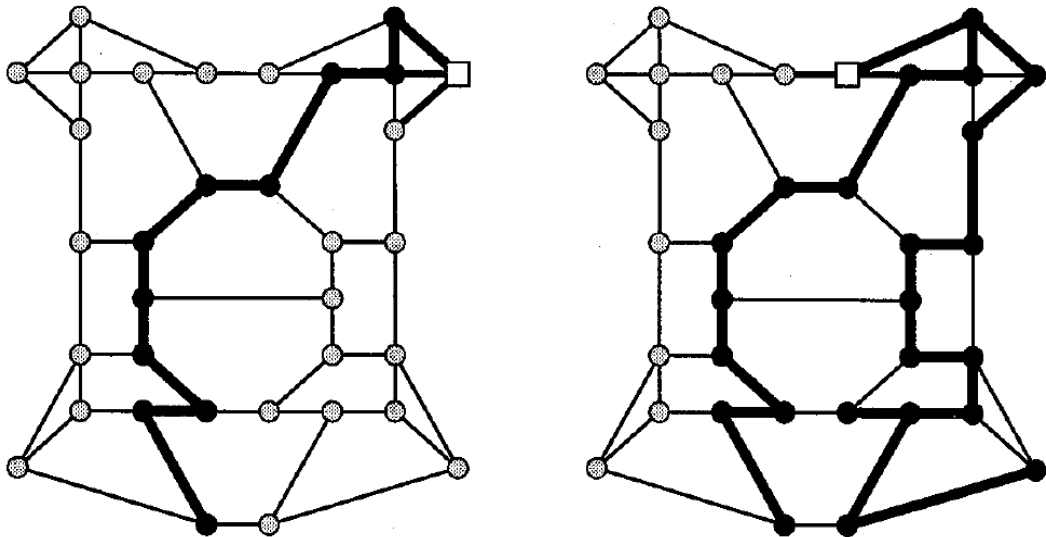
Όπως φαίνεται και στην παραπάνω εικόνα, η αναζήτηση του δικτύου ξεκινάει από τον κόμβο **A** και συνεχίζει προς τα κάτω, όπως δείχνει η κόκκινη γραμμή, η οποία επιδεικνύει την πορεία αναζήτησης. Έστω, λοιπόν, πως οι κόμβοι – στόχος είναι οι κόμβοι που είναι κυκλωμένοι με πράσινο χρώμα. Η μέθοδος αναζήτησης κατά πλάτος εξερευνά τους κόμβους πλησιέστερους στη ρίζα του δέντρου και έπειτα εξερευνά κόμβους μακριά. Για παράδειγμα, μετά τον κόμβο **A, B** και **C** η αναζήτηση συνεχίζει με τους κόμβους **D, E, F** και **G**. Στο συγκεκριμένο δίκτυο οι κόμβοι επισκέπτονται με τη σειρά **A, B, C, D, E, F, G, H, I, J, K, L, M, N, O, P, Q**. Σε αυτή την περίπτωση ο κόμβος **J** επισκέπτεται πριν από τον κόμβο **N**.

Για τη συστηματική εύρεση ενός συνδεδεμένου στοιχείου ενός δικτύου (εφαρμογή μίας διαδικασίας «επίσκεψης»), ξεκινούμε με μία κορυφή από το περιφέρεια, ενώ όλες οι άλλες ανήκουν ακόμα στην κατηγορία των αφανών κορυφών και μέχρι να «επισκεφτούμε» όλες τις πιθανές κορυφές, πραγματοποιούμε το επόμενο βήμα: “κούνα μία κορυφή (έστω  $x$ ) από την περιφέρεια στο δέντρο και βάζουμε όλες τις αφανείς κορυφές, γειτονικές στη  $x$  κορυφή στην περιφέρεια”. Οι μέθοδοι εξερεύνησης των δικτύων διαφέρουν στο πως θα αποφασιστεί ποια κορυφή πρέπει να κουνηθεί πρώτα από την περιφέρεια στο δέντρο. Στη μέθοδο αναζήτησης κατά βάθος επιθυμούμε να διαλέξουμε την κορυφή την πιο πρόσφατα εφαρμόσιμη από την περιφέρεια; αυτό αντιστοιχεί στήλης για την κράτηση των κορυφών στην περιφέρεια. Στη μέθοδο αναζήτησης κατά πλάτος θέλουμε να διαλέξουμε την κορυφή, η οποία ήταν πιο παλιά εφαρμόσιμη; αυτό αντιστοιχεί στη χρήση μίας σειράς για την κράτηση κορυφών στην περιφέρεια ([2]).

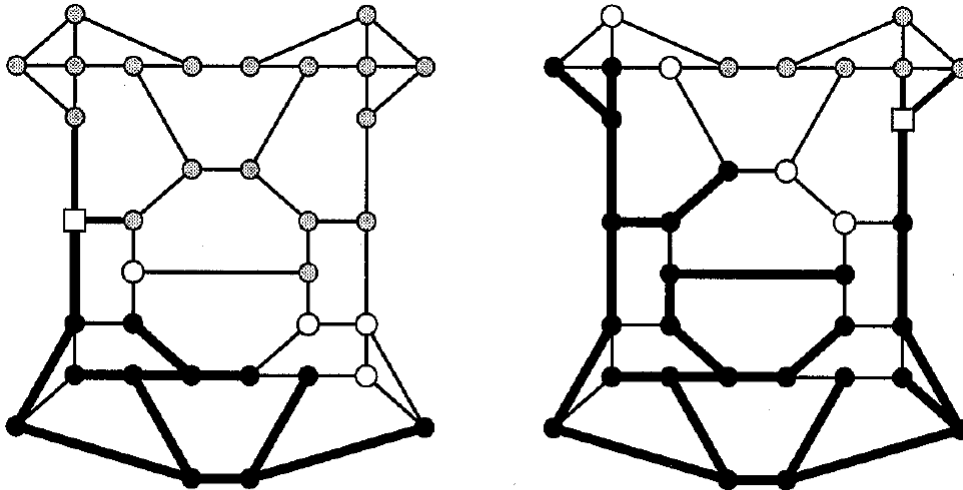
## Σύγκριση των δύο μεθόδων

Η σύγκριση μεταξύ των δύο μεθόδων είναι αρκετά προφανής, όταν εξετάζουμε ένα μεγαλύτερο δίκτυο. Και στις δύο περιπτώσεις, η αναζήτηση ξεκινάει από τον κόμβο κάτω αριστερά. Η κατά βάθος αναζήτηση διευθύνεται μέσω του δικτύου, αποθηκεύοντας σε σειρά τα σημεία στα οποία άλλα μονοπάτια διακλαδίζονται; Η κατά πλάτος αναζήτηση σαρώνει το δίκτυο, χρησιμοποιώντας μία σειρά για την ενθύμηση των συνόρων των επισκεπτόμενων τμημάτων. Η αναζήτηση κατά βάθος εξερευνά το δίκτυο ψάχνοντας για νέες κορυφές αρκετά μακριά από το σημείο έναρξης, λαμβάνοντας στενότερες κορυφές μόνο όταν εμφανίζονται αδιέξοδα; η αναζήτηση κατά πλάτος καλύπτει τελείως την περιοχή κοντά στο σημείο έναρξης, κινούμενη πιο μακριά μόνο όταν όλα τα κοντινά στοιχεία έχουν εξεταστεί. Αναφέρεται ξανά πως η σειρά με την οποία επισκέπτονται οι κόμβοι εξαρτάται αρκετά από τη σειρά με την οποία οι ακμές εμφανίζονται στην είσοδο και από τα αποτελέσματα αυτής της ταξινόμησης, από τη σειρά με την οποία οι κορυφές εμφανίζονται στις λίστες γειτνίασης ([2]).

Πέρα από αυτές τις λειτουργικές διαφορές, είναι ενδιαφέρον να προβληματιστούμε σχετικά με τις θεμελιώδεις διαφορές στην εφαρμογή αυτών των μεθόδων. Η αναζήτηση κατά βάθος είναι πολύ απλά εκφρασμένη αναδρομικά (επειδή η υποκείμενη δομή δεδομένων της είναι μία στήλη), ενώ η κατά πλάτος αναζήτηση επιδέχεται μία απλή μη-αναδρομική εφαρμογή (επειδή η υποκείμενη δομή δεδομένων της είναι μία σειρά ([2])).



Εικόνα 38: Αναζήτηση κατά βάθος σε ένα μεγαλύτερο δίκτυο ([2])

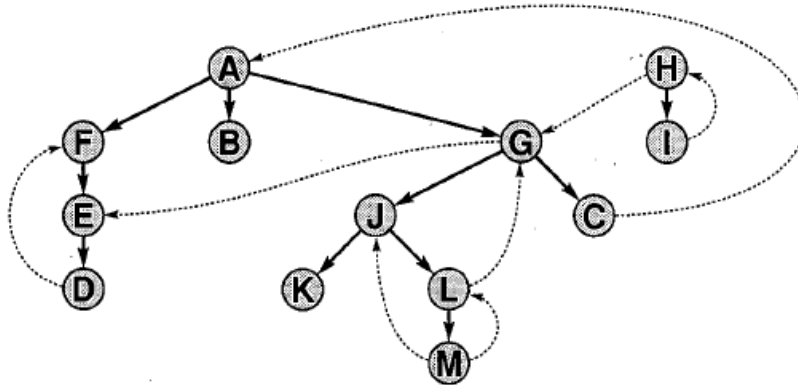


Εικόνα 39: Αναζήτηση κατά πλάτος σε ένα μεγαλύτερο δίκτυο ([2])

➤ **Μεταβατική Κλειστότητα (transitive closure)**

Στα μη-προσανατολισμένα δίκτυα, η απλή συνδεσιμότητα παρέχει τις κορυφές οι οποίες μπορούν να φταστούν από μία δεδομένη κορυφή διασχίζοντας ακμές από το δίκτυο: είναι όλες στο ίδιο συνδεδεμένο στοιχείο. Παρόμοια, στα προσανατολισμένα δίκτυα, ενδιαφερόμαστε, συχνά, για το σύνολο των κορυφών, οι οποίες μπορούν να φταστούν από μία δεδομένη κορυφή διασχίζοντας ακμές από το δίκτυο με την υποδεικνυόμενη φορά. Το πρόβλημα με τα προσανατολισμένα δίκτυα είναι κάπως πιο πολύπλοκο από την απλά συνδεσιμότητα ([2]).

Είναι εύκολο να αποδειχθεί πως η αναδρομική διαδικασία της «επίσκεψης» ενός κόμβου στη μέθοδο της αναζήτησης κατά βάθος, επισκέπτεται όλους τους κόμβους οι οποίοι μπορούν να προσπελαστούν από τον κόμβο έναρξης. Έτσι, εάν τροποποιήσουμε αυτή τη διαδικασία ώστε να τυπώνουμε τους κόμβους τους οποίους επισκέπτεται κάθε φορά ο αλγόριθμός μας, ουσιαστικά τυπώνουμε όλους τους κόμβους που μπορούν να προσπελαστούν από τον αρχικό κόμβο. Ας σημειώσουμε πως δεν είναι απαραίτητα αληθές το ότι κάθε δέντρο στο δάσος της αναζήτησης κατά βάθος περιέχει όλους τους κόμβους που μπορούν να προσπελαστούν από τις ρίζες του δέντρου. Παρακάτω παρουσιάζουμε ένα παράδειγμα, ως συνέχεια του δικτύου που παρουσιάστηκε στην εικόνα 29 για να γίνουν οι έννοιες που επεξηγούμε καλύτερα κατανοητές ([51]).



Εικόνα 40: Δάσος αναζήτησης κατά πλάτος ([2])

Στο παράδειγμά μας, λοιπόν, όλοι οι κόμβοι μπορούν να προσπελαστούν από τον κόμβο H, όχι μόνο από το I. Για να έχουμε όλους τους κόμβους οι οποίοι μπορούν να προσπελαστούν από τον κάθε κόμβο, απλά καλείται μία συνάρτηση «επίσκεψη»  $V$  φορές, μία για κάθε κόμβο ([2]):

```

for (k=1;k<=V;k++)
{
  id=0;

  for (j=1;j<=V;j++)  val[j]=0;

  visit (k);

  printf("\n");
}

```

Αυτό το πρόγραμμα επιστρέφει το ακόλουθο αποτέλεσμα για το προσανατολισμένο δίκτυο της εικόνας 29. Όπως και πριν, η σειρά των ονομάτων των κόμβων για κάθε γραμμή είναι ένα τεχνούργημα της συγκεκριμένης αναπαράστασης του δικτύου και της χρησιμοποιούμενης διαδικασίας αναζήτησης. Το σύνολο των κόμβων σε κάθε γραμμή είναι μία δομημένη ιδιότητα του ίδιου του δικτύου: κάθε γραμμή έχει τους κόμβους αυτούς προσπελάσιμους μέσω ενός προσανατολισμένου δικτύου από τον πρώτο κόμβο της γραμμής ([2]).

```

A F E D B G J K L M C
B
C A F E D B G J K L M
D F E
E D F
F E D
G J K L M C A F E D B
H G J K L M C A F E D B I
I H G J K L M C A F E D B
J K L G C A F E D B M
K
L G J K M C A F E D B
M L G J K C A F E D B

```

Εικόνα 41: Αποτέλεσμα κώδικα επίσκεψης ([2])

Για μη-προσανατολισμένα δίκτυα, αυτή ο υπολογισμός θα παρήγαγε έναν πίνακα με την ιδιότητα ότι κάθε γραμμή που αντιστοιχεί στους κόμβους σε ένα συνδεδεμένο στοιχείο, καταγράφει όλους τους κόμβους σε αυτό το στοιχείο. Ο πίνακας στην παραπάνω εικόνα έχει μία παρόμοια ιδιότητα: ορισμένες από τις γραμμές καταγράφει πανομοιότυπα σύνολα κόμβων. Παρακάτω θα αναλύσουμε τη γενίκευση της συνεκτικότητας που εξηγεί αυτή η ιδιότητα ([2]).

Μία λειτουργία η οποία θα θέλαμε να πραγματοποιήσουμε είναι να προσθέσουμε μία ακμή άμεσα από το  $x$  στο  $y$ , εάν υπάρχει κάποιος τρόπος να πάμε από το  $x$  στο  $y$ . Το δίκτυο το οποίο προκύπτει από την πρόσθεση όλων των ακμών αυτής της φύσεως σε ένα προσανατολισμένο δίκτυο αποκαλείται *μεταβατική κλειστότητα* του δικτύου. Κανονικά, ένας μεγάλος αριθμός ακμών θα προστεθεί και η μεταβατική κλειστότητα είναι πιθανόν να είναι αρκετά πυκνή, έτσι μία αναπαράσταση μητρώου γειτνίασης χρειάζεται. Αυτό είναι ανάλογο με τα συνδεδεμένα στοιχεία σε ένα μη-προσανατολισμένο δίκτυο; άπαξ και παρουσιάσουμε αυτόν τον υπολογισμό μία φορά, μπορούμε μετά γρήγορα να απαντήσουμε ερωτήσεις όπως “υπάρχει κάποιος δρόμος για να πάμε από τον κόμβο  $x$  στον κόμβο  $y$  ;” ([2]).

***Ιδιότητα:*** Η αναζήτηση κατά πλάτος μπορεί να χρησιμοποιηθεί για τον υπολογισμό της μεταβατικής κλειστότητας ενός προσανατολισμένου δικτύου σε βήματα της τάξεως του  $V(E+V)$  για τα αραιά δίκτυα και της τάξεως του  $V^3$  για πυκνά δίκτυα ([2]).

Παρακάτω παρουσιάζεται ένας σημαντικά απλός κώδικας για τον υπολογισμό της μεταβατικής κλειστότητας ενός δικτύου, το οποίο αντιπροσωπεύεται με μητρώο γειτνίασης ([2]):

```

for ( y = 1; y <= V; y++ )
  for ( x = 1; x <= V; x++ )
    if ( a [ x ][ y ] )

      for ( j = 1; j <= V ; j++ )

        if ( a [ y ][ j ] a [ x ][ j ] = 1 ;

```

Ο Stephen Warshall επινόησε αυτή τη μέθοδο το 1962 χρησιμοποιώντας την απλή παρατήρηση ότι “ εάν υπάρχει ένας τρόπος να πάμε από τον κόμβο  $x$  στον κόμβο  $y$  και ένας τρόπος να πάμε από τον κόμβο  $y$  στον κόμβο  $j$ , τότε υπάρχει σίγουρα ένας τρόπος να πάμε από τον κόμβο  $x$  στον κόμβο  $j$ .” Το σημαντικό είναι να κάνουμε αυτήν την παρατήρηση λίγο πιο εύστοχη, έτσι ώστε να γίνει ο υπολογισμός σε ένα μόνο πέρασμα μέσω του μητρώου: “ εάν υπάρχει ένας τρόπος να πάμε από τον κόμβο  $x$  στον κόμβο  $y$  χρησιμοποιώντας μόνο κόμβους με λιγότερους δείκτες από  $y$  και ένας τρόπος να πάμε από τον κόμβο  $y$  στον κόμβο  $j$ , τότε υπάρχει ένας τρόπος να πάμε από τον κόμβο  $x$  στον κόμβο  $j$  χρησιμοποιώντας μόνο κόμβους με λιγότερους δείκτες από  $y+1$ .” Ο παραπάνω κώδικας είναι μία άμεση εφαρμογή των παρακάτω ([2]).

	A	B	C	D	E	F	G	H	I	J	K	L	M
A	1	0	0	0	1	1	0	0	0	0	0	0	0
B	0	1	0	0	0	0	0	0	0	0	0	0	0
C	1	0	1	0	0	0	0	0	0	0	0	0	0
D	0	0	0	1	0	1	0	0	0	0	0	0	0
E	0	0	0	1	1	0	0	0	0	0	0	0	0
F	0	0	0	0	1	1	0	0	0	0	0	0	0
G	0	0	1	0	1	0	1	0	0	1	0	0	0
H	0	0	0	0	0	1	1	1	0	0	0	0	0
I	0	0	0	0	0	0	1	1	0	0	0	0	0
J	0	0	0	0	0	0	0	0	1	1	1	1	1
K	0	0	0	0	0	0	0	0	0	1	0	0	0
L	0	0	0	0	0	1	0	0	0	0	1	1	1
M	0	0	0	0	0	0	0	0	0	0	0	1	1

Εικόνα 42: Αρχικά στάδια της μεθόδου του Warshall ([2])

	A	B	C	D	E	F	G	H	I	J	K	L	M
A	1	1	1	1	1	1	0	0	1	1	1	1	1
B	0	1	0	0	0	0	0	0	0	0	0	0	0
C	1	1	1	1	1	1	0	0	1	1	1	1	1
D	0	0	0	1	1	1	0	0	0	0	0	0	0
E	0	0	0	1	1	1	0	0	0	0	0	0	0
F	0	0	0	1	1	1	0	0	0	0	0	0	0
G	1	1	1	1	1	1	0	0	1	1	1	1	1
H	1	1	1	1	1	1	1	1	1	1	1	1	1
I	1	1	1	1	1	1	1	1	1	1	1	1	1
J	0	0	0	0	0	0	0	0	1	1	1	1	1
K	0	0	0	0	0	0	0	0	0	1	0	0	0
L	1	1	1	1	1	1	0	0	1	1	1	1	1
M	0	0	0	0	0	0	0	0	0	1	1	1	1

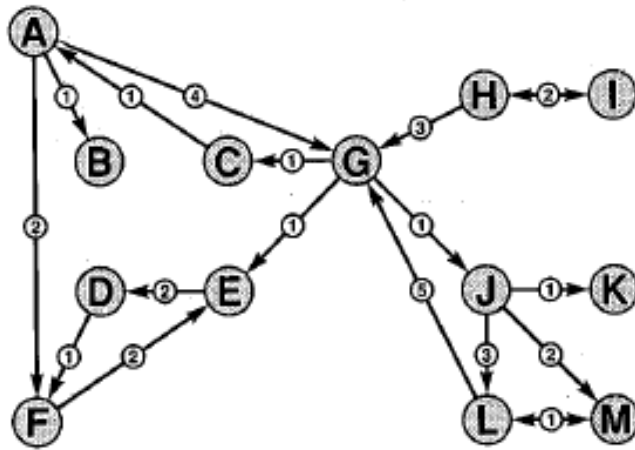
Εικόνα 43: Τελικά στάδια της μεθόδου του Warshall ([2])

Η μέθοδος του Warshall ουσιαστικά μετατρέπει το μητρώο γειτνίασης ενός δικτύου στο μητρώο γειτνίασης της μεταβατικής κλειστότητας του.

**Ιδιότητα:** Ο αλγόριθμος του Warshall βρίσκει τη μεταβατική κλειστότητα σε βήματα της τάξεως του  $V^3$  ([2]).

➤ **Το πρόβλημα της συντομότερης διαδρομής (Shortest path)**

Η μεταβατική κλειστότητα ενός δικτύου χωρίς βάρη (προσανατολισμένο ή όχι) απαντάει σε ερωτήσεις του τύπου “Υπάρχει κάποιο μονοπάτι από τον κόμβο  $x$  στον κόμβο  $y$  ;” για όλα τα ζευγάρια κορυφών  $x,y$ . Για τα δίκτυα με βάρη (προσανατολισμένα ή μη) είναι αναγκαία η δημιουργία ενός πίνακα που επιτρέπει την εύρεση όλων των σύντομων διαδρομών από τον κόμβο  $x$  στον κόμβο  $y$  για όλα τα ζευγάρια κορυφών. Αυτό είναι ουσιαστικά και το πρόβλημα όλων των σύντομων διαδρομών (*all-pairs shortest-path problem*). Για παράδειγμα στο δίκτυο με βάρη που φαίνεται στην παρακάτω εικόνα, θέλουμε να μπορούμε να λαμβάνουμε από τον πίνακα αυτό την πληροφορία πως η συντομότερη διαδρομή από τον κόμβο  $M$  στον κόμβο  $K$  έχει μήκος 8 και πως η συντομότερη διαδρομή από τον κόμβο  $J$  στον κόμβο  $F$  έχει μήκος 12 κτλ ([2]).



Εικόνα 44: Ένα προσανατολισμένο δίκτυο με βάρη ([2])

Όπως και παραπάνω, ο αλγόριθμος της σύντομης διαδρομής βρίσκει τη πιο σύντομη διαδρομή από τον κόμβο έναρξης σε κάθε άλλη κορυφή, έτσι ώστε να τρέξουμε αυτή τη διαδικασία μόνο  $V$  φορές, ξεκινώντας από κάθε κόμβο. Αυτό ουσιαστικά δίνει έναν τελικό αλγόριθμο, ο οποίος τρέχει σε βήματα της τάξης του  $(E+V)V \log V$ . Αλλά είναι επίσης δυνατό να χρησιμοποιηθεί και μία μέθοδος ακριβώς σαν τη μέθοδο του Warshall, η οποία γενικά οφείλεται στο Robert W (Bob) Floyd και παρουσιάζεται παρακάτω ([2]):

```

for ( y = 1; y <= V; y++ )
  for ( x = 1; x <= V; x++ )
    if ( a[x][y] )
      for ( j=1; j <= V; j++ )
        if ( a[y][j] > 0 )
          if ( !a[x][j] || a[x][y] + a[y][j] < a[x][j] )
            a[x][j] = a[x][y] + a[y][j];

```



Η δομή του αλγορίθμου είναι ακριβώς η ίδια με αυτή στη μέθοδο του Warshall. Αντί, όμως, να χρησιμοποιήσουμε τη διάζευξη για να παρακολουθήσουμε τα μονοπάτια, κάνουμε ένα μικρό υπολογισμό για κάθε ακμή ώστε να καθορίσουμε εάν είναι μέρος μίας καινούργιας σύντομης διαδρομής: “ η σύντομη διαδρομή από τον κόμβο  $x$  στον κόμβο  $j$  χρησιμοποιώντας μόνο κόμβους με λιγότερους δείκτες από  $y+1$  είναι είτε η πιο σύντομη διαδρομή από τον κόμβο  $x$  στον κόμβο  $j$  χρησιμοποιώντας μόνο κόμβους με λιγότερους από  $y$  δείκτες, είτε εάν είναι πιο σύντομη, η πιο σύντομη διαδρομή από τον κόμβο  $x$  στον κόμβο  $y$  συν την απόσταση από τον κόμβο  $y$  στον κόμβο  $j$ .” Ως συνήθως, μία μηδενική τιμή μητρώου, αντιστοιχεί στην απουσία της υποδεικνυόμενης ακμής ([2]).

**Ιδιότητα:** Ο αλγόριθμος του Floyd λύνει το πρόβλημα των σύντομων διαδρομών σε βήματα της τάξεως του  $V^3$  ([2]).

	A	B	C	D	E	F	G	H	I	J	K	L	M
A	0	1	0	0	0	2	4	0	0	0	0	0	0
B	0	0	0	0	0	0	0	0	0	0	0	0	0
C	1	0	0	0	0	0	0	0	0	0	0	0	0
D	0	0	0	0	0	1	0	0	0	0	0	0	0
E	0	0	0	2	0	0	0	0	0	0	0	0	0
F	0	0	0	2	0	0	0	0	0	0	0	0	0
G	0	0	1	0	1	0	0	0	0	1	0	0	0
H	0	0	0	0	0	3	0	1	0	0	0	0	0
I	0	0	0	0	0	0	0	1	0	0	0	0	0
J	0	0	0	0	0	0	0	0	0	1	3	2	
K	0	0	0	0	0	0	0	0	0	0	0	0	0
L	0	0	0	0	0	5	0	0	0	0	0	1	
M	0	0	0	0	0	0	0	0	0	0	0	1	0

Εικόνα 45: Αρχικά στάδια του αλγορίθμου του Floyd ([2])

	A	B	C	D	E	F	G	H	I	J	K	L	M
A	6	1	5	6	4	2	4	0	0	5	6	8	7
B	0	0	0	0	0	0	0	0	0	0	0	0	0
C	1	2	6	7	5	3	5	0	0	6	7	9	8
D	0	0	0	5	3	1	0	0	0	0	0	0	0
E	0	0	0	2	5	3	0	0	0	0	0	0	0
F	0	0	0	4	2	5	0	0	0	0	0	0	0
G	2	3	1	3	1	4	6	0	0	1	2	4	3
H	5	6	4	6	4	7	3	2	1	4	5	7	6
I	6	7	5	7	5	8	4	1	2	5	6	8	7
J	0	0	0	0	0	0	0	0	0	1	3	2	
K	0	0	0	0	0	0	0	0	0	0	0	0	0
L	7	8	6	8	6	9	5	0	0	6	7	9	1
M	0	0	0	0	0	0	0	0	0	0	1	0	

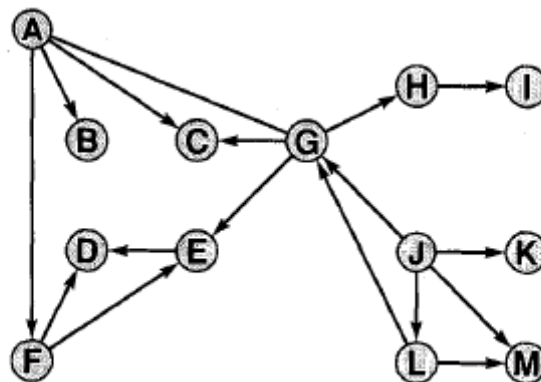
Εικόνα 46: Τελικά στάδια του αλγορίθμου του Floyd ([2])

Όπως και στην παραπάνω περίπτωση οι μηδενικές τιμές των μητρώων, που αντιστοιχούν στην απουσία ενός μονοπατιού ανάμεσα στις περιεχόμενες

κορυφές, είναι όμοιες και για τους δύο αλγόριθμους. Οι μη μηδενικές τιμές για τον αλγόριθμο του Warshall υποδεικνύουν την ύπαρξη ενός μονοπατιού μεταξύ των δύο κορυφών; για τον αλγόριθμο του Floyd δίνουν το μήκος της πιο σύντομης, μέχρι εκείνη τη στιγμή, διαδρομής. Η πραγματική πιο σύντομη διαδρομή μπορεί να υπολογισθεί, επίσης, χρησιμοποιώντας μία μορφή μητρώου του μητρώου που αναφέρθηκε και παραπάνω: θέτοντας την τιμή στη θέση  $(x, j)$  με το όνομα της προηγούμενης κορυφής κάθε φορά στο πιο σύντομο μονοπάτι από τον κόμβο  $x$  στον κόμβο  $j$  ([2]).

➤ **Τοπολογική ταξινόμηση (Topological Sorting)**

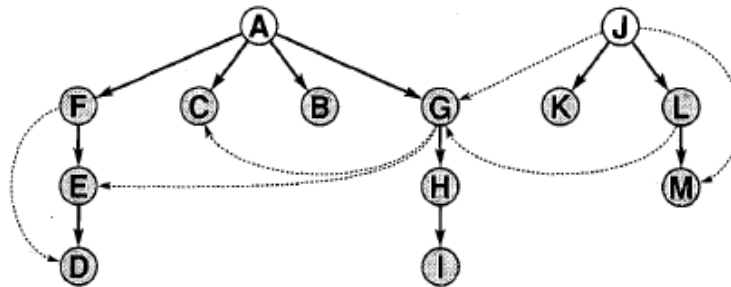
Τα κυκλικά δίκτυα (Cyclic Graphs) προκύπτουν σε πολλές εφαρμογές που αφορούν τα προσανατολισμένα δίκτυα. Εάν για παράδειγμα το δίκτυο της εικόνας 29 μοντελοποιούσε μία γραμμή παραγωγής, τότε αυτό θα υπονοούσε, ως πούμε, πως η εργασία A πρέπει να γίνει πριν από την εργασία G, η οποία πρέπει να γίνει πριν από την εργασία C, η οποία πρέπει με τη σειρά της να γίνει πριν από την εργασία A. Αλλά μία τέτοια κατάσταση είναι αδύνατη: για αυτή και άλλες πολλές εφαρμογές, χρειάζονται προσανατολισμένα δίκτυα με μη προσανατολισμένους κύκλους (no directed cycles) (κύκλοι με όλες τις ακμές τους να δείχνουν προς την ίδια κατεύθυνση). Αυτού του τύπου τα δίκτυα ονομάζονται προσανατολισμένα μη κυκλικά δίκτυα (Directed Acyclic Graphs - DAGs). Τα δίκτυα αυτά μπορούν να έχουν πολλούς κύκλους, εάν οι κατευθύνσεις των ακμών δεν λαμβάνονται υπόψη; η καθορισμένη τους ιδιότητα είναι απλά πως κανείς δεν μπορεί να μπει μέσα σε έναν κύκλο ακολουθώντας τις ακμές με την προκαθορισμένη κατεύθυνση. Η παρακάτω εικόνα παρουσιάζει ένα τέτοιο δίκτυο παρόμοιο με αυτό της εικόνας 29, με λίγες ακμές αφαιρούμενες ή φορές αλλαγμένες με σκοπό να αφαιρεθούν κύκλοι. Στην παρακάτω εικόνα παρουσιάζεται ένα τέτοιου τύπου δίκτυο ([2]).



Εικόνα 47: Ένα προσανατολισμένο μη-κυκλικό δίκτυο (dag) ([2])

Τα DAGs είναι αλήθεια κάπως διαφορετικά από τα γενικά προσανατολισμένα δίκτυα: κατά μία έννοια, είναι μισά δέντρο, μισά δίκτυα. Μπορούμε σίγουρα να ωφεληθούμε από την ειδική τους δομή, όταν τα επεξεργαζόμαστε. Εάν τα

δούμε από οποιαδήποτε κορυφή, ένα dag μοιάζει με ένα δέντρο; και για να ειπωθεί και διαφορετικά, το δάσος αναζήτησης κατά πλάτος για τα dag δεν έχει πάνω ακμές. Στην παρακάτω εικόνα παρουσιάζεται το δάσος αναζήτησης κατά πλάτος, το οποίο περιγράφει τη λειτουργία της μεθόδου αναζήτησης κατά πλάτος στο dag της εικόνας 41 ([2]).



Εικόνα 48: Αναζήτηση κατά πλάτος σε ένα dag ([2])

Μία θεμελιώδης λειτουργία στα dags είναι η επεξεργασία των κορυφών του δικτύου με τέτοια σειρά με την οποία καμία κορυφή δεν επεξεργάζεται πριν από οποιαδήποτε κορυφή η οποία δείχνει σε αυτή. Για παράδειγμα, οι κορυφές στο παραπάνω δίκτυο μπορούν να προσπελαστούν με την ακόλουθη σειρά:

J K L M A G H I F E D B C

Εάν οι ακμές είχαν σχεδιαστεί με τις κορυφές σε αυτές τις θέσεις, θα πήγαιναν όλοι από τα αριστερά προς τα δεξιά. Όπως αναφέρθηκε και παραπάνω, αυτό το χαρακτηριστικό έχει προφανής εφαρμογή. Για παράδειγμα, στα δίκτυα τα οποία παριστάνουν παραγωγικές διαδικασίες, δείχνει ένα συγκεκριμένο δρόμο για να επεξεργασθούν οι περιορισμοί που αναπαρίστανται από το δίκτυο. Αυτή η λειτουργία ονομάζεται *τοπολογική ταξινόμηση (topological sorting)*, επειδή περιλαμβάνει την ταξινόμηση των κορυφών του δικτύου.

Γενικά, η σειρά των κορυφών που παράγεται από την τοπολογική ταξινόμηση δεν είναι μοναδική. Για παράδειγμα, η σειρά:

A J G F K L E M B H C I D

είναι μία κανονική τοπολογική ταξινόμηση του παραδείγματος μας (και υπάρχουν και πολλές άλλες). Στην αναφερόμενη εφαρμογή της παραγωγικής διαδικασίας, αυτή η κατάσταση ανακύπτει όταν μία δουλειά δεν έχει άμεση ή έμμεση εξάρτηση η μία στην άλλη και έτσι, μπορούν παρουσιαστούν με οποιαδήποτε σειρά.

Είναι ευκαιριακά χρήσιμο να ερμηνεύσουμε τις ακμές σε ένα δίκτυο με τον άλλο τρόπο: λέγοντας πως μία ακμή προσανατολισμένη από τον κόμβο x

στον κόμβο  $y$  σημαίνει πως η κορυφή  $x$  εξαρτάται από την κορυφή  $y$ . Για παράδειγμα, οι κορυφές μπορούν να αποτελούν όρους για να καθορισθούν με μία γλώσσα προγραμματισμού με μία ακμή από τον κόμβο  $x$  στον κόμβο  $y$ , εάν ο ορισμός του  $x$  χρησιμοποιεί το  $y$ . Σε αυτή την περίπτωση, θα ήταν χρήσιμο να ευρεθεί μία σειρά με την ιδιότητα ότι κάθε όρος ορίζεται πριν από χρησιμοποιηθεί σε άλλο ορισμό. Αυτό αντιστοιχεί στην τοποθέτηση των κορυφών σε μία γραμμή, έτσι ώστε όλες οι ακμές να πηγαινούν από τα δεξιά από τα αριστερά. Μία *αντίστροφη τοπολογική ταξινόμηση (reverse topological sorting)* για το δίκτυο μας είναι:

D E F C B I H G A K M L J

Ο διαχωρισμός αυτός δεν είναι ζωτικής σημασίας: η εκτέλεση μίας αντίστροφης τοπολογικής ταξινόμησης είναι σε ένα δίκτυο είναι ισοδύναμη με την εκτέλεση της τοπολογικής ταξινόμησης του δικτύου, το οποίο λαμβάνεται από την αντιστροφή όλων των ακμών ([2]).

### ➤ **Ισχυρά συνεκτικές συνιστώσες (Strongly Connected Components)**

Εάν ένα δίκτυο περιλαμβάνει έναν *προσανατολισμένο κύκλο* (εάν πάμε από τον έναν κόμβο πίσω στον ίδιο ακολουθώντας τις ακμές στην υποδεικνυόμενη κατεύθυνση), τότε δεν είναι ένα dag και δεν μπορεί να ταξινομηθεί τοπολογικά: οποιαδήποτε κορυφή του κύκλου τυπωθεί πρώτα θα υπάρξει άλλη μία κορυφή η οποία θα δείχνει σε αυτή και δεν θα έχει, όμως, τυπωθεί ακόμα. Οι κόμβοι στον κύκλο είναι αμοιβαία προσβιβάσιμοι με την έννοια του ότι υπάρχει ένα τρόπος να πάμε από κάθε κόμβο του κύκλου σε οποιαδήποτε άλλο κόμβο του κύκλου και πίσω. Από την άλλη μεριά, μολονότι ένα δίκτυο μπορεί να είναι συνδεδεμένο, είναι απίθανο οποιοσδήποτε κόμβος να προσπελασθεί από οποιαδήποτε άλλον μέσω ενός προσανατολισμένου μονοπατιού. Στην πραγματικότητα, οι κόμβοι διαχωρίζονται σε σύνολα, τα οποία ονομάζονται *ισχυρά συνεκτικές συνιστώσες (Strongly Connected Components)* με την ιδιότητα ότι όλοι οι κόμβοι μέσα σε μία συνιστώσα είναι αμοιβαία προσβιβάσιμοι, αλλά δεν υπάρχει τρόπος να πάμε από τον ένα κόμβο της μίας συνιστώσας σε έναν κόμβο μίας άλλης συνιστώσας και πίσω ([2]).

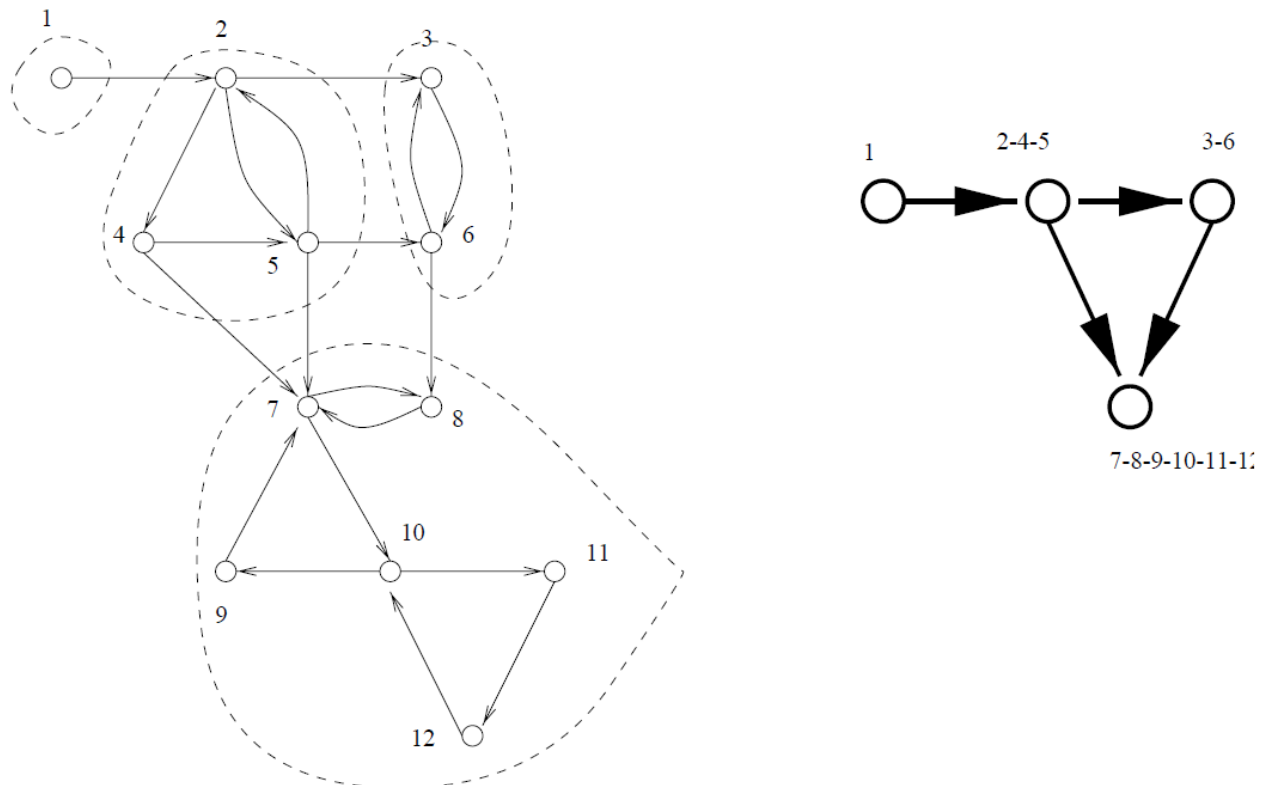
Οι ισχυρά συνεκτικές συνιστώσες του προσανατολισμένου δικτύου της εικόνας 29 είναι δύο μόνοι κόμβοι, ο B και ο K, ένα ζευγάρι κόμβων H I, μία τριπλέτα κόμβων D E F και μία μεγάλη συνιστώσα με έξι κόμβους A C G J L M. Για παράδειγμα, η κορυφή A είναι σε μία διαφορετική συνιστώσα από την κορυφή F επειδή παρόλο που υπάρχει ένα μονοπάτι από τον κόμβο A στον κόμβο F, δεν υπάρχει κανένας τρόπος να πάμε από τον κόμβο F στον κόμβο A.

Οι ισχυρά συνεκτικές συνιστώσες ενός προσανατολισμένου δικτύου μπορούν να βρεθούν με τη χρήση διαδοχικών αναζητήσεων κατά πλάτος. Η μέθοδος που θα εξετασθεί επινοήθηκε από τον Robert Endre Tarjan το 1972. Μιας και είναι βασισμένη στην αναζήτηση κατά πλάτος, τρέχει σε χρόνο ανάλογο της

ποσότητας  $V+E$ , αλλά είναι πραγματικά μία πολύ έξυπνη μέθοδος. Απαιτεί μόνο λίγες απλές τροποποιήσεις της βασικής συνάρτησης «επίσκεψης» κόμβων (έχουμε αναφερθεί και παραπάνω σε αυτή), αλλά πριν ο Tarjan να την παρουσιάσει, κανένας μη γραμμικού χρόνου αλγόριθμος δεν ήταν γνωστός για αυτό το πρόβλημα, παρόλο που αρκετοί άνθρωποι είχαν ασχοληθεί με το θέμα ([2]).

**Ιδιότητα:** Οι ισχυρά συνεκτικές συνιστώσες ενός δικτύου μπορούν να βρεθούν σε γραμμικό χρόνο ([2]).

Παρακάτω παρουσιάζεται ένα προσανατολισμένο δίκτυο και οι ισχυρά συνεκτικές συνιστώσες του.



Εικόνα 49: Προσανατολισμένο δίκτυο και οι ισχυρά συνεκτικές συνιστώσες του ([22])

### **2.5.5 Προσέγγιση κατά Boole για τη μοντελοποίηση σηματοδότησης δικτύων**

Η μαθηματική μοντελοποίηση των βιολογικών διαδικασιών παρέχει βαθιές γνώσεις για σύνθετα κυτταρικά συστήματα. Μολονότι τα ποσοτικά και συνεχή μοντέλα όπως είναι οι διαφορικές εξισώσεις είναι ευρέως χρησιμοποιούμενες, η χρήση τους εμποδίζεται σε συστήματα, όπου η γνώση των μηχανιστικών στοιχείων και των κινητικών παραμέτρων είναι σπάνια. Από την άλλη πλευρά, ένας πλούτος δεδομένων μοριακού και ποιοτικού επιπέδου πάνω σε μεμονωμένα συστατικά και αντιδράσεις, μπορούν να ληφθούν από την πειραματική λογοτεχνία και τις τεχνολογίες υψηλής απόδοσης, κάτι το οποίο κάνει τις ποιοτικές προσεγγίσεις, όπως είναι η μοντελοποίηση κατά Boole, ιδιαίτερα χρήσιμες. Παρακάτω θα προσπαθήσουμε να επιδείξουμε πως η μοντελοποίηση κατά Boole μπορεί να εφαρμοσθεί για την παρουσίαση της δομικής ανάλυσης των κυτταρικών δικτύων ([20]).

Η συστημική βιολογία στοχεύει στο να διαφωτίσει στο πως οι πολύπλοκες συμπεριφορές των βιολογικών συστημάτων αναδύονται από τις ιδιότητες των στοιχείων και των αντιδράσεων του συστήματος. Χρησιμοποιεί ένα συνδυασμό πειραματικών τεχνικών και υπολογιστικών προσεγγίσεων για το κέρδος γενικών γνώσεων σχετικά με τα βιολογικά συστήματα. Οι πειραματικές τεχνικές που χρησιμοποιούνται στα βιολογικά συστήματα τείνουν να έχουν δυνατότητες υψηλής απόδοσης και είναι έτσι, ικανά να καθορίζουν την αφθονία ή τη δραστηριότητα πολλών συνιστωσών την ίδια στιγμή. Για παράδειγμα, η αφθονία της μεταγραφής του mRNA χιλιάδων γονιδίων μπορεί να μετρηθεί με μικρό-πίνακες ή ακολουθίες RNA. Ποσοτικές συγκεντρώσεις πρωτεϊνών και μετά-μεταφραστικές τροποποιήσεις μπορούν να καθοριστούν με πρωτεομικές και φωσφοπρωτομικές μελέτες διεξαγόμενες μέσω φασματομετρία μάζας ([20]).

Υπάρχουν, επίσης, πειράματα υψηλής απόδοσης τα οποία μπορούν να ανιχνεύσουν τις αντιδράσεις μεταξύ των στοιχείων των συστημάτων, όπως είναι οι αντιδράσεις πρωτεϊνών, μεταγραφικές ρυθμίσεις (αντιδράσεις μεταξύ πρωτεϊνών και γενετικού υλικού) και γενετικές αντιδράσεις. Επιπροσθέτως, με τις χημικές δοκιμές υψηλής απόδοσης, τα πειράματα μικρής κλίμακας τα οποία μελετούν λιγότερα στοιχεία και αντιδράσεις, που συμπεριλαμβάνονται σε συγκεκριμένες βιολογικές διαδικασίες, οι οποίες παρέχουν υψηλής ποιότητας και αξιόπιστες γνώσεις για τα βιολογικά συστήματα ([20]).

Τα πειραματικά δεδομένα από τις τεχνολογίες υψηλής απόδοσης και τις μελέτες μικρής κλίμακας παρέχουν μία πλούσια πηγή κατανόησης των επιπέδων μηχανισμών των βιολογικών διαδικασιών. Παρόλα αυτά, ένα συμπλήρωμα των υπολογιστικών προσεγγίσεων μοντελοποίησης είναι απαραίτητο για την ανάκτηση των μηχανιστικών γνώσεων από τα δεδομένα και παράγουν ελέγξιμες υποθέσεις. Οι υπολογιστικές μέθοδοι που έχουν χρησιμοποιηθεί στη συστημική βιολογία, μπορούν να ταξινομηθούν σε top-down μεθόδους και σε bottom-up μεθόδους. Οι top-down μέθοδοι όπως είναι οι στατιστικές αναλύσεις και τα στατικά μοντέλα δικτύων, εφαρμόζονται σε πρωτεομικά δεδομένα υψηλής απόδοσης και στοχεύουν στην αποκρυπτογράφηση της οργάνωσης των υποκείμενων συστημάτων και στην

εκλογή συγκεκριμένων πληροφοριών για τις βιολογικές διαδικασίες. Οι μέθοδοι αυτής της τάξης δεν απαιτούν κινετικές παραμέτρους και μπορούν να εφαρμοσθούν στην ανάλυση γονιδιακών δεδομένων με χιλιάδες στοιχεία ή αντιδράσεις να λαμβάνουν γνώσεις για τα βιολογικά συστήματα. Οι bottom-up μέθοδοι μοντελοποιούν το πώς αντιδρούν στοιχεία, όπως τα γονίδια, οι πρωτεΐνες και οι μεταβολίτες, επιτυγχάνουν δυναμικές συμπεριφορές κυτταρικών συστημάτων. Αυτή η τάξη μεθόδων συνήθως ξεκινάει με υποθέσεις βιολογικών μηχανισμών που παράγονται από μεμονωμένα πειράματα μικρής κλίμακας. Η συνεχής δυναμική μοντελοποίησης, η πιο ευρέως χρησιμοποιούμενη bottom-up μέθοδος, απαιτεί επαρκείς μηχανιστικές ιδιότητες και κινετικές παραμέτρους, όπως οι τιμές σύνθεσης και αποσύνθεσης, κάνοντάς το πρακτικό για συστήματα με μόνο δέκα στοιχεία ή λιγότερα. Η διακριτή δυναμική μοντελοποίηση, όπως τα μοντέλα δικτύων κατά Boole, τα λογικά μοντέλα με πολλές τιμές και τα δίκτυα Petri, δεν απαιτεί κινετικές παραμέτρους και δύναται να παρέχει ποσοτικές δυναμικές περιγραφές της συμπεριφοράς των συστημάτων. Αυτές οι προσεγγίσεις μπορούν να εφαρμοσθούν για συστήματα με εκατοντάδες στοιχεία και έχουν χρησιμοποιηθεί ευρέως στη μοντελοποίηση των βιολογικών δικτύων ([27]).

Τα μοντέλα δικτύων κατά Boole, τα οποία είχαν προταθεί αρχικά σαν πρωτότυπα μοντέλα των συνήθων γονιδιακών δικτύων, είναι μία ειδική περίπτωση διακριτών δυναμικών μοντέλων. Ένα δίκτυο κατά Boole αποτελείται από ένα σύνολο κόμβων, των οποίων η κατάσταση είναι δυαδική και καθορισμένη από άλλους κόμβους στο δίκτυο, μέσω συναρτήσεων Boole. Από άποψη πολυπλοκότητας, τα δίκτυα κατά Boole βρίσκονται μεταξύ των στατικών μοντέλων δικτύων και των συνεχών δυναμικών μοντέλων, κάτι το οποίο τα καθιστά μία προσιτή και δυνατή προσέγγιση για τη μοντελοποίηση μεγάλης κλίμακας βιολογικών συστημάτων. Μετά τη συναρμολόγηση επιμέρους στοιχείων και ρυθμιστικών αλληλεπιδράσεων που συμπεριλαμβάνονται σε ένα σύστημα σε μία συνεκτική αναπαράσταση του δικτύου. Τα μοντέλα κατά Boole μπορούν να χρησιμοποιηθούν για την χρονική ποιοτική συμπεριφορά του συστήματος και για την κατανόηση του πώς οι διαταραχές ίσως να τροποποιήσουν τη συμπεριφορά του. Επίσης, οδηγούν σε προβλέψιμες-ελέγξιμες υποθέσεις, οι οποίες είναι ιδιαίτερα πολύτιμα σε συστήματα υψηλής κλίμακας ελαχίστως κατανοητέ. Τα δίκτυα κατά Boole έχουν εφαρμοσθεί επιτυχώς στη μοντελοποίηση πολλών ρυθμιστικών γονιδίων και δικτύων σηματοδότησης μίας ποικιλίας οργανισμών. Επιπλέον, τα δίκτυα κατά Boole έχουν χρησιμοποιηθεί σαν μοντέλα της αντίστροφης μηχανικής των βιολογικών δικτύων, ώστε να παραχθούν ρυθμιστικές αντιδράσεις και μονοπάτια σηματοδότησης γονιδιακών εκφράσεων ή προτεομικών δεδομένων ([27]).

Παρακάτω παρουσιάζεται μία επισκόπηση της μεθοδολογίας της μοντελοποίησης κατά Boole στη συστημική βιολογία. Θα περιγραφούν τα βασικά βήματα της δυναμικής μοντελοποίησης των κυτταρικών δικτύων και των υλοποιήσεών τους ([20]).

## Δυναμική μοντελοποίηση των κυτταρικών δικτύων κατά Boole

Μία Boolean μεταβλητή μπορεί να πάρει μόνο **δύο τιμές (0 ή 1)**, οι οποίες αντιστοιχούν στις λογικές τιμές **ψέμα** ή **αλήθεια**. Η αναπαράσταση της κατάστασης μίας βιολογικής ποσότητας από μία δυαδική μεταβλητή, με τις δύο καταστάσεις συνήθως να αναφέρονται σαν **OFF** ή **ON**. Μία Boolean συνάρτηση με k μεταβλητές είναι μία χαρτογράφηση  $B: \{0,1\}^k \rightarrow \{0,1\}$  από το σύνολο όλων των k-απλών πάνω από  $\{0,1\}$  σε μία δυαδική έξοδο. Αυτή η συνάρτηση περιγράφει τον καθορισμό μίας Boolean τιμής-εξόδου βασισμένη σε κάποιες λογικές πράξεις από k δυαδικές εισόδους. Οι βασικές λογικές πράξεις συμπεριλαμβάνουν τις **πράξεις AND, OR και NOT**. Για παράδειγμα, η παράσταση ([26]):

$$D = (A \text{ OR } B) \text{ AND NOT } C$$

είναι μία συνάρτηση κατά Boole με τρεις μεταβλητές. Μία Boolean συνάρτηση μπορεί επίσης να αντιπροσωπευθεί από ένα «μητρώο αληθείας», όπου κάθε γραμμή αντιστοιχεί σε ένα συνδυασμό τιμών Boolean μεταβλητών και στις συναφείς τιμές εξόδου των. Το μητρώο αληθείας μίας Boolean συνάρτησης με k μεταβλητές έχει  $2^k$  σειρές και k+1 στήλες ([26]).

Ένα μοντέλο δικτύου κατά Boole αποτελείται από ένα σύνολο Boolean μεταβλητών  $\{\sigma_1, \sigma_2, \dots, \sigma_n\}$ , των οποίων η τιμή καθορίζεται από άλλες μεταβλητές στο δίκτυο μέσω ενός συνόλου Boolean συναρτήσεων  $B = \{B_1, B_2, \dots, B_n\}$ , κάθε μία αντιστοιχούσα σε κάθε μεταβλητή. Σε ένα δυναμικό μοντέλο κατά Boole, η τιμή της κάθε μεταβλητής  $\sigma_i$  καθορίζεται από τις τρέχουσες ή τις προηγούμενες τιμές των ρυθμιστών τους (είσοδοι), ανάλογα με τα τρέχοντα χρησιμοποιούμενα συστήματα στο μοντέλο. Το σύγχρονο σύστημα ενημερώνει όλες τις μεταβλητές του μοντέλου ταυτόχρονα, δηλαδή η τιμή κάθε μεταβλητής η τιμή κάθε μεταβλητής τη χρονική στιγμή t +1, καθορίζεται από τις  $k_i$  εισόδους τη στιγμή t ([26]):

$$\sigma_i^{t+1} = B_i(\sigma_{i_1}^t, \sigma_{i_2}^t, \dots, \sigma_{i_{k_i}}^t)$$

Αυτή η λειτουργία ενημέρωσης είναι ντετερμινιστική. Στα ασύγχρονα συστήματα, οι μεταβλητές ενημερώνονται με έναν μη-σύγχρονο τρόπο ([26]):

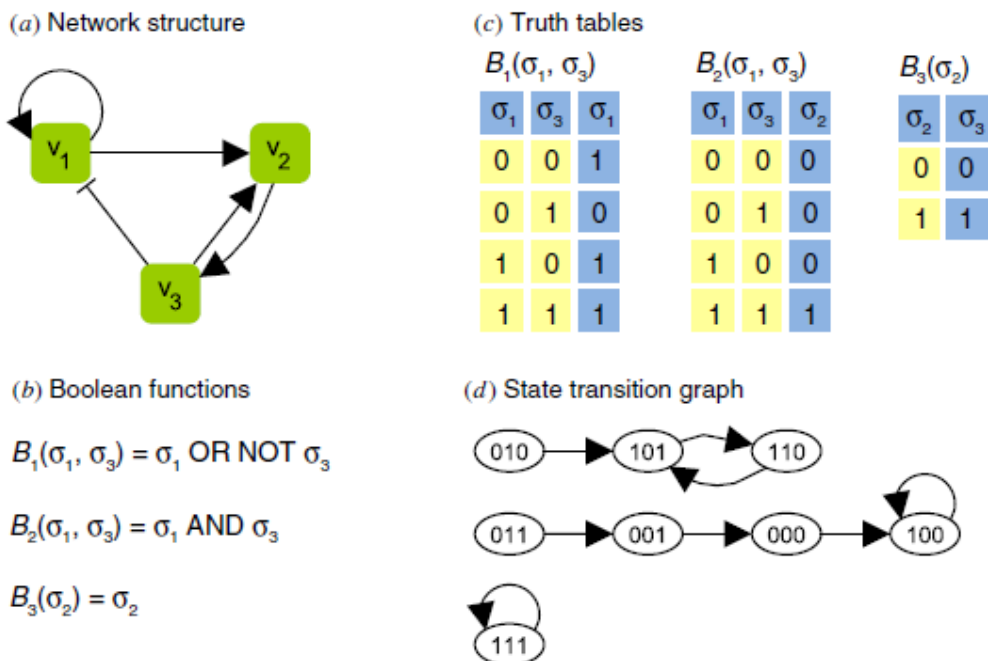
$$\sigma_i^* = B_i(\sigma_{i_1}, \sigma_{i_2}, \dots, \sigma_{i_{k_i}})$$

Όπου ο αστερίσκος υποδηλώνει τη νέα τιμή της μεταβλητής  $\sigma_i$  με  $i = 1, 2, \dots, n$ . Οι τιμές των εισόδων στη δεξιά πλευρά της εξίσωσης μπορεί να είναι οι τρέχοντες ή οι προγενέστερες, εξαρτώμενες από τα μεμονωμένα χρονοδιαγράμματα. Όπως θα παρουσιασθεί και παρακάτω, και τα ντετερμινιστικά και τα στοχαστικά ασύγχρονα συστήματα έχουν χρησιμοποιηθεί για τη μοντελοποίηση των βιολογικών συστημάτων ([26]).



Ένα μοντέλο δικτύου κατά Boole μπορεί να προβληθεί σε ένα προσανατολισμένο δίκτυο  $G(V,E)$ , όπου το σύνολο των κόμβων  $V = \{v_1, v_2, \dots, v_n\}$  αντιστοιχεί σε Boolean μεταβλητές και το σύνολο των ακμών  $E$  ορίζεται έμμεσα από τις Boolean συναρτήσεις στο μοντέλο. Κάθε ακμή έχει ένα σημάδι που υποδηλώνει εάν ο κόμβος εισόδου έχει θετικό ή αρνητικό αποτέλεσμα στον κόμβο – στόχο. Ας σημειωθεί πως τα μοντέλα κατά Boole περιέχουν πληροφορίες επιπρόσθετες στο διάγραμμα του προσανατολισμένου και σηματοδοτημένου δικτύου, μιας και το ίδιο το διάγραμμα αντιστοιχεί σε πολλές εναλλακτικές Boolean συναρτήσεις και το μοντέλο κατά Boole προσθέτει ένα δυναμικό στρώμα στο δίκτυο με τη μορφή μίας μεταβλητής κατάστασης  $\sigma_i(t)$  για κάθε κόμβο  $v_i$ . Η κατάσταση του συστήματος τη στιγμή  $t$  μπορεί να αναπαρασταθεί από ένα διάνυσμα  $(\sigma_1(t), \sigma_2(t), \dots, \sigma_n(t))$  με το  $i$ -οστό στοιχείο να αναπαριστά την κατάσταση του κόμβου  $v_i$  τη στιγμή  $t$ . Όλες οι πιθανές καταστάσεις του συστήματος, συνολικά  $2^n$ , συνθέτουν το χώρο κατάστασης. Οι πιθανές τροχιές στο χώρο κατάστασης μπορεί να αντιπροσωπευθεί από μία μεταβατική κατάσταση δικτύου, όπου οι κόμβοι είναι καταστάσεις του συστήματος και οι ακμές αντιπροσωπεύουν τις επιτρεπόμενες μεταβάσεις μεταξύ των καταστάσεων. Ενημερώνοντας τις καταστάσεις των κόμβων σε κάθε χρονικό βήμα, το μοντέλο εξελίσσεται κατά τη διάρκεια του χρόνου ακολουθώντας μία τροχιά καταστάσεων και τελικώς φτάνει σε μία σταθερή κατάσταση (σταθερό σημείο) ή ένα σύνολο επαναλαμβανόμενων καταστάσεων. Αυτές οι σταθερές ή επαναλαμβανόμενες καταστάσεις, αναφέρονται συλλογικά σαν ελκυστές. Το σύνολο των αρχικών καταστάσεων που οδηγεί το μοντέλο σε ένα συγκεκριμένο ελκυστή αποκαλείται η λεκάνη της έλξης αυτού του ελκυστή ([26]).

Στην παρακάτω εικόνα, παρουσιάζεται ένα απλό δίκτυο κατά Boole και οι συναφείς συναρτήσεις κατά Boole, τα μητρώα αληθείας και οι καταστάσεις μετάβασής του, βασισμένες στα σύγχρονα συστήματα ([26]).

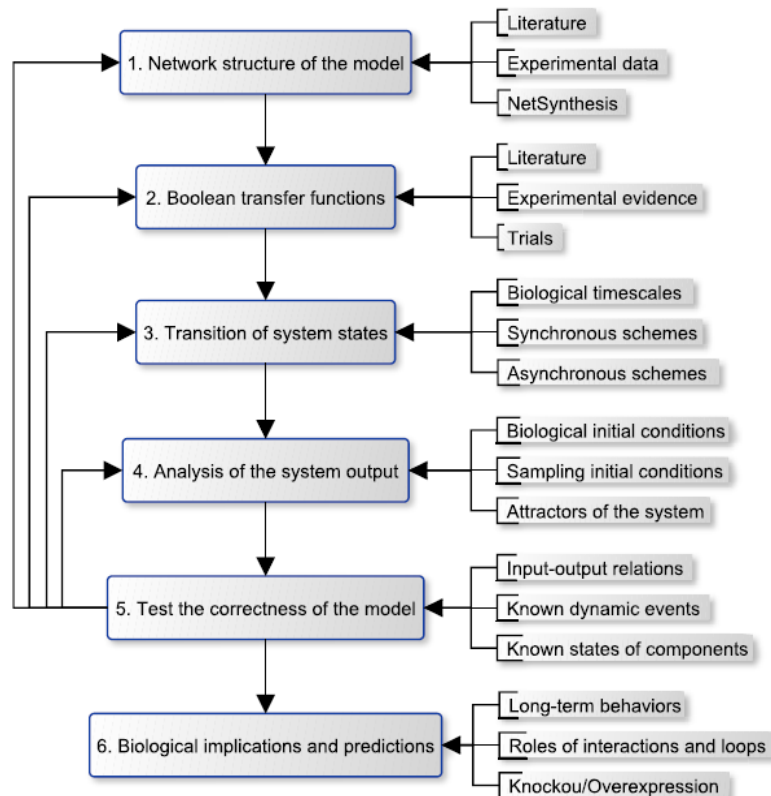


Εικόνα 50: Ένα απλό μοντέλο δικτύου κατά Boole ([26])

(a) Το προσανατολισμένο δίκτυο (δομή δικτύου) αντιστοιχεί με το μοντέλο του Boole. Οι ακμές με τα αιχμηρά βέλη εκπροσωπούν τα θετικά αποτελέσματα και οι ακμές με τα αμβλύ βέλη υποδεικνύουν τα αρνητικά αποτελέσματα. Ας σημειωθεί πως το δίκτυο δεν καθορίζει μοναδικώς τις συναρτήσεις κατά Boole για τους κόμβους  $v_1$  και  $v_2$ . (b) Οι συναρτήσεις κατά Boole στο μοντέλο. Ας σημειωθεί πως το δίκτυο στην εικόνα (a) μπορεί να υποστηρίξει τις εναλλακτικές Boolean συναρτήσεις, ειδικά  $B_1(\sigma_1, \sigma_3) = \sigma_1 \text{ AND NOT } \sigma_3$ ,  $B_2(\sigma_1, \sigma_3) = \sigma_1 \text{ OR } \sigma_3$ . (c) Τα μητρώα αληθείας των συναρτήσεων κατά Boole δίνονται στην εικόνα (b). (d) Η μεταβατική κατάσταση δικτύου του Boolean μοντέλου, το οποίο κατασκευάζεται χρησιμοποιώντας το σύγχρονο σύστημα. Οι καταστάσεις 100 και 111 είναι τα σταθερά σημεία του συστήματος, και οι καταστάσεις 101 και 110 διαμορφώνουν έναν περιορισμένο κύκλο ([26]).

Τα Boolean δίκτυα, σαν μία ειδική περίπτωση της διακριτής δυναμικής μοντελοποίησης, παρέχουν έναν αποτελεσματικό φορμαλισμό για την περιγραφή της δυναμικής των βιολογικών συστημάτων. Η προβολή των σηματοδοτημένων και προσανατολισμένων δικτύων ενός Boolean μοντέλου μπορεί να συνδέονται άμεσα με το διάγραμμα μονοπατιών ενός ρυθμιστικού βιολογικού συστήματος. Κάθε κόμβος  $v_i$  στο Boolean δίκτυο αντιπροσωπεύει ένα βιολογικό στοιχείο, όπως είναι ένα γονίδιο, μία πρωτεΐνη, ένας μεταβολίτης, ένα κανάλι ιόντων ή ένα ερέθισμα (σήμα), το οποίο αντιστοιχεί σε μία δυαδική κατάσταση (επίπεδο συγκέντρωση, δραστηριότητας και έκφρασης)  $\sigma_i$ . Η κατάσταση  $\sigma_i = 1$  (ON) παριστάνει ότι το στοιχείο  $v_i$  ενεργοποιείται ή εκφράζεται ή έχει κάποια παραπάνω όρια συγκέντρωσης και η κατάσταση  $\sigma_i = 0$  (OFF) υποδηλώνει ότι το στοιχείο  $v_i$  δεν ενεργοποιείται ή δεν εκφράζεται ή έχει ένα κάτω όριο συγκέντρωσης. Κάθε Boolean συνάρτηση αντιπροσωπεύει τους όρους εξάρτησης των στοιχείων εισόδου στη ρύθμιση της συγκέντρωσης του κατόντη στοιχείου-στόχου. Η μη-παραμετρική φύση και τα ποιοτικά χαρακτηριστικά της μοντελοποίησης κατά Boole την κάνει κατάλληλη για την ανάλυση σύνθετων συμπεριφορών ενός συστήματος μεγάλης κλίμακας, σαν τη δραστηριότητα των στοιχείων στη σταθερή κατάσταση, οι αλλαγές της δραστηριότητας των στοιχείων μετά από μία διαταραχή, οι σχέσεις εισόδου-εξόδου ενός συστήματος και η σταθερότητα των κυτταρικών αποκρίσεων σε ένα σήμα ([26]).

Η δυναμική μοντελοποίηση κατά Boole των βιολογικών δικτύων συνεπάγεται έξι βασικά βήματα, όπως φαίνεται και στην παρακάτω εικόνα. Στην παρούσα επισκόπηση, παρέχεται μία σύντομη περιγραφή καθενός από αυτά τα βήματα ([26]).



Εικόνα 51: Τα κύρια βήματα στη δυναμική μοντελοποίηση κατά Boole των βιολογικών συστημάτων ([26])

Το πρώτο βήμα είναι η *σύνθεση της δομής του δικτύου* από την εκτενή συλλογή της σχετικής βιβλιογραφίας (όπως υποστηρίζεται από τη διεθνή βάση δεδομένων) και των πειραματικών δεδομένων που αφορούν τα ενδιαφερόμενα βιολογικά συστήματα. Παρόλο που για πολλές βιολογικές διαδικασίες, τα διαφορετικά πειράματα παράγουν μία αφθονία σχετικών στοιχείων και αιτιωδών αντιδράσεων, υπάρχει ανεπαρκής πληροφόρηση για τη συνολική δομή και τους μηχανισμούς αυτών των διαδικασιών. Ως εκ τούτου, πηγές πληροφόρησης από μεμονωμένα πειράματα πρέπει να συλλέγονται και να ενσωματώνονται. Πειραματικά στοιχεία σχετικά με τη συμμετοχή ενός στοιχείου ή σχετικά με τις ρυθμιστικές σχέσεις μίας βιολογικής διαδικασίας, έχουν πολλά είδη. Για παράδειγμα, η αλλαγή της συγκέντρωσης μίας πρωτεΐνης, μετά την επεξεργασία του συστήματος με την είσοδο ενός σήματος υποδεικνύει ότι αυτή η πρωτεΐνη μπορεί να αποτελεί ένα στοιχείο αυτού του δικτύου μεταγωγής. Τέτοια στοιχεία μπορούν να συλλεχθούν από γονιδιακές εκφράσεις υψηλής απόδοσης, από πρωτεομικά και μεταβολομικά δεδομένα. Επιπροσθέτως, εάν η υπέρ-έκφραση ενός στοιχείου οδηγεί σε αλλαγές της σχετικής κυτταρικής απόκρισης, μπορεί να συναχθεί το συμπέρασμα ότι αυτό το στοιχείο εμπλέκεται στις βιολογικές διαδικασίες. Αιτιώδεις σχέσεις μεταξύ των στοιχείων μπορούν να συλλεχθούν από φωσφοπρωτεομικά δεδομένα υψηλής απόδοσης, από αντιδράσεις μεταξύ πρωτεϊνών και γενετικού υλικού καθώς και από μελέτες γενετικών αντιδράσεων. Αυτές οι πηγές δεδομένων μπορεί να μην είναι συγκεκριμένες για τα ενδιαφερόμενα βιολογικά συστήματα. Για να γίνουν αυτά τα συστήματα

συγκεκριμένου πλαισίου, μπορεί κάποιος να τα συμπληρώσει με τις άμεσες βιοχημικές ενδείξεις, από μικρής κλίμακας πειράματα. Εάν οι σχετικές πληροφορίες είναι επαρκείς, αυτές οι αιτιώδεις σχέσεις μπορούν να αντιπροσωπευθούν με προσανατολισμένες ακμές από ντο ένα στοιχείο στο άλλο που χαρακτηρίζονται από ένα από τα δύο σήματα: ενεργοποίησης (θετικά) ή ανασταλτικά (αρνητικά). Σε κάποιες περιπτώσεις, γενετικά στοιχεία από πολλαπλά πειράματα οδηγούν σε σύνθετες αιτιώδεις σχέσεις, οι οποίες μπορούν να αναλυθούν σε σχέσεις μεταξύ στοιχείων ανάλογα με τη συγκεκριμένη περίπτωση. Η ενασχόληση με τέτοιες περιστάσεις γίνεται πιο απλή με τη χρήση του λογισμικού «*NET-SYNTHESIS*», το οποίο βρίσκει το πιο φειδωλό δίκτυο, το οποίο ενσωματώνει όλα τα γνωστά στοιχεία και τις αιτιώδεις σχέσεις μεταξύ των ([26]).

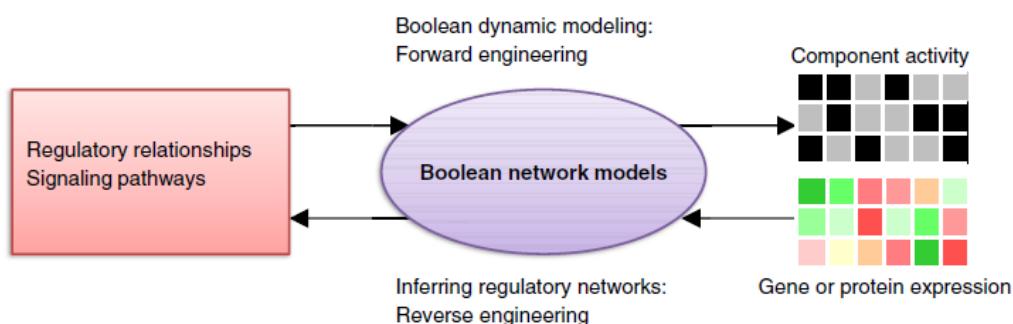
Η δομή του δικτύου που συναρμολογείται στο πρώτο βήμα περιορίζει σημαντικά, αλλά δεν καθορίζει μοναδικά, τις σχέσεις εξάρτησης μεταξύ των κόμβων κατάστασης. Έτσι, το δεύτερο βήμα είναι ο *καθορισμός των Boolean συναρτήσεων μεταφοράς*, οι οποίες βασίζονται στη βιβλιογραφία και στις πειραματικές παρατηρήσεις. Πολλά βιολογικά γεγονότα μπορούν να αντιπροσωπευθούν ποιοτικά από Boolean συναρτήσεις. Για παράδειγμα, θεωρούμε μία πρωτεΐνη P των οποίων η φωσφορυλιωμένη μορφή ρυθμίζει κατάντη διεργασίες (πιο απλά λέει πως είναι ενεργή). Μπορούμε να ορίσουμε την κατάσταση της πρωτεΐνης P με 1 (ON), εάν είναι κυρίως στη φωσφορυλιωμένη μορφή και με 0 (OFF), εάν είναι κυρίως στη μη-φωσφορυλιωμένη μορφή. Εάν η πρωτεΐνη P ρυθμίζεται αποκλειστικά από μία K κινάση, η οποία τη φωσφορυλιώνει, η Boolean συνάρτηση μεταφοράς για την κατάσταση της P μπορεί να γραφεί σαν  $P^* = K$ , όπου για λόγους απλότητας οι καταστάσεις των κόμβων συμβολίζονται από τα ονόματα των κόμβων και ο αστερίσκος σαν σύμβολο αναφέρεται στη νέα κατάσταση της P. Εάν εκτός από μία κινάση η πρωτεΐνη P αλληλεπιδρά και αποφωσφορυλιώνεται με φωσφατάση R, η συνάρτηση μεταφοράς για την πρωτεΐνη P μπορεί να γραφεί σαν  $P^* = \text{NOT } R$ . Σε πολλές περιπτώσεις, η ενεργοποίηση ενός στοιχείου απαιτεί δύο ή περισσότερους ρυθμιστές. Για παράδειγμα, η διαδικασία μεταγραφής ενός γονιδίου G μπορεί να απαιτεί ένα μεταγραφικό σύμπλοκο, το οποίο αποτελείται από δύο πρωτεΐνες  $P_1$  και  $P_2$ . Αυτό μπορεί να αναπαρίσταται από το τελεστή AND περιγράφοντας την ταυτόχρονη παρουσία δύο πρωτεϊνών:  $G^* = P_1 \text{ AND } P_2$ . Εάν ένα στοιχείο ρυθμίζεται από πολλαπλούς ρυθμιστές και οποιοσδήποτε από αυτούς μπορεί να ενεργοποιήσει το στοιχείο αυτό ανεξάρτητα, π.χ. μία πρωτεΐνη με πολλαπλές θέσεις φωσφορυλίωσης, τα ανεξάρτητα αποτελέσματα αυτών των ρυθμιστών στο στοιχείο-στόχο μπορούν να συλληφθούν από τον τελεστή OR. Η Boolean συνάρτηση μεταφοράς για ένα στοιχείο μπορεί να είναι ένας πολύπλοκος συνδυασμός των τελεστών AND, OR και NOT. Εάν η συνάρτηση μεταφοράς για ένα στοιχείο δεν είναι ευρέως γνωστή, διάφορες παραλλαγές μπορούν δοκιμαστούν από τη σύγκριση των δυναμικών ακολουθιών με την παρατήρηση του αληθινού συστήματος. Η δυναμική ενός Boolean μοντέλου πηγάζει από τη *μετάβαση της κατάστασης ενός συστήματος σε μία άλλη* και καθορίζεται από τις συναρτήσεις μεταφοράς και επιλέγεται από το επιλεγθέν σύστημα. Το σύγχρονο σύστημα είναι ο απλούστερος τρόπος ενημέρωσης, όπου οι καταστάσεις όλων των κόμβων ενημερώνονται συγχρόνως σύμφωνα με την τελευταία κατάσταση του συστήματος .

Αυτός ο τύπος ενημέρωσης υποδηλώνει ότι οι χρονικές κλίμακες όλων των βιολογικών συμβάντων στο σύστημα είναι παρόμοιες και οι καταστάσεις μετάβασης των στοιχείων συγχρονίζονται. Η κατάσταση τροχιάς του συστήματος καθορίζεται υπό το σύγχρονο σύστημα και οποιαδήποτε κατάσταση του συστήματος μπορεί να έχει το πολύ ένα διάδοχο στη μετάβαση κατάστασης δικτύου. Παρόλα αυτά, τα κυτταρικά συστήματα είναι αρκετά πολύπλοκα, και συχνά οι χρονικές κλίμακες των βιολογικών συμβάντων είναι διαφορετικές και μπορούν να ποικίλουν ευρέως από κλάσματα του δευτερολέπτου σε ώρες. Το σύγχρονο σύστημα δεν μπορεί να ευθύνεται για τέτοιες παραλλαγές. Στα ασύγχρονα μοντέλα, οι καταστάσεις των κόμβων ενημερώνονται με ένα μη-σύγχρονο τρόπο, ανάλογα με τα χρονοδιαγράμματα των επιμέρους βιολογικών συμβάντων. Υπάρχουν ντετερμινιστικά ασύγχρονα συστήματα με σταθερά μεμονωμένα χρονοδιαγράμματα και στοχαστικά ασύγχρονα συστήματα, όπου όλοι οι κόμβοι ενημερώνονται σύμφωνα με μία τυχαία σειρά ή ένας κόμβος είναι τυχαία επιλεγείς για να ενημερωθεί. Στα στοχαστικά ασύγχρονα μοντέλα, η ίδια αρχική κατάσταση μπορεί να οδηγήσει σε διάφορους διαδόχους στα δίκτυα μετάβασης κατάστασης, χάρις στην τυχαιότητα που διακρίνει το σύστημα ενημέρωσης. Τα ασύγχρονα συστήματα μπορούν να ενημερώνονται βάσει της γνώσης των χρονοδιαγραμμάτων κάποιων στοιχείων, εάν τέτοια γνώση είναι διαθέσιμη. Τα συστήματα ενημέρωσης έχουν μία σημαντική επίδραση στη δυναμική ενός συστήματος. Κάποιος μπορεί να διαλέξει ένα σύστημα το οποίο είναι πιο ρεαλιστικό για το ενδιαφερόμενο βιολογικό σύστημα ή μπορεί να συγκρίνει διαφορετικά συστήματα για το ίδιο σύστημα ([20]).

Ξεκινώντας από μία αρχική κατάσταση, το μοντέλο εξελίσσεται με την πάροδο του χρόνου μεταβαίνοντας από τη μία κατάσταση στην άλλη και τελικά σταθεροποιείται σε ένα σημείο έλξης, το οποίο αντιπροσωπεύει τη μακροπρόθεσμη συμπεριφορά του συστήματος. Τα σημεία έλξης των δικτύων ρύθμισης και σηματοδότησης αντιστοιχεί συνήθως στη σταθερή ενεργοποίηση της κατάστασης των στοιχείων ή στους κυτταρικούς φαινοτύπους. Ως εκ τούτου, η *αναγνώριση των πιθανών σημείων έλξης* είναι χρήσιμη και βιολογικά σχετική. Ιδιαίτερα, επιτρέπει την εξέταση των δραστηριοτήτων των στοιχείων σε μία σταθερή κυτταρική κατάσταση και τη σύγκριση των με πειραματικές παρατηρήσεις. Είναι, επίσης, χρήσιμος ο καθορισμός των κρίσιμων στοιχείων για τους κυτταρικούς φαινοτύπους εξετάζοντας τις αλλαγές στα σημεία έλξης του συστήματος, εάν ένα στοιχείο είναι σταθερό στην κατάσταση OFF ή υπερεκφρασμένες (σταθερά στην κατάσταση ON). Διαφορετικές αρχικές συνθήκες μπορούν να οδηγήσουν το σύστημα σε διαφορετικά σημεία έλξης. Ιδανικά, μπορούμε να ξεκινήσουμε από μία βιολογική σχετικά αρχική συνθήκη, εάν είναι γνωστή εκ των προτέρων. Εάν η διαθέσιμη πληροφορία είναι ανεπαρκής, μπορεί κανείς να δοκιμάσει ένα μεγάλο αριθμό αρχικών συνθηκών και να υπολογίζει τα κλάσματα υλοποίησης ενός συγκεκριμένου σημείου έλξης, αντιπροσωπεύοντας την πιθανότητα ότι το σύστημα επιταχύνει τον αντίστοιχο κυτταρικό φαινότυπο. Ως μοντέλο κατά Boole ενός δικτύου με  $n$  κόμβους έχει  $2^n$  πιθανές αρχικές συνθήκες ([20]).

Ένα σημαντικό βήμα της δυναμικής μοντελοποίησης των βιολογικών συστημάτων κατά Boole, είναι η *δοκιμή της ορθότητας του μοντέλο*.

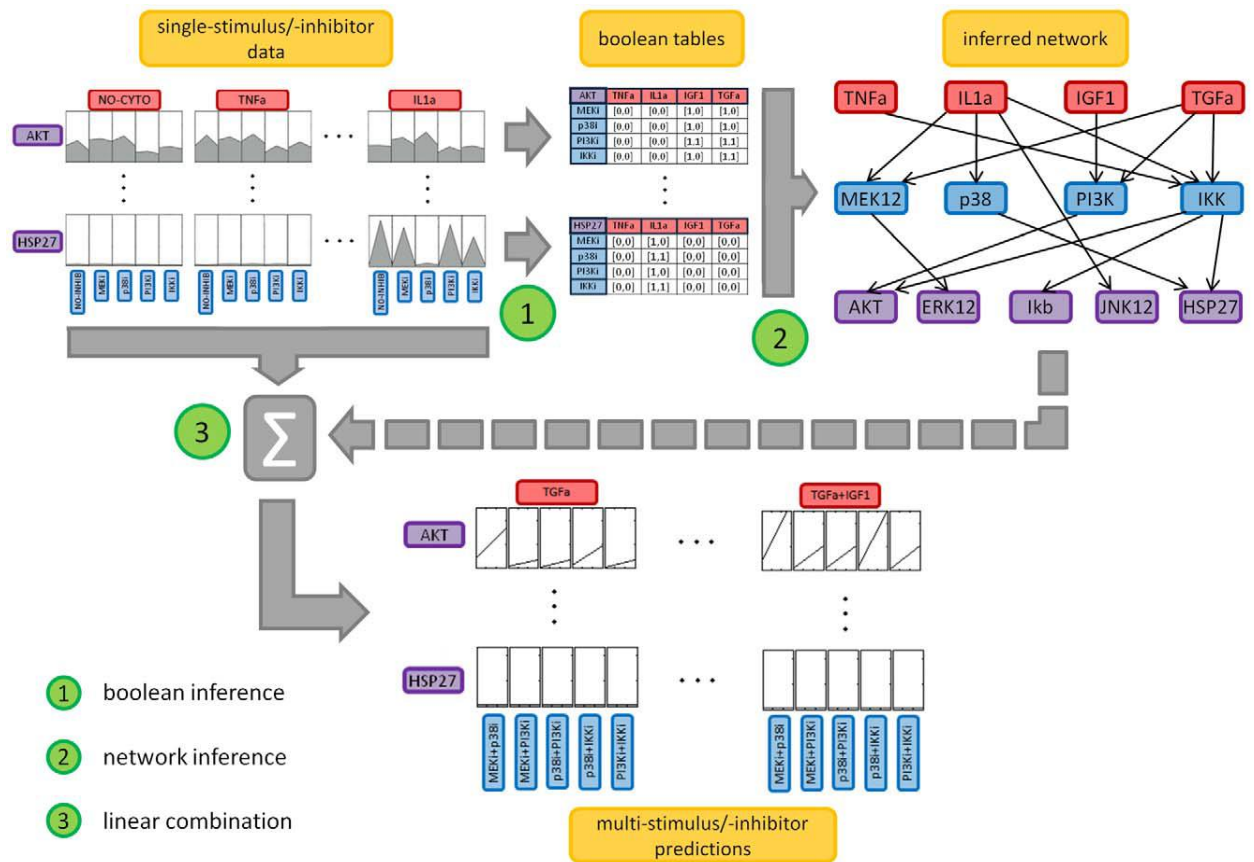
Υπάρχουν διάφοροι τρόποι να επικυρωθούν ορισμένα στοιχεία του μοντέλου. Για παράδειγμα, το μοντέλο πρέπει να είναι σε θέση να αναπαράγει εκ των προτέρων πειραματικές παρατηρήσεις, όπως είναι οι σχέσεις εισόδου-εξόδου, οι δυναμικές συμπεριφορές και οι κυτταρικές αποκρίσεις. Εάν το μοντέλο αποτύχει να το κάνει αυτό, χρειάζεται να πάμε πίσω και να επανεξετάσουμε εάν ορισμένα σημαντικά στοιχεία ή αντιδράσεις λείπουν από τη δομή του δικτύου ή εάν κάποιες Boolean συναρτήσεις μεταφοράς είναι ελλιπείς ή λανθασμένες (π.χ. χρήση AND αντί του OR). Η αποτυχία μπορεί, επίσης, να προκύψει από την ακαταλληλότητα των συστημάτων ενημέρωσης ή των αρχικών συνθηκών. Μετά από αρκετούς γύρους επαναλήψεων, μπορεί να ληφθεί ένα Boolean δυναμικό μοντέλο με γνωστές πειραματικές παρατηρήσεις. Μία προηγμένη στρατηγική για την επικύρωση του μοντέλου είναι η διεξαγωγή νέων πειραμάτων, σχετικά με τη δραστηριότητα κάποιων στοιχείων. Παρόλα αυτά, κάποια πειράματα μπορεί να είναι πολύ χρονοβόρα και μπορεί να μην είναι η πιο κατάλληλη επένδυση στο στάδιο της δοκιμής μοντέλων. Έχει αποκαλυφθεί πως πολλά βιολογικά συστήματα είναι εύρωστα, ώστε ένας έμμεσος τρόπος να επικυρωθεί το μοντέλο να είναι η ευρωστία του σε μικρές διαταραχές, όπως είναι οι εναλλαγές του OR και οι κανόνες του AND, η αλλαγή των πρόσημων των αντιδράσεων, η πρόσθεση ή η διαγραφή ενός στοιχείου ή μίας αντίδρασης. Ένα καλό μοντέλο μπορεί να φιλοξενήσει την πλειονότητα των μικρών διαταραχών, που αντανακλούν την προσαρμοστικότητα του συστήματος υπό διάφορες περιστάσεις ([26]).



Εικόνα 52: Εφαρμογές της μοντελοποίησης κατά Boole στα βιολογικά συστήματα ([26])

Αυτό το δυναμικό μοντέλο μπορεί να προβλέψει την ποιοτική δραστηριότητα των στοιχείων, όπως είναι οι πρωτεΐνες και οι μεταβολίτες. Η bottom-up προσέγγιση που χρησιμοποιείται για την κατασκευή του δυναμικού μοντέλου αναφέρεται πολλές φορές σαν «εμπρόσθια» μηχανική των βιολογικών ρυθμιστικών δικτύων. Από την άλλη μεριά, τεχνικές υψηλής απόδοσης μπορούν να παράγουν άφθονα προφίλ γονιδίων, πρωτεϊνών και μεταβολιτών σε μία σταθερή ή μεταβατική κατάσταση του κυττάρου. Οι μετρήσεις αυτές αντανακλούν τις ποιοτικές δραστηριότητες αυτών των βιολογικών στοιχείων στις αντίστοιχες βιολογικές διαδικασίες, οι οποίες προκύπτουν από τη ρύθμιση των δυναμικών επιπέδων του συστήματος διαφόρων μορίων. Η «αντίστροφη» μηχανική των βιολογικών δικτύων, όπως φαίνεται στην παραπάνω εικόνα, στοχεύει στην ανάκτηση αυτών των βασικών ρυθμιστικών

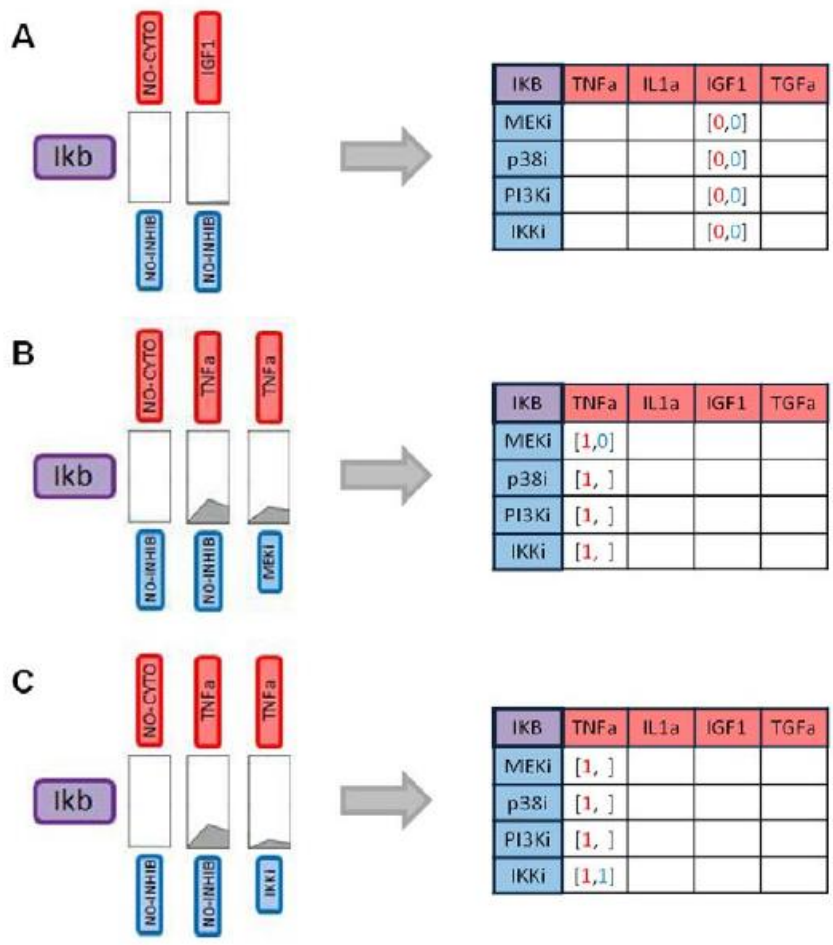
σχέσεων ή μονοπατιών σηματοδότησης από τα παρατηρούμενα δεδομένα δραστηριότητας.



Εικόνα 53: Διάγραμμα ροής της μεθοδολογίας Boole ([23])

Όπως φαίνεται και στην παραπάνω εικόνα, η μέθοδος που περιγράφεται αποτελείται από τρία βήματα βάσει: 1) εξαγωγή των μητρώων Boole από μονο-πειραματικά δεδομένα ερεθισμάτων-αναστολέων, 2) ανακατασκευή ενός δικτύου αιτίας-αποτελέσματος από τα μητρώα Boole, 3) γραμμικός συνδυασμός δεδομένων ερεθισμάτων-αναστολέων για την πρόβλεψη των επιπέδων πρωτεϊνικής δραστηριότητας, βάσει της δομής των δικτύων ([26]).

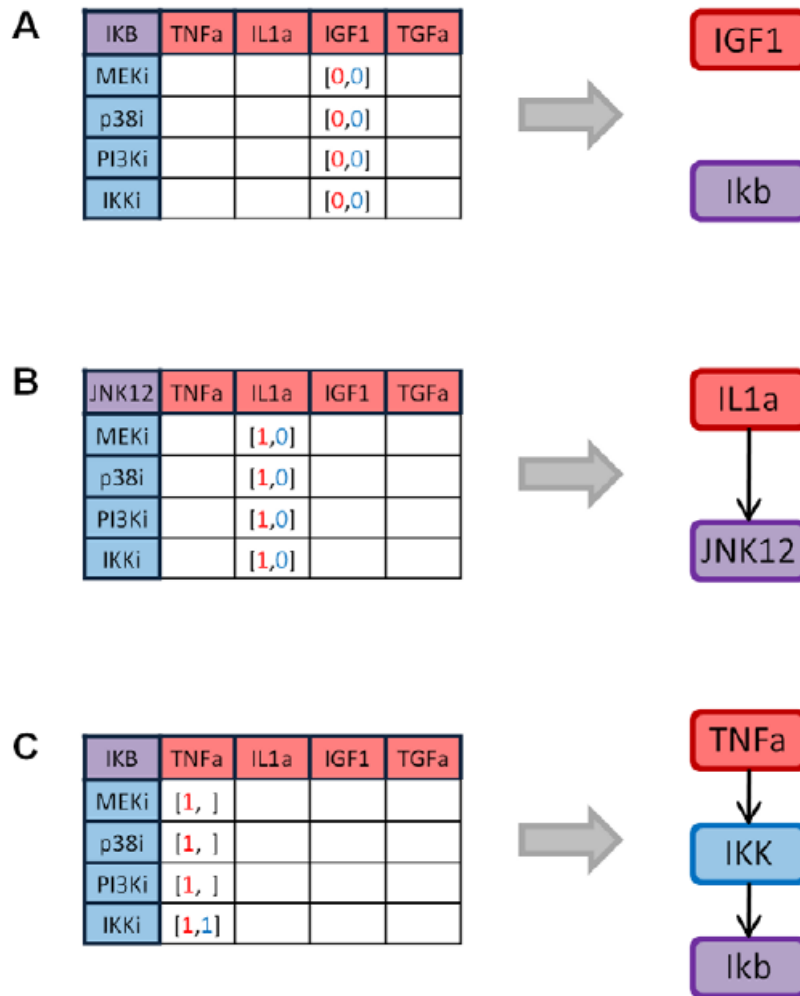
Στις παρακάτω εικόνες παρουσιάζονται πιο αναλυτικά και επεξηγηματικά τα ενδιάμεσα αυτά βήματα.



Εικόνα 54: Εξαγωγή μητρώων Boole ([23])

Στην παραπάνω εικόνα παρουσιάζονται τρία παραδείγματα: A. Το ερέθισμα IGF1 δεν επηρεάζει την πρωτεΐνη Ikb, B. Το ερέθισμα TNF $\alpha$  επηρεάζει την πρωτεΐνη Ikb, αλλά η παρουσία του αναστολέα MEK, δεν αλλάζει τα επίπεδα της πρωτεΐνης, C. Το ερέθισμα TNF $\alpha$  επηρεάζει την πρωτεΐνη Ikb και η παρουσία του αναστολέα IKK, μειώνει τα επίπεδα πρωτεΐνης.





Εικόνα 55: Ανακατασκευή δικτύων ([23])

Στην παραπάνω εικόνα παρουσιάζονται τρία παραδείγματα: A. Το ερέθισμα IGF1 δεν επηρεάζει την πρωτεΐνη Ikb, B. Το ερέθισμα IL1 $\alpha$  επηρεάζει την πρωτεΐνη JNK12, αλλά κανένας από τους αναστολείς δεν ασκεί σημαντική επίδραση, C. Το ερέθισμα TNF $\alpha$  επηρεάζει την πρωτεΐνη Ikb και στη δράση του διαμεσολαβεί η πρωτεΐνη IKK.

### 2.5.6 Περιορισμοί μεθόδων

Η προβλεπτική ικανότητα όλων των μεθοδολογιών μοντελοποίησης λανθάνει ακριβείας σύνδεσης πρωτεϊνών στο δίκτυο. Εάν το δίκτυο σηματοδότησης που χρησιμοποιείται για τον αλγόριθμο μοντελοποίησης δεν είναι ακριβές, οι προβλέψεις του μοντέλου θα είναι δυσνόητες. Για να παρακαμφθεί αυτό, πρέπει να ενσωματωθούν στη μοντελοποίηση του δικτύου τα πειραματικά δεδομένα, για τη σύλληψη της σηματοδοτημένης απόκρισης του εξεταζόμενου

συστήματος. Έπειτα, ένας αναπτυσσόμενος αλγόριθμος ετοιμάζει το μοντέλο σχετικά με τα ειδικά δεδομένα του συστήματος, για την εξασφάλιση μοντέλων πρόβλεψης.

## 2.6 Πειραματικά δεδομένα στη συστημική βιολογία

Στην ενότητα αυτή θα αναλυθούν οι βασικές αρχές της πρωτεϊνωματικής. Επίσης, θα αναλυθούν εκτενώς οι βασικές έννοιες περί πρωτεϊνών και της σημασίας που έχουν για τον επιστημονικό κλάδο της συστημικής βιολογίας και κυρίως για τη μοντελοποίηση των δικτύων μεταγωγής σημάτων. Τέλος, θα παρουσιαστεί το υπόβαθρο των δεδομένων που χρησιμοποιήθηκαν στην παρούσα διπλωματική και γενικά των πειραματικών δεδομένων που παράγονται από το εργαστήριο «Συστημικής βιολογίας και Εμβιομηχανικής» της σχολής μηχανολόγων μηχανικών του Εθνικού Μετσόβιου Πολυτεχνείου, μέσω της δημοφιλούς τεχνικής μετρήσεως πρωτεϊνών που ονομάζεται ELISA.

### 2.6.1 Πρωτεϊνωματική

Πρωτεϊνωματική (proteomics) είναι η μεγάλης κλίμακας ανάλυση των πρωτεϊνών ενός κυττάρου ή οργανισμού. Γενικά η κατάληξη «-ωματική» (-omics) χρησιμοποιείται για να περιγράψει την ανάλυση μακρομορίων ενός ολόκληρου συστήματος. Η ορολογία προκύπτει ως επέκταση της έννοιας του αρχικού όρου γονιδιωματική (genomics), ο οποίος αναφέρεται στην ανάλυση του γονιδιώματος (genome), του συνόλου δηλαδή των γονιδίων του οργανισμού ([16]).

Σύμφωνα με το κεντρικό δόγμα, η διαδρομή της γενετικής πληροφορίας ξεκινά από το γονιδίωμα και τερματίζει στις πρωτεΐνες. Οι πρωτεΐνες (το σύνολο των πρωτεϊνών που εκφράζονται σε ένα βιολογικό σύστημα) είναι οι τελικοί αποδέκτες της γενετικής πληροφορίας και ως εκ τούτου αποτελούν, από βιολογικής άποψης τουλάχιστον, τον φαινότυπο<sup>1</sup> που βλέπουμε σε κάθε σύστημα ([16]).

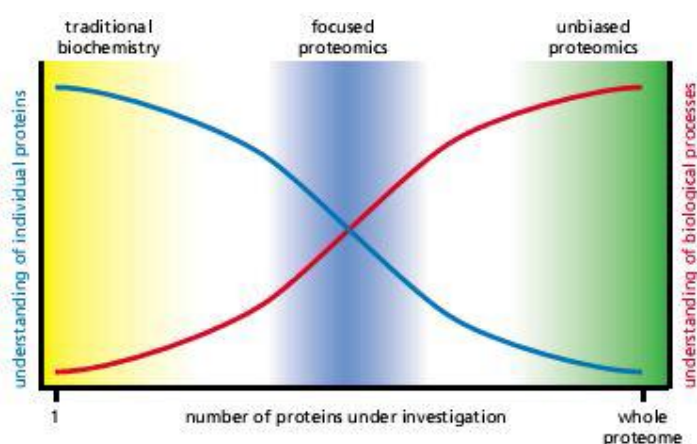
### Γιατί είναι αναγκαία η μελέτη ενός κυττάρου σε πρωτεϊνικό επίπεδο;

Η απάντηση είναι ότι η μελέτη σε γονιδιακό επίπεδο δεν αρκεί. Η λειτουργία ενός κυττάρου είναι πολύ πολύπλοκη και δεν μπορεί να ερμηνευτεί μόνο μέσω της έκφρασης του γονιδιώματος. Το κύτταρο βασίζεται, επιπλέον, σε διάφορα μονοπάτια (σειρές χημικών αντιδράσεων) για τη λειτουργία του (μεταβολικά, μετάδοσης σήματος κ.ά.). Κατά συνέπεια, η τελική έκφραση του πρωτεϊνώματος<sup>2</sup> δεν μπορεί να εξαχθεί από την ευθεία γραμμή αιτιότητας, που υπονοεί το κεντρικό δόγμα, αφού υπάρχουν πολλά μετα-μεταφραστικά (post-translational) φαινόμενα που το καθορίζουν ([40]).

<sup>1</sup> **Φαινότυπος** είναι η σύνθεση των παρατηρήσιμων χαρακτηριστικών και των γνωρισμάτων ενός οργανισμού, όπως η μορφολογία του, η βιοχημικές ή φυσιολογικές του ιδιότητες, η συμπεριφορά του ή τα αποτελέσματα της συμπεριφοράς του (πχ η φωλιά των πουλιών), σε μια δεδομένη χρονική στιγμή. Αποτελεί δηλαδή το μέρος του γονότυπου του οργανισμού το οποίο μπορούμε (άμεσα ή έμμεσα) να παρατηρήσουμε. (Wikipedia) ([40])

Η «παραδοσιακή» πρωτεϊνωματική ανάλυση συνίσταται στο διαχωρισμό-απομόνωση και αναγνώριση της δομής και, κατά συνέπεια, της λειτουργίας των πρωτεϊνών. Ο διαχωρισμός των πρωτεϊνών γίνεται συνήθως με την τεχνική της μονοδιάστατης ή δισδιάστατης ηλεκτροφόρησης σε τζελ (1οr2-DGelElectrophoresis), ενώ η ανάλυση με την τεχνική της φασματομετρίας μάζας (MassSpectrometry). Οι τεχνικές αυτές συνδυαζόμενες επιτρέπουν να διαχωριστούν και αναλυθούν μέχρι 1.000 πρωτεΐνες, αριθμός που δεν θεωρείται ιδιαίτερα μεγάλος, δεδομένου ότι σε κάθε δείγμα ο αριθμός των πρωτεϊνών μπορεί να είναι της τάξης του  $10^6$  ([40]).

Η συστημική ανάλυση δεν ξεκινάει από το μηδέν όπως η παραδοσιακή ανάλυση. Αντιθέτως χτίζει πάνω στην υπάρχουσα γνώση. Στόχος της συστημικής ανάλυσης είναι η κατανόηση συγκεκριμένων φαινοτύπων, γι' αυτό και δεν στοχεύει στην αδιάκριτη αναγνώριση πρωτεϊνών, αλλά στην παρακολούθηση συγκεκριμένων πρωτεϊνών στόχων για τις οποίες εκτιμάται, εκ των προτέρων, ότι σχετίζονται (ή είναι πιθανό να σχετίζονται) με τον υπό ανάλυση φαινότυπο ([40]).



Εικόνα 56: " Φάσμα "πρωτεϊνικής ανάλυσης ([40])

Ως εκ τούτου, πειράματα που εντάσσονται σε μια συστημική ανάλυση δεν αποσκοπούν στην άκριτη αναγνώριση πρωτεϊνών, αλλά στην ποσοτικοποίηση της έκφρασης μερικών κρίσιμων εξ αυτών. Για μια συστημική ανάλυση, κρίσιμο μέγεθος δεν είναι το ποιες πρωτεΐνες υπάρχουν σε ένα κύτταρο, αλλά ποιες είναι ενεργές, ποιων η έκφραση, δηλαδή, διαφοροποιείται σε μια χρονική περίοδο ως απάντηση σε κάποιο ερέθισμα. Επομένως, ο φαινότυπος, υπό το πρίσμα της πρωτεϊνωματικής, είναι το αποτέλεσμα «μονοπατιών» αλληλεπίδρασης, δηλαδή της σειριακής ενεργοποίησης διαφόρων πρωτεϊνών. Μια πρωτεΐνη «ενεργοποιείται» μέσω της διαδικασίας της φωσφορυλίωσης ([40]).

<sup>2</sup>Πρωτεϊνωμα (κατά το γονίδιο – γονιδίωμα), είναι το σύνολο το πρωτεϊνών που εκφράζονται από ένα συγκεκριμένο κύτταρο (ή οργανισμό) σε κάποιο δεδομένο χρόνο για δεδομένες συνθήκες ([40]).

## 2.6.2 Μετρήσεις πρωτεϊνών

Όπως είδαμε τα γονίδια περιέχουν συμπυκνωμένη όλη την πληροφορία που απαιτείται για τη λειτουργία ενός οργανισμού. Από την άλλη πλευρά οι τελικοί εκφραστές, που θέτουν σε εφαρμογή τις διάφορες εντολές, είναι οι πρωτεΐνες. Οι πρωτεΐνες είναι υπεύθυνες για την εκτέλεση κάθε λειτουργίας καθώς και για τις φυσικές ιδιότητες, την εμφάνιση και το σχήμα κάθε μικροσκοπικού ή μεγάλου χαρακτηριστικού ενός οργανισμού. Επομένως, αν μπορέσουμε να βρούμε πως επηρεάζεται η παραγωγή και διακίνηση των πρωτεϊνών, αλλά και ποιες πρωτεΐνες είναι υπεύθυνες για ποια χαρακτηριστικά, θα μπορέσουμε να προβλέψουμε και να καθορίσουμε σε μεγάλο βαθμό τη λειτουργία ενός οργανισμού ([9]).

Φυσικά, το να μετρήσουμε όλες τις πρωτεΐνες, σε όλα τα δυνατά σχήματα και όλους τους δυνατούς συνδυασμούς πρωτεϊνών την κάθε χρονική στιγμή, όπως μπορούμε να μετρήσουμε το συνολικό γονιδίωμα ενός ατόμου, είναι απλά αδύνατο. Οι πρωτεΐνες δεν καθορίζονται απλά από την ακολουθία των αμινοξέων από τα οποία αποτελούνται. Η δομή και το σχήμα τους, που δεν είναι απαραίτητως σταθερά στο χρόνο και καθορίζονται από διάφορους εξωτερικούς παράγοντες, όπως το χημικό περιβάλλον, η θερμοκρασία κ.λπ., παίζουν καθοριστικό ρόλο στη λειτουργία διαφόρων πρωτεϊνών. Επίσης, σύνθεση διαφόρων πρωτεϊνών μεταξύ του μπορεί να δημιουργήσει πρωτεΐνες με διαφορετικές ιδιότητες. Με τον τρόπο αυτό δημιουργούνται άπειροι συνδυασμοί από τους οποίους προκύπτει και η μεγάλη ποικιλία στον φαινότυπο των έμβιων οργανισμών ([9]).

Ένα ακόμα μεγάλο ζήτημα με τη μέτρηση πρωτεϊνών είναι ότι κάθε φορά μπορούμε να μετράμε μόνο μια συγκεκριμένη δομή μιας πρωτεΐνης και δεν μπορούμε να μετρήσουμε ταυτόχρονα ακόμα και το ίδιο είδος πρωτεΐνης, αν έχει παρεμβληθεί κάποιος μετασχηματισμός. Επιπλέον, κάποιες πρωτεΐνες βρίσκονται σε μεγάλη πληθώρα στα κύτταρα, ενώ άλλες σε απειροελάχιστες ποσότητες, κάνοντας πολύ δύσκολο τον εντοπισμό των δευτέρων. Τέλος, ένα άλλο πρόβλημα στη μέτρηση των πρωτεϊνών είναι ότι δεν μπορούμε να τις αναπαράγουμε στο εργαστήριο όπως κάνουμε με το DNA και έτσι δεν μπορούμε να ενισχύσουμε το σήμα ενός δείγματος ([9]).

Στη συνέχεια θα προχωρήσουμε στην παρουσίαση ορισμένων μόνο μεθόδων που χρησιμοποιούνται για την μέτρηση πρωτεϊνών. Όπως και στην περίπτωση της μέτρησης του DNA, οι μέθοδοι αυτές χωρίζονται σε μαζικές και μη μαζικές, ανάλογα με το αν μπορεί να γίνει παράλληλη επεξεργασία μεγάλου αριθμού δειγμάτων. Μια άλλη διάκριση που θα μας απασχολήσει περισσότερο στο σημείο αυτό είναι με βάση το αν ξεκινώντας τη μέτρηση γνωρίζουμε ποιες πρωτεΐνες θα μετρήσουμε ή όχι ([9]).

Αρχικά θα προχωρήσουμε στην συνοπτική παρουσίαση των μεθόδων όπου δεν μετράμε συγκεκριμένες πρωτεΐνες, αλλά προσπαθούμε να μετρήσουμε όσο το δυνατόν το συνολικό σώμα των πρωτεϊνών σε ένα δείγμα. Στις μεθόδους αυτές θέλουμε να μετρήσουμε όσο το δυνατόν περισσότερα, αλλά μπορούμε να επεξεργαστούμε παράλληλα πολύ μικρό αριθμό δειγμάτων. Για το λόγο αυτό, οι μέθοδοι αυτές είναι γενικά μη μαζικές μέθοδοι ([9]).

Στη συνέχεια θα παρουσιάσουμε συνοπτικά και τις μεθόδους στις οποίες θέλουμε να μετρήσουμε την ύπαρξη συγκεκριμένων πρωτεϊνών στο δείγμα. Στην περίπτωση αυτή είναι περιορισμένος ο αριθμός των πρωτεϊνών, την ύπαρξη των οποίων μπορούμε να μετρήσουμε κάθε φορά, αλλά είναι πολλά τα δείγματα που μπορούμε να επεξεργαστούμε παράλληλα. Έτσι οι μέθοδοι αυτές είναι γενικά μαζικές (high-throughput).

### **Μέτρηση άγνωστων πρωτεϊνών (μη μαζικές μέθοδοι)**

---

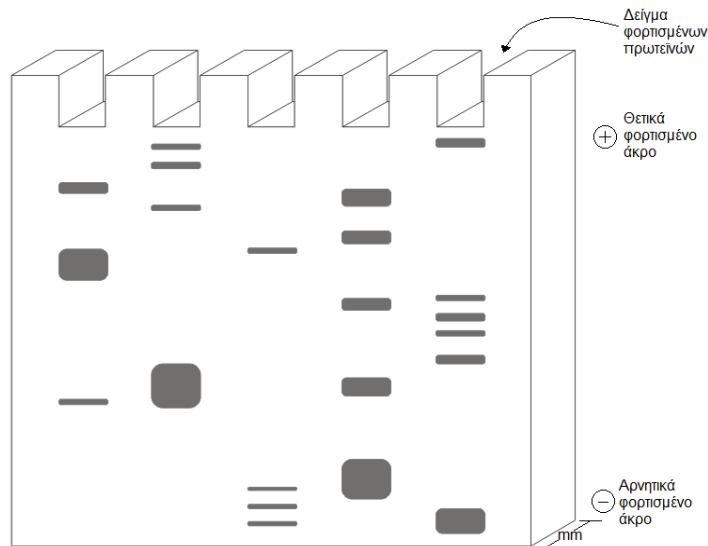
Στις μεθόδους αυτές χρησιμοποιούμε κάποιες εκ των προτέρων γνωστές ιδιότητες των πρωτεϊνών, όπως είναι η μάζα και το ηλεκτρικό φορτίο τους. Οι μέθοδοι αυτές είναι γενικά ποιοτικές και έχουν χαμηλή ακρίβεια. Χρησιμοποιούνται κυρίως για να βρεθούν μεγάλες διαφορές στην παράγωγη πρωτεϊνών ανάμεσα σε υγιή και ασθενή κύτταρα. Στη συνέχεια παρουσιάζουμε συνοπτικά ορισμένες από αυτές τις μεθόδους ([9]).

#### ➤ Western blot

Η μέθοδος αυτή κυριάρχησε για πολλά χρόνια στο χώρο της μέτρησης πρωτεϊνών. Είναι σχετικά απλή μέθοδος, αλλά είναι γενικά δύσκολο να αυτοματοποιηθεί, ενώ τα αποτελέσματα που δίνει είναι κυρίως ποιοτικά. Επίσης, σε σχέση με τις μαζικές μεθόδους που θα εξετάσουμε στη συνέχεια είναι και αρκετά χρονοβόρα. Πάνω στη μέθοδο αυτή έχουν βασιστεί και οι υπόλοιπες μη μαζικές μέθοδοι, αναπτύσσοντας κατά καιρούς διάφορες παραλλαγές της ([9]).

Η μέθοδος αυτή χρησιμοποιεί μια διαδικασία που ονομάζεται ηλεκτροφόρηση γέλης (gel electrophoresis) για τον διαχωρισμό των πρωτεϊνών του δείγματος. Γενικά ο διαχωρισμός μπορεί να γίνει ανάλογα με τις ηλεκτρικές ιδιότητες μιας πρωτεΐνης, το μοριακό της βάρος ή συνδυασμό παραγόντων ([9]).

Η διαδικασία για την μέτρηση πρωτεϊνών σε western blot έχει ως εξής: Εφαρμόζουμε διαφορά δυναμικού σε σώμα γέλης όπου έχουμε τοποθετήσει και το δείγμα των ηλεκτρικά φορτισμένων πρωτεϊνών. Οι πρωτεΐνες κινούνται προς τα κάτω μέσα στη γέλη με ταχύτητα που εξαρτάται από τη μάζα τους. Αν εμείς γνωρίζουμε τις μάζες διάφορων πρωτεϊνών μπορούμε ανάλογα με το βάθος όπου έφτασε το δείγμα στη γέλη να τις ταυτοποιήσουμε. Έτσι αναγνωρίζουμε πρωτεΐνες με περίπου γνωστό μέγεθος, ενώ υπολογίζουμε ποιοτικά από το αποτύπωμα που δημιουργείται και την ποσότητά τους. Τα αποτελέσματα αποτυπώνονται σε διαφανές χαρτί που ακουμπάμε πάνω στη γέλη. Ο αριθμός των δειγμάτων που μπορούμε να διαχειριστούμε παράλληλα είναι γενικά περιορισμένος ([9]).



Εικόνα 57: Μέθοδος Western blot ([9])

➤ 2D Page

Η μέθοδος αυτή είναι εντελώς παρόμοια με την προηγούμενη, μόνο που στην περίπτωση αυτή έχουμε μόνο ένα δείγμα για το οποίο μπορούμε να μετρήσουμε παράλληλα δύο ιδιότητες. Στην μια διεύθυνση κίνησης του δείγματος μπορούμε για παράδειγμα να μετράμε τη μάζα των πρωτεϊνών και στην άλλη διεύθυνση το ηλεκτρικό τους φορτίο ([9]).

➤ Φασματοσκοπία μάζας (Mass spectrometry - MS)

Η φασματοσκοπία μάζας είναι μια μέθοδος που χρησιμοποιείται ευρέως για τον προσδιορισμό της μάζας σωματιδίων σε ένα μίγμα ή μόριο. Το δείγμα ιοντίζεται και στη συνέχεια μπαίνει σε ένα σωλήνα όπου εφαρμόζεται ηλεκτρικό πεδίο. Οι πρωτεΐνες του δείγματος επιταχύνονται ανάλογα με τη μάζα τους και η συσκευή μετρά και αποδίδει το φάσμα των μαζών του δείγματος. Αντιστοιχίζοντας γνωστές μάζες σε πρωτεΐνες αναγνωρίζουμε τελικώς τη σύσταση του δείγματος. Η μέθοδος αυτή δεν αποδίδει ούτε καν ποιοτικά αποτελέσματα για την ποσότητα των διάφορων πρωτεϊνών ([9]).

**Μέτρηση γνωστών πρωτεϊνών (μαζικές μέθοδοι)**

Στις μεθόδους αυτές μετράμε συγκεκριμένες πρωτεΐνες μέσα σε ένα δείγμα. Η μέτρηση των πρωτεϊνών αυτών γίνεται με χρήση συγκεκριμένων αντισωμάτων. Όπως είδαμε στο κεφάλαιο 2 όλες οι πρωτεΐνες έχουν κάποιες περιοχές σύνδεσης, όπου συνδέονται ισχυρά με άλλες πρωτεΐνες. Οι περιοχές αυτές είναι συγκεκριμένες για κάθε ζεύγος πρωτεϊνών. Το χαρακτηριστικό αυτό εκμεταλλευόμαστε και παράγουμε αντισώματα που

συνδέονται αποκλειστικά και μόνο με συγκεκριμένες πρωτεΐνες. Επειδή οι μέθοδοι αυτές βασίζονται σε αυτή την ιδιότητα ονομάζονται και «μέθοδοι βασισμένες συνάφεια» (affinity based methods) ([9]).

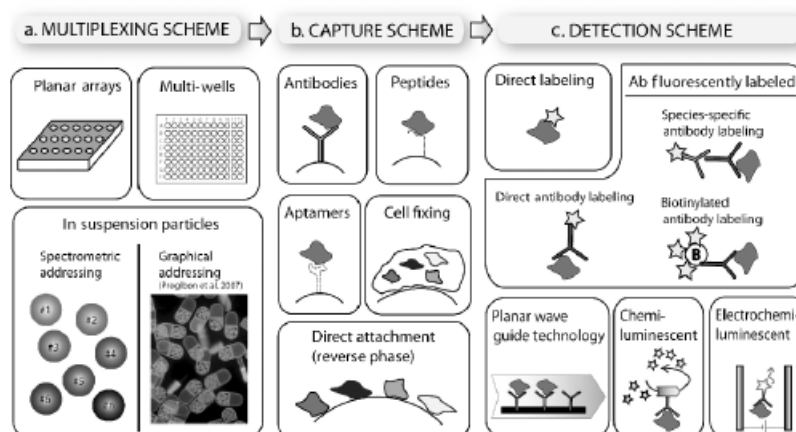
Η συνάφεια, δηλαδή το πόσο ισχυρός είναι ο δεσμός, μεταξύ του αντισώματος και της εν λόγω πρωτεΐνης είναι κρίσιμης σημασίας για τις μεθόδους αυτές. Για το λόγο αυτό έχουν αναπτυχθεί διάφορες τεχνολογίες που παρέχουν μεγαλύτερη ή μικρότερη ακρίβεια μετρήσεων ανάλογα με τα αντισώματα τα οποία χρησιμοποιούν. Σε κάθε περίπτωση, μόνο οι μέθοδοι αυτές δίνουν τη δυνατότητα να μετρήσουμε ακόμα και πρωτεΐνες που βρίσκονται σε πολύ μικρή ποσότητα στο δείγμα, πράγμα αδύνατο για τις μεθόδους που αναλύσαμε μέχρι τώρα ([9]).

Υπάρχουν τρία βασικά χαρακτηριστικά τα οποία ικανοποιεί κάθε μέθοδος που βασίζεται στη συνάφεια ([15]).

1. Ένας τρόπος αναγνώρισης (identification or multiplexing scheme) είναι απαραίτητος για να μπορεί να διαχειριστεί κανείς ταυτόχρονα πολλά δείγματα και να τα ξεχωρίσει από το σύνολο κατά τη μέτρηση. Το χαρακτηριστικό αυτό επιτυγχάνεται χωρίζοντας το δείγμα σε διαφορετικά πηγαδάκια (wells), σε διαφορετικά σημεία σε μια γυάλινη πλάκα ή επάνω σε μικροσφαιρίδια με διαφορετικές ιδιότητες.

2. Ένας τρόπος σύνδεσης (capture scheme) που σκοπό έχει να ακινητοποιήσει τις πρωτεΐνες που θέλουμε να μετρήσουμε στα σημεία αναγνώρισης που περιγράψαμε παραπάνω. Για τον σκοπό αυτό χρησιμοποιούνται κυρίως τα επιλεγμένα αντισώματα, αλλά και άλλες μέθοδοι.

3. Ένας τρόπος αναγνώρισης (detection scheme) που ιδανικά μπορεί να μετρήσει μια ιδιότητα που είναι ανάλογη της ποσότητας της εκάστοτε πρωτεΐνης στο δείγμα.



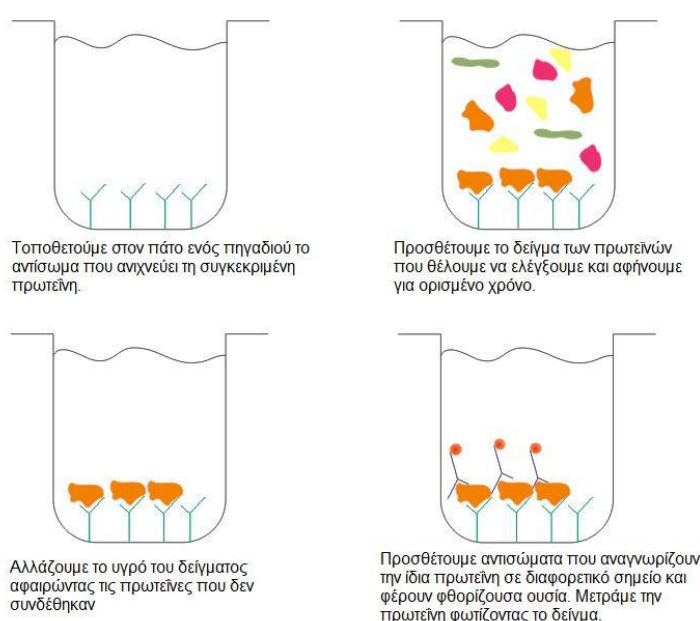
Εικόνα 58: Βασικά χαρακτηριστικά μεθόδων συνάφειας ([9])



Στη συνέχεια θα προχωρήσουμε στη συνοπτική παρουσίαση ορισμένων μεθόδων αυτής της κατηγορίας. Οι μέθοδοι αυτές με τη σειρά τους μπορεί να έχουν διάφορες παραλλαγές αλλά εμείς θα περιοριστούμε σε μια συνοπτική και γενική παρουσίαση.

### ➤ ELISA

Η μέθοδος αυτή είναι μια από τις πιο διαδεδομένες ενώ έχει και πολλές παραλλαγές. Το όνομά της προέρχεται από το ακρωνύμιο Enzyme Linked Immunosorbent Assay. Στη συνέχεια θα περιγράψουμε σχηματικά μια παραλλαγή που ονομάζεται sandwich ELISA ως ένα παράδειγμα.



**Εικόνα 59: Μέθοδος ELISA ([9])**

### ➤ Protein Microarrays

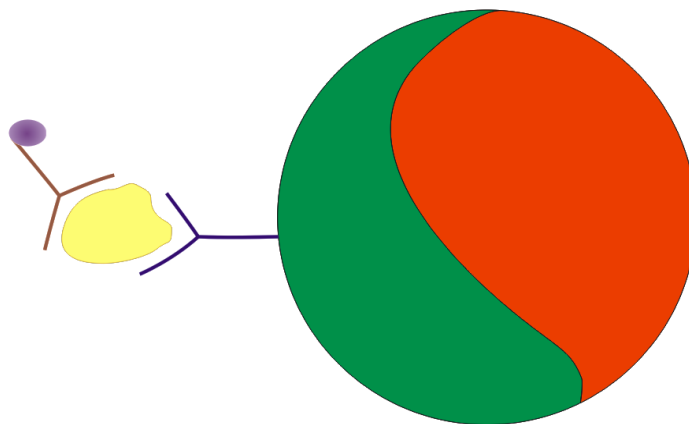
Η μέθοδος αυτή μοιάζει πολύ με τη μέθοδο ELISA με την μόνη διαφορά ότι το αρχικό αντίσωμα δεν στερεώνεται σε ξεχωριστά πηγαδάκια αλλά πάνω ε μια γυάλινη πλάκα. Με τον τρόπο αυτό μπορούμε να μετρήσουμε μεγάλες ποσότητες δείγματος, αν και υπάρχουν περιορισμοί ως προς την ακρίβεια της μεθόδου. Η στοίχιση του δείγματος στην πλάκα μπορεί να έχει κάποιο πρόσθετο νόημα και να έχει προκύψει για παράδειγμα με τη μέθοδο 2D ([9]).

➤ Meso-Scale Discovery (MSD)

Η μέθοδος αυτή είναι παρόμοια με τη μέθοδο ELISA και διαφέρει στον τρόπο που γίνεται η αναγνώριση των αντισωμάτων. Λόγω αυτής της διαφορετικής διαδικασίας μπορούν να τοποθετηθούν σε κάθε πηγαδάκι μέχρι 10 διαφορετικά αντισώματα, τα οποία θα αναγνωρίσουν 10 διαφορετικές πρωτεΐνες και μάλιστα με μεγάλη ακρίβεια. Το μειονέκτημα της μεθόδου αυτής είναι ότι απαιτεί συγκεκριμένα αναλώσιμα που παρέχονται μόνο από συγκεκριμένες εταιρίες ([9]).

➤ xMap Technology

Η τεχνολογία αυτή, που είναι προϊόν της εταιρείας Luminex Corporation, έχει ένα ιδιαίτερο πλεονέκτημα σε σχέση με όλες τις άλλες μεθόδους. Χρησιμοποιεί μικροσφαιρίδια (beads) διαμέτρου της τάξης των μερικών μm πάνω στα οποία τοποθετούνται τα αντισώματα που προσδένουν τις πρωτεΐνες. Τα μικροσφαιρίδια αυτά είναι από πολυμερή υλικά και περιέχουν, σε διάφορα ποσοστά φθορίζουσες ουσίες, δίνοντας έτσι τη δυνατότητα ανίχνευσης 100 διαφορετικών χρωμάτων. Επομένως, σε κάθε δείγμα θα μπορούσαν να μετρηθούν 100 διαφορετικές πρωτεΐνες, αν και αυτό δεν είναι πρακτικά επιτεύξιμο λόγω άλλων περιορισμών. Τα μικροσφαιρίδια προστίθενται στο δείγμα και το κάθε σφαιρίδιο δρα σαν ένα ξεχωριστό πηγαδάκι της μεθόδου ELISA. Στη συνέχεια τα σφαιρίδια με τις πρωτεΐνες που έχουν προσδεθεί περνάν από ένα ανιχνευτή laser όπου εντοπίζεται το χρώμα του σφαιριδίου, άρα ο τύπος της πρωτεΐνης που έχει προσδεθεί, και ο φθορισμός του δεύτερου αντισώματος που αποδεικνύει ότι πράγματι το σφαιρίδιο φέρει την πρωτεΐνη. Μετρώντας τον αριθμό των αριθμό των σφαιριδίων κάθε χρώματος έχουμε και μια εικόνα για την ποσότητα της πρωτεΐνης. Μια παραλλαγή της μεθόδου αυτής χρησιμοποιεί μαγνητικά μικροσφαιρίδια που διευκολύνουν την επεξεργασία του δείγματος και αυξάνουν την ταχύτητα της διαδικασίας ([9]).

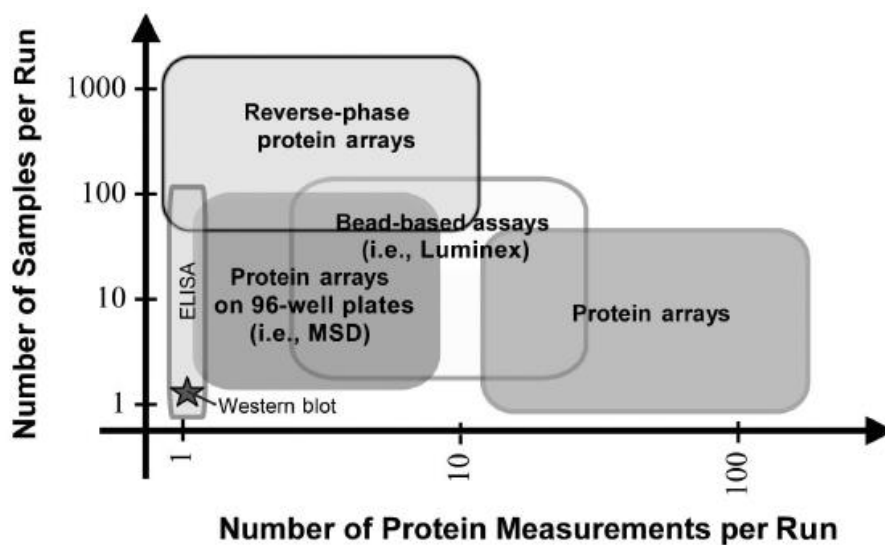


Εικόνα 60: Μικροσφαιρίδιο ([9])

➤ Reverse Phase

Η μέθοδος αυτή, όπως και η προηγούμενες, χρησιμοποιεί αντισώματα για την ανίχνευση των πρωτεϊνών, μόνο που στην περίπτωση αυτή απαιτείται μόνο ένα αντίσωμα. Το δείγμα προσκολλάται απευθείας σε μια πλάκα ή μεμβράνη και στη συνέχεια μπορεί να χρησιμοποιηθούν διάφορες μέθοδοι για την ανίχνευση των πρωτεϊνών. Μπορεί να χρησιμοποιηθούν διαφορετικά δευτερεύοντα αντισώματα σε διαφορετικά βήματα της διαδικασίας και να μετρηθούν με τον τρόπο αυτό πολλές πρωτεΐνες χωρίς να μειώνεται η ισχύς του δείγματος. Τα μειονεκτήματα του σχετίζονται με την πιθανότητα να μετράμε δευτερεύοντα αντισώματα που όμως δεν έχουν προσδεθεί σε κάποια πρωτεΐνη ([9]).

Έχοντας αναλύσει τις πιο διαδεδομένες από τις μεθόδους μέτρησης πρωτεϊνών προχωράμε σε μια σχηματική σύγκρισή τους. Γενικά δεν υπάρχει αμφιβολία ότι οι μαζικές μέθοδοι σύντομα θα βελτιωθούν και θα κυριαρχήσουν, δίνοντας όλο και μεγαλύτερες δυνατότητες στο πεδίο της μέτρησης πρωτεϊνών (proteomics) ([9]).



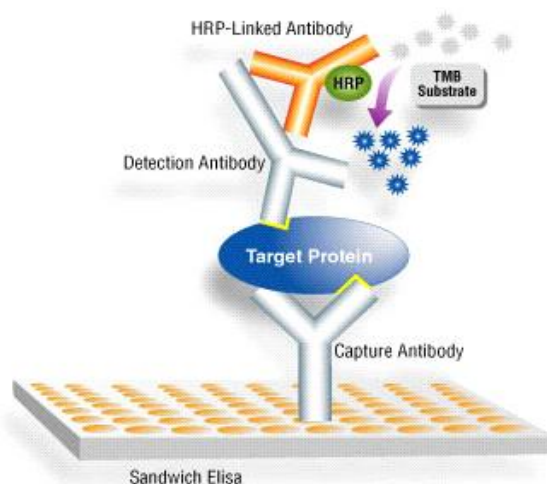
Εικόνα 61: Σύγκριση μεθόδων ([15])

Παρακάτω θα αναφερθούμε πιο εκτενώς στη μέθοδο μέτρησης πρωτεϊνών ELISA, η οποία χρησιμοποιείται τα τελευταία χρόνια στο εργαστήριο «Συστημικής βιολογίας και Εμβιομηχανικής» του Εθνικού Μετσόβιου Πολυτεχνείου με μεγάλη επιτυχία για τη μέτρηση πρωτεϊνικών μορίων σε κυτταρικά δείγματα. Οι μετρήσεις αυτές αποσκοπούν στην παραγωγή κατάλληλων πειραματικών δεδομένων, τα οποία προάγουν την έρευνα στο συγκεκριμένο επιστημονικό πεδίο της συστημικής βιολογίας.

### 2.6.3 Τεχνική ELISA

Η ELISA (**Enzyme-linked immunosorbent assay**) είναι μία διαδομένη μορφή αναλυτικής βιοχημικής δοκιμής η οποία χρησιμοποιεί ένα υπό-είδος ετερογενούς και στερεάς φάσης ενζυματικής ανοσοανίχνευσης (EIA) με σκοπό την ανίχνευση μίας ουσίας, συνήθως ενός αντιγόνου, σε ένα υγρό δείγμα. Η υλοποίηση μίας τέτοιας δοκιμής προϋποθέτει τουλάχιστον ένα αντίσωμα με εξειδίκευση σε ένα συγκεκριμένο αντιγόνο. Η διαδικασία ανίχνευσης και αξιολόγησης της παρουσίας ενός αντιγόνου στο εκάστοτε δείγμα είναι η εξής ([40]):

Σε πρώτο στάδιο, το δείγμα με έναν άγνωστο αριθμό αντιγόνων ακινητοποιείται πάνω σε μία στερεή βάση/πλάκα. Έπειτα προστίθεται το κατάλληλο αντίσωμα ανίχνευσης για το είδος των αντιγόνων του δείγματος σχηματίζοντας ένα σύμπλεγμα. Το αντίσωμα ανίχνευσης μπορεί να είναι ομοιοπολικώς συνδεδεμένο με ένα ένζυμο ή μπορεί να ανιχνευθεί με τη σειρά του από ένα δευτερεύον αντίσωμα συνδεδεμένο με ένα ένζυμο με βιοσύζευξη. Ανάμεσα σε κάθε στάδιο, η πλάκα πλένεται με ένα ήπιο διάλυμα απορρυπαντικού ώστε να απομακρυνθούν τυχούσες πρωτεΐνες ή αντισώματα που δεν δεσμεύτηκαν. Μετά το τελικό στάδιο πλύσης, προστίθεται στην πλάκα ένα ενζυματικό υπόστρωμα για να παράγει ένα ορατό σήμα, το οποίο υποδεικνύει την ποσότητα του αντιγόνου στο δείγμα. Το σήμα αυτό είναι συνήθως η αλλαγή χρώματος του υποστρώματος ([40]).



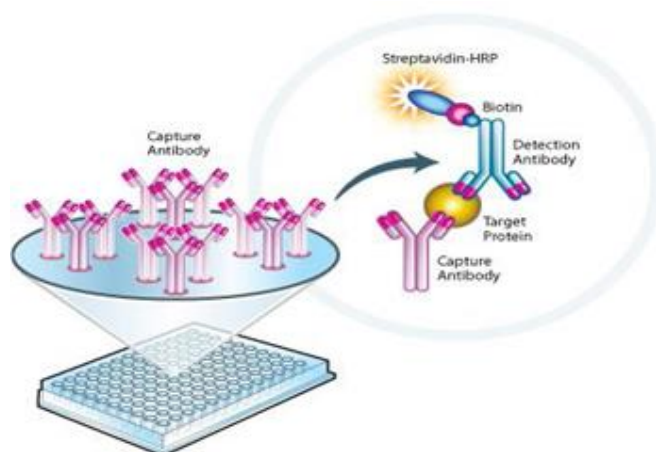
Εικόνα 62: Αναπαράσταση της μεθόδου ([40])

Ως βιοχημική δοκιμή τρυβλίου (plate-based biochemical assay) εννοείται μια τυπική δοκιμή-πείραμα (assay) στη βάση ενός τρυβλίου (plate), συνήθως 96 βοθρίων (πηγαδιών-wells), στο οποίο κάθε βοθρίο περιέχει ένα δείγμα (υγρό διάλυμα/μίγμα), του οποίου τη βιοχημική σύσταση θέλουμε να αναλύσουμε ([40]).

Η ονομασία ELISA (Enzyme-Linked Immuno Sorbent Assay), συνοψίζει μόνο την αρχική μορφή της τεχνικής. Πρακτικώς δεν αναφέρεται σε μια συγκεκριμένη τεχνική, καθώς υπάρχει μια μεγάλη ποικιλία τεχνικών που χαρακτηρίζονται με αυτό το όνομα, αλλά στην ιδέα που βασίζονται όλες αυτές οι τεχνικές ([41]).

Η βασική ιδέα την οποία όλες αυτές οι τεχνικές υλοποιούν ως ένα βαθμό είναι η εξής ([41]):

1. οι αναλύτες του δείγματος ακινητοποιούνται πάνω σε μία στερεή επιφάνεια.
2. εν συνεχεία προστίθενται, στις περισσότερες περιπτώσεις, ένα ή δύο αντισώματα (εξου και “immunosorbent”) για την ανίχνευση των ακινητοποιημένων αναλυτών, τα οποία δημιουργούν ένα σύμπλεγμα με τον αναλύτη. Το αντίσωμα αυτό καλείται Αντίσωμα Ανίχνευσης ή δευτερεύον (Detection/Secondary Antibody). Στην περίπτωση που χρησιμοποιούνται αντισώματα, οι αναλύτες είναι εξ ορισμού αντιγόνα.
3. Το αντίσωμα αυτό πρέπει με τη σειρά του να σημανθεί (labeled) μέσω μιας ανιχνεύσιμης ουσίας, η οποία παραδοσιακά ήταν κάποιο ένζυμο (εξου και το enzymelinked), ώστε να μπορεί να ανιχνευτεί το σύμπλεγμα.
4. Μεταξύ των διαφόρων βημάτων συνήθως υπάρχει πλύση των βοθρίων, ώστε να απομακρυνθούν οι διάφορες ουσίες. Ο αριθμός των πλύσεων ποικίλει ανάλογα με τα πρωτόκολλα που χρησιμοποιούνται.
5. Η ποσοτικοποίηση γίνεται μέσω της συσχέτισης του αριθμού των ανιχνευμένων συμπλεγμάτων με την αρχική συγκέντρωση του μίγματος. Η σχέση αυτή δεν είναι γνωστή εκ των προτέρων, οπότε κάθε δοκιμή χρειάζεται και την αντίστοιχη βαθμονόμηση (calibration).

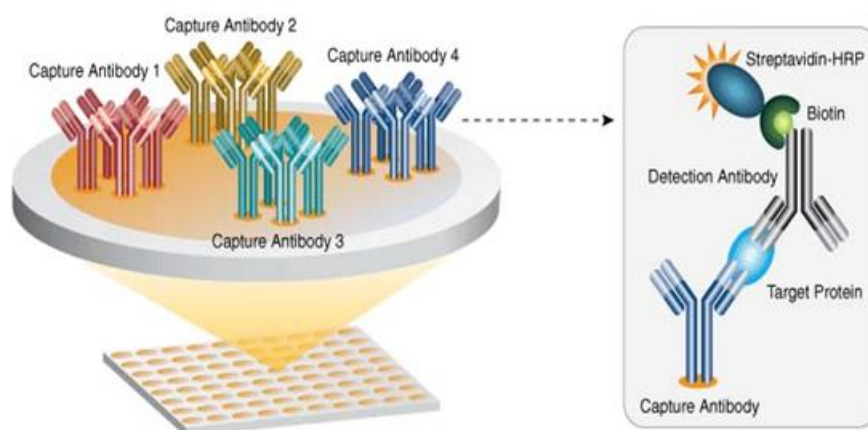


**Εικόνα 63: Αρχή λειτουργίας ELISA ([41])**

## Πολυπλεξία

Στην κλασική εκδοχή της, μια μέτρηση με την τεχνική ELISA αναγνωρίζει και ποσοτικοποιεί μόνο έναν αναλύτη ανά δείγμα. Για τις ανάγκες μιας «συστημικής» προσέγγισης, όμως, κάτι τέτοιο δεν είναι αρκετό, καθώς θα έπρεπε για κάθε μετρούμενο σε κάθε διαφορετικό υπό μελέτη ερέθισμα (stimulus), να πραγματοποιείται μια ξεχωριστή μέτρηση. Μια τέτοια διαδικασία θα ήταν εξαιρετικά χρονοβόρα και απαιτητική γι' αυτό και χρειάστηκε να αναπτυχθούν ειδικές τεχνολογίες για την εκτέλεση πολυπλεκτικών (multiplexed) μετρήσεων ([40]).

Πολυπλεκτική (multiplexed) ονομάζεται μια μέτρηση κατά την οποία μετρώνται στο ίδιο δείγμα και ταυτόχρονα πολλοί αναλύτες (τυπικά πάνω από έναν). Στο εργαστήριο η πειραματική δοκιμή ELISA που πραγματοποιήθηκε είναι η λεγόμενη Bead-based Sandwich ELISA ([40]).



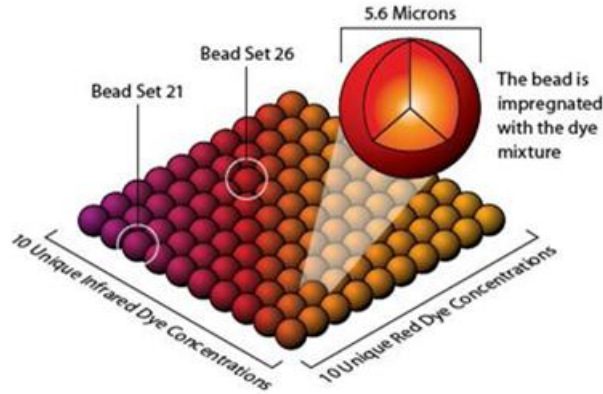
Εικόνα 64: Αρχή λειτουργίας Multiplexed ELISA ([40])

### Χαρακτηριστικά Bead-based Sandwich ELISA

Οι αναλύτες, που στην περίπτωση μας είναι κυτταροκίνες (cytokines) και φωσφοπρωτεΐνες (phosphoproteins), ακινητοποιούνται πάνω στην επιφάνεια μικροσφαιριδίων (beads). Σε αντίθεση με τη συνήθη τεχνική της παθητικής προσρόφησης (adsorption) πάνω σε ακίνητη, επίπεδη επιφάνεια, τα μικροσφαιρίδια (διαμέτρου 5,6μm) έχουν τη δυνατότητα να κινούνται μέσα στο δείγμα καθώς και να επηρεάζονται από τα μαγνητικά πεδία λόγω του σιδήρου που περιέχουν (σε ποσοστό 2%-4%) ([41]).

Τα μικροσφαιρίδια είναι χρωματισμένα με ειδική φθορίζουσα βαφή, ώστε το κάθε ένα (ή η κάθε ομάδα) να χαρακτηρίζεται από μια ξεχωριστή χρωματική ταυτότητα. Ο διαχωρισμός των μικροσφαιριδίων αποτελεί αναγκαία

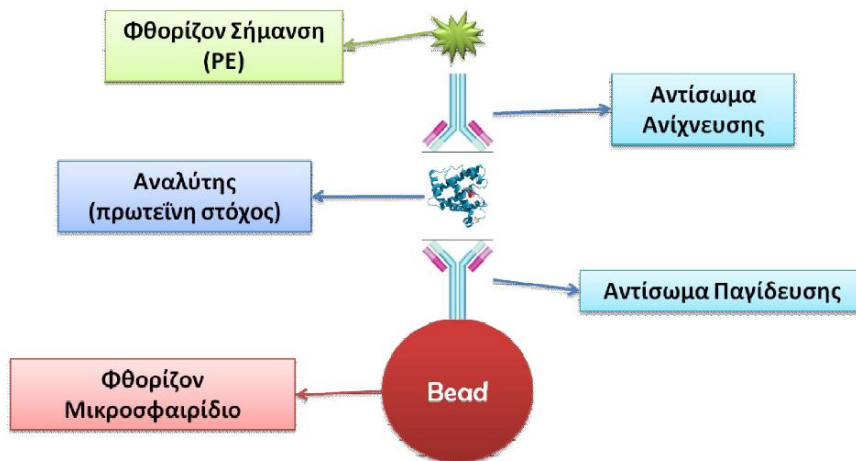
προϋπόθεση για την πολυπλεξία της μεθόδου, διότι έτσι μπορούμε να αντιστοιχίσουμε τα μικροσφαιρίδια στους διάφορους αναλυτές που μας ενδιαφέρουν ([41]).



Εικόνα 65: set micro-beads ([40])

Η προσρόφηση δεν γίνεται απευθείας στην επιφάνεια, αλλά από τα αντισώματα παγίδευσης (ή πρωτεύον αντίσωμα – Capture/Primary Antibody) με τα οποία έχουν προ-επικαλυφθεί τα μικροσφαιρίδια ([40])

Αυτό το είδος προσρόφησης λέγεται επιλεκτική (specific binding), διότι τα αντισώματα έχουν την ιδιότητα να αντιδρούν επιλεκτικά, ως ένα βαθμό, με τους αναλυτές, ενώ όταν η προσρόφηση γίνεται απ' ευθείας στην επιφάνεια τότε λέγεται μη-επιλεκτική, διότι η επιφάνεια δεν έχει την ιδιότητα να ξεχωρίζει τους αναλυτές ή οποιαδήποτε άλλη βιομοριακή ουσία μπορεί να υπάρχει στο δείγμα. Λόγω της χρήσης δύο αντισωμάτων για την ανίχνευση ενός αναλύτη, η τεχνική καλείται Sandwich ELISA ([40]).

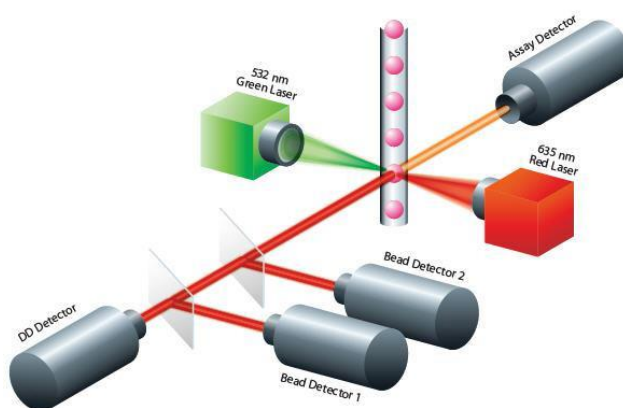


Εικόνα 66: ELISA Μικροσφαιριδίων ([40])

Η χρήση μεθόδων επιλεκτικής προσρόφησης ενδείκνυται για μια συστημική ανάλυση, γιατί ο ερευνητής έχει τη δυνατότητα να κατευθύνει το πείραμα, χρησιμοποιώντας τις αριστί γνώσεις του για το φαινόμενο, ώστε να μετρήσει μόνο τους αναλύτες που τον ενδιαφέρουν. Αντίθετα, η χρήση μη-επιλεκτικών μεθόδων συνίσταται για περιπτώσεις όπου δεν υπάρχει προγενέστερη γνώση του φαινομένου και θέλουμε να ξεκινήσουμε μια ανάλυση από το μηδέν, βασιζόμενη μόνο στα δεδομένα μας ([40]).

**Τα αντισώματα ανίχνευσης σημαίνονται με διαφορετικού χρώματος βαφή (πράσινη) από τα μικροσφαιρίδια (συνδυασμός ερυθρής και υπέρυθρης) ([40]).**

Για την μέτρηση, τα μικροσφαιρίδια αφήνονται να επωαστούν πρώτα μαζί με το δείγμα και στη συνέχεια με τα αντισώματα ανίχνευσης και κατόπιν αναγκάζονται να περάσουν μέσα από δύο συστήματα ανίχνευσης φωτεινότητας: ένα για την αναγνώριση της χρωματικής ταυτότητας του μικροσφαιριδίου, η οποία γίνεται μέσω της αναγνώρισης έντασης φωτεινότητας κάθε βαφής και του συνδυασμού τους, και ένα για την αναγνώριση της έντασης φωτεινότητας της βαφής των σημασμένων αντισωμάτων ανίχνευσης, η οποία θεωρείται ανάλογη του πλήθους τους (θεωρείται ότι κάθε αντίσωμα εκπέμπει στην ίδια ένταση) ([40]).



Εικόνα 67: Υπόδειγμα συστήματος ανίχνευσης ([40])

### Τα βασικά πλεονεκτήματα της μεθόδου

Τα βασικά πλεονεκτήματα της μεθόδου είναι δύο ([40]):

1. Η δυνατότητα καθοδήγησης των μικροσφαιριδίων (είτε μαγνητικά είτε μηχανικά) και άρα αυτοματοποίησης της διαδικασίας.
2. Η εξάλειψη των προβλημάτων της αντίδρασης υγρής-στερεάς φάσης.



Αντίδραση υγρής-στερεάς φάσης παρατηρείται σε μια τυπική δοκιμή ELISA, καθώς οι αναλύτες που βρίσκονται σε υγρή φάση πρέπει να αντιδράσουν (έμμεσα ή άμεσα) με μια στερεή επιφάνεια. Στην αντίδραση υγρής-στερεάς φάσης υπάρχουν οι εξής περιορισμοί ([40]):

- Σε αντίθεση με τα μικροσφαιρίδια που κινούνται ελεύθερα στο δείγμα, η αντίδραση δε λαμβάνει χώρα σε όλο τον όγκο του δείγματος, αλλά σε μια περιορισμένη ζώνη κοντά στη στερεή επιφάνεια (ζώνη αντίδρασης). Συνεπώς, η μετρούμενη συγκέντρωση δεν ανταποκρίνεται σε όλο το δείγμα, αλλά στη συγκέντρωση της συγκεκριμένης ζώνης. Επιπλέον, ο ρυθμός εξέλιξης του φαινομένου δεν καθορίζεται αυστηρά από το ρυθμό της αντίδρασης, αλλά και από το ρυθμό προσέλευσης των αντιδρώντων στη ζώνη αντίδρασης.
- Παρατηρούνται στερικά φαινόμενα (stericeffects), δηλαδή περιορισμοί λόγω του όγκου των αντιδρώντων και του κατάλληλου προσανατολισμού που πρέπει να έχουν μεταξύ τους (μείωση των βαθμών ελευθερίας στις κινήσεις των αντιδρώντων). Με την χρήση μικροσφαιριδίων μπορούμε να αυξήσουμε την συνολική επιφάνεια κρατώντας τον λόγο επιφάνειας όγκου σταθερό και έτσι να περιορίσουμε σημαντικά αυτό το πρόβλημα.

Πέραν των βασικών θεωρητικών πλεονεκτημάτων, στην πράξη παρατηρούνται και τα εξής πλεονεκτήματα ([40]):

- Ευκολότερη πολυπλεξία, καθώς αρκεί να προστεθούν στο μίγμα μικροσφαιρίδια διαφορετικών χρωμάτων, ενώ στην επίπεδη επιφάνεια πρέπει να χωριστεί η επιφάνεια σε περιοχές μειώνοντας έτσι την συνολική επιφάνεια αντίδρασης.
- Μεγαλύτερη οικονομία σε χρόνο και χρήμα για τον ίδιο όγκο δεδομένων.

Από την περιγραφείσα μέθοδο προκύπτουν τα πειραματικά δεδομένα, τα οποία χρησιμοποιήθηκαν για τη μοντελοποίηση που περιγράφεται στην παρούσα διπλωματική.

## **2.7 Βελτιστοποίηση μονοπατιών σηματοδότησης με υπολογιστικές μεθόδους**

### **2.7.1 Εισαγωγικά**

Όπως έχει, ήδη, αναφερθεί και παραπάνω πολλάκις, τα κύτταρα είναι πολύπλοκες λειτουργικές μονάδες. Η μεταγωγή σήματος αναφέρεται στον υποκείμενο μηχανισμό, ο οποίος ρυθμίζει τη λειτουργία των κυττάρων και απεικονίζεται συνήθως στα μονοπάτια σηματοδότησης. Κάθε τύπος κυττάρων έχει συγκεκριμένους μηχανισμούς μεταγωγής σήματος και διάφορες ασθένειες προκύπτουν από μεταλλάξεις στα μονοπάτια σηματοδότησης. Μικρό-μόρια αναστολείς έχουν αναδειχθεί ως νέες φαρμακευτικές παρεμβάσεις που στοχεύουν να μπλοκάρουν ορισμένα μονοπάτια, σε μία προσπάθεια να αναστραφεί η ανωμαλία του φαινοτύπου των ασθενών κυττάρων. Παρόλο που ενώσεις έχουν σχεδιαστεί για να αναστέλλουν διάφορα μόρια, λίγα είναι γνωστά για τον τρόπο δράσης των σε ένα «λειτουργικό» δίκτυο σηματοδότησης. Στην επιστημονική προσέγγιση που θα περιγραφεί παρακάτω, συνδυάζονται νέες υψηλής απόδοσης μετρήσεις πρωτεϊνών-σημάτων και σοφιστικές υπολογιστικές τεχνικές για την αξιολόγηση των επιδράσεων των φαρμάκων στα κύτταρα. Η προσέγγιση αυτή, αποτελείται από δύο βήματα: δημιουργία μονοπατιών που προσομοιώνουν την κυτταρική λειτουργία και αναγνωρίζουν τις επαγόμενες από τα φάρμακα μεταβολές των εν λόγω μονοπατιών. Κατανοώντας τις επιπτώσεις των φαρμάκων στα φυσιολογικά και τα ασθενή κύτταρα, μπορούμε να αποκτήσουμε σημαντικές πληροφορίες για την ανάλυση των κλινικών αποτελεσμάτων, με σκοπό τη βελτίωση της αποτελεσματικότητας και της ασφάλειας των φαρμάκων ([33]).

### **2.7.2 Μέθοδος Ακέραιου Γραμμικού Προγραμματισμού για τη μοντελοποίηση μονοπατιών μεταγωγής σημάτων πρωτεϊνών (ILP)**

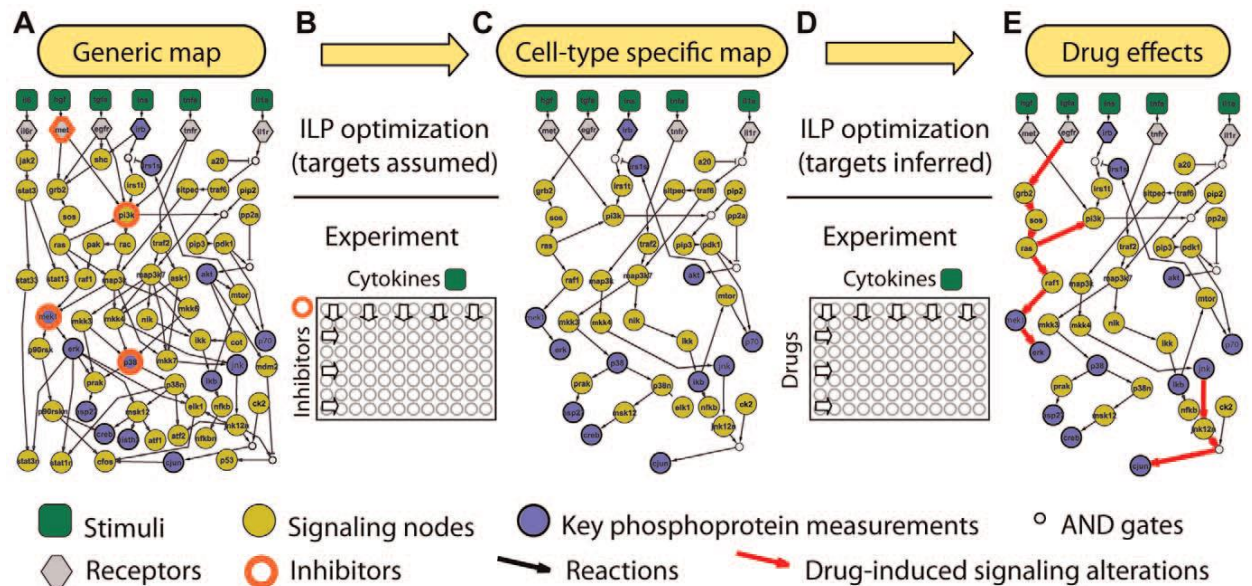
Η στόχο-κεντρική ανακάλυψη φαρμάκων είναι ο τομέας κυρίαρχης εστίασης, σήμερα, των φαρμακευτικών βιομηχανιών. Ο πρωταρχικός σκοπός τους, είναι η επιλεκτική στοχοποίηση πρωτεϊνών μέσα στα ασθενή κύτταρα, με σκοπό τη βελτίωση ενός ανεπιθύμητου φαινοτύπου, π.χ. ανεξέλεγκτος πολλαπλασιασμός κυττάρων ή ανεξέλεγκτη απελευθέρωση κυτοκινών (cytokines). Ίδανικά, άλλα μονοπάτια μέσα στα ασθενή κύτταρα, καθώς και άλλοι φαινότυποι άλλων κυτταρικών τύπων, πρέπει να παραμείνουν ανεπηρέαστοι από τη θεραπευτική προσέγγιση. Παρόλα αυτά, παρόλη την πληθώρα των νέων πιθανών στόχων που προκύπτουν από την ακολουθία του ανθρώπινου γωμιώματος, μόνο λίγοι αποδείχτηκαν αποτελεσματικοί στην κλινική. Ένας σημαντικός περιορισμός είναι η ανικανότητα για την κατανόηση των μηχανισμών ή της ενεργοποίησης των φαρμάκων, είτε λόγω στην πολύπλοκη μεταγωγή σημάτων των κυτταρικών δικτύων, είτε λόγω του πολύπλοκου προφίλ της επιλεκτικότητας και της δραστικότητας των φαρμάκων ([31]).

Η εύρεση των στόχων των φαρμάκων βασίζεται παραδοσιακά σε υψηλής απόδοσης αναλύσεις που χρησιμοποιούν ανά-συνδυασμένα ένζυμα ή πρωτεϊνικά θραύσματα. Ο κύριος στόχος είναι ο χαρακτηρισμός των βιοχημικών δραστηριοτήτων των φαρμάκων και η απεικόνισή των σε χάρτες φαρμακευτικών αλληλεπιδράσεων. Στις περισσότερες περιπτώσεις, μιας και οι στόχοι είναι γνωστοί, η επίδραση των μονοπατιών σηματοδότησης επικυρώνεται από τη μέτρηση της αποτελεσματικότητας των φαρμάκων να αναστέλλουν την ενεργοποίηση (συνήθως μετρούμενη ως επίπεδο φωσφορυλίωσης) της κατάντη πρωτεΐνης. Παρόλα αυτά, πέραν αυτής της μέτρησης, λίγα είναι γνωστά στο πώς το υπόλοιπο δίκτυο σηματοδότησης επηρεάζεται ([31]).

Για την αντιμετώπιση των επιδράσεων των φαρμάκων σε πιο φυσιολογικές συνθήκες, νέα γονιδιωμικά και πρωτεομικά εργαλεία έχουν αναπτυχθεί. Στο πεδίο των γονιδιωμάτων, έχει αναπτυχθεί η μακράς κλίμακας ανάλυση του mRNA ενισχυόμενη από τις υπολογιστικές προσεγγίσεις για την από-συνέλιξη των φαρμάκων-στόχων. Παρόλο των ολιστικών πλεονεκτημάτων που οι γονιδιακές προσεγγίσεις έχουν να προσφέρουν, η πρωτεϊνικό-κεντρική προσέγγιση βρίσκεται ένα βήμα πιο κοντά στη λειτουργία του κυττάρου. Η μελέτη, λοιπόν, σε πρωτεϊνικό επίπεδο μπορεί να προσφέρει περισσότερες πληροφορίες από ότι η ανάλυση του γενετικού υλικού ([31]).

Στην ενότητα αυτή, περιγράφεται μία σημαντικά διαφορετική προσέγγιση για την αναγνώριση των επιδράσεων των φαρμάκων, τα οποία αξιολογούνται από τις μεταλλάξεις που προκαλούν στα μονοπάτια σηματοδότησης. Αντί για την αναγνώριση των δεσμευτικών συμμετοχών, παρακολουθούνται οι μεταλλάξεις των μονοπατιών ακολουθώντας τα γεγονότα φωσφορυλίωσης με διάφορες θεραπείες με κυτοκίνες. Το διάγραμμα ροής της εν λόγω μεθόδου παρουσιάζεται στην παρακάτω εικόνα. Από πειραματικής απόψεως, χρησιμοποιώντας πολύ-πλεκτικές μεθόδους με μικροσφαιρίδια, μπορούμε να μετρήσουμε δεκατρία γεγονότα φωσφορυλίωσης υπό την επίδραση πενήντα διαφορετικών συνθηκών, οι οποίες παράγονται από τη συνδυαστική μεταχείριση ερεθισμάτων και επιλεκτικών αναστολέων. Βάσει της απόκρισης σηματοδότησης και ενός συνδυασμού πιθανών αντιδράσεων (γενετικά μονοπάτια), δημιουργείται ένα συγκεκριμένο μονοπάτι κυτταρικού τύπου χρησιμοποιώντας μία αποδοτική φόρμουλα βελτιστοποίησης, γνωστή και ως **Ακέραιος Γραμμικός Προγραμματισμός (Integer Linear Programming)**. Αυτή η προσέγγιση δημιουργείται σύμφωνα με τη μοντελοποίηση κατά Boole, όπως περιγράφηκε εκτενώς σε παραπάνω ενότητα. Η ILP μοντελοποίηση λύνεται με τη χρήση τυποποιημένων εμπορικών πακέτων λογισμικού με εγγυημένη αποδοτικότητα (εντός μίας ορισμένης από το χρήστη μικρής, αριθμητικής ανοχής). Για την αξιολόγηση των επιδράσεων των φαρμάκων, τα κύτταρα υπόκεινται στο ίδιο ερέθισμα υπό την παρουσία διάφορων φαρμάκων και ανιχνεύουμε τις μεταλλάξεις των ιδίων γεγονότων φωσφορυλίωσης. Μετά, επαναεφαρμόζεται το μοντέλο ILP, χωρίς κάποια εκ των προτέρων υπόθεση του φάρμακου-στόχου και παρακολουθούνται οι αλλαγές στην τοπολογία του μονοπατιού με ή χωρίς την παρουσία φαρμάκων ([31]).

Σε αντίθεση με άλλες αναπτυσσόμενες τεχνικές, η μέθοδος αυτή βασίζεται στην πραγματική επίδραση των συμβάντων φωσφορυλίωσης, προσεκτικά εξαπλωμένη στο δίκτυο σηματοδότησης. Από υπολογιστικής άποψης, η προσέγγιση με ILP παρουσιάζει μία πιο γρήγορη και πιο αποδοτική βελτιστοποίηση μονοπατιών σε σχέση με τους σύγχρονους αλγορίθμους και μπορεί να αναγνωρίσει τις βασικές επιδράσεις των φαρμάκων, καθώς και άγνωστες επιδράσεις στους τομείς των μονοπατιών που περιορίζονται μεταξύ των ενεργών υποδοχέων και των μετρούμενων φωσφορυλιωμένων πρωτεϊνών ([31]).



**Εικόνα 68:** Πειραματικό και υπολογιστικό διάγραμμα ροής για την αξιολόγηση της επίδρασης των φαρμάκων ([31])

(A) Μία γενική απεικόνιση κατά *Boole* συναρμολογείται από μία διεθνή βάση δεδομένων μονοπατιών και συμπεριλαμβάνει ερεθίσματα (πράσινα τετράγωνα), μετρούμενες φωσφοπρωτεΐνες (μοβ κύκλοι) και γειτονικές πρωτεΐνες (κίτρινοι κύκλοι). (B) Τα κύτταρα αντιμετωπίζονται με ένα συνδυασμό κυτοκινών και επιλεκτικών αναστολέων (κόκκινοι κύκλοι) γνωστών επιδράσεων και μία μέθοδος Ακέραιου Γραμμικού Προγραμματισμού (ILP) χρησιμοποιείται για ταιριάζει τα δεδομένα στα μονοπάτια κατά *Boole*. (C) Δημιουργία ενός συγκεκριμένου μονοπατιού κυττάρων. (D) Τα κύτταρα αντιμετωπίζονται με ένα συνδυασμό κυτοκινών και φαρμάκων – η επίδρασή των παραμένει άγνωστη – και η μέθοδος ILP χρησιμοποιείται για δεύτερη φορά για να ταιριάζει τα δεδομένα φωσφορυλίωσης. (E) Μεταλλάξεις συγκεκριμένης κυτταρικής τοπολογίας αποκαλύπτει επίδραση φαρμάκων (κόκκινα βέλη ([31])).

## Υπολογιστική διαδικασία: διαμόρφωση ILP

---

Στην παράγραφο αυτή περιγράφεται το πώς το μοντέλο Boole (όπως περιγράφηκε και στην προηγούμενη ενότητα), μπορεί να αναδιατυπωθεί σαν ένα μοντέλο ILP. Ένα μονοπάτι καθορίζεται σαν ένα σύνολο αντιδράσεων  $i=1, \dots, n_r$  και ειδών  $j=1, \dots, n_s$ . Κάθε αντίδραση έχει τρία αντίστοιχα σύνολα δεικτών, δηλαδή το σύνολο δεικτών των μορίων σηματοδότησης  $R_i$ , των αναστολέων  $I_i$  και των «παραγώγων»  $P_i$  (σαν «παραγώγο» μπορεί να ορισθεί, επίσης, το επίπεδο της φωσφορυλίωσης των πρωτεϊνών). Αυτά τα σύνολα, είναι όλα υποσύνολα του συνόλου των ειδών των δεικτών ( $R_i, I_i, P_i \subseteq \{1, \dots, n_s\}$ ). Μία αντίδραση διεξάγεται, εάν και μόνο εάν, παρευρίσκονται όλα τα αντιδρώντα και κανένας αναστολέας. Εάν μία αντίδραση διεξάγεται, σχηματίζονται όλα τα παράγωγα. Ας σημειωθεί πως αντιδράσεις χωρίς παράγωγα, όπως και αντιδράσεις χωρίς αντιδρώντα και αναστολείς, δεν θα μας απασχολήσουν εδώ ([31]).

Αν και τυπικά το σύνολο των ειδών είναι γνωστό, το σύνολο των αντιδράσεων δεν είναι γνωστό. Αντίθετα, μόνο ένα υπερσύνολο πιθανών αντιδράσεων είναι αξιωματικό. Ο στόχος της προτεινόμενης διαμόρφωσης είναι η εύρεση ενός βέλτιστου (κατά κάποια έννοια) συνόλου αντιδράσεων από ένα τέτοιο υπερσύνολο. Σε αυτό το βαθμό, δυαδικές μεταβλητές  $y_i$  εισάγονται, δείχνοντας εάν μία αντίδραση είναι πιθανή ή όχι ( $y_i = 0$  η σύνδεση δεν είναι παρούσα,  $y_i = 1$  η σύνδεση είναι παρούσα) ([31]).

Ένα σύνολο πειραμάτων πραγματοποιείται, αναπροσαρμόζονται από τον εκθέτη  $k = 1, \dots, n_e$ . Σε κάθε πείραμα ένα υποσύνολο ειδών εισάγεται στο σύστημα, και ένα άλλο υποσύνολο εξάγεται από το σύστημα. Αυτά συνοψίζονται από τα σύνολα δεικτών  $M^{k,1}$  και  $M^{k,0}$  αντίστοιχα (δύο για κάθε πείραμα). Στην παρούσα προτεινόμενη μέθοδο, εισάγονται σταθερές για όλα τα εν λόγω είδη, αντίστοιχα  $x_j^k = 1$  και  $x_j^k = 0$ . Στα ακόλουθα θα υποθεθεί πως αυτά τα είδη δεν εμφανίζονται σαν παράγωγα σε καμία αντίδραση; αυτή η υπόθεση δεν περιορίζεται, καθώς στα πειράματα που εκτελούνται μόνο εξωκυτταρικά είδη και αναστολείς χρησιμοποιούνται. Στα πειράματα μετράται ένα τρίτο υποσύνολο των ειδών (σύνολο δεικτών  $M^{k,2}$ ) και για τα εναπομείναντα είδη καμία πληροφορία δεν είναι διαθέσιμη. Στην προτεινόμενη μέθοδο, για κάθε ένα από τα πειράματα και για κάθε ένα από αυτά τα είδη, εισάγεται μία δυαδική μεταβλητή απόφασης  $x_j^k \in \{0,1\}$ , η οποία δείχνει εάν το είδος  $j$  είναι παρόν ( $x_j^k = 1$ ) ή όχι ( $x_j^k = 0$ ) στο πείραμα  $k$  σύμφωνα με τις προβλέψεις του μοντέλου. Αποδεικνύεται ότι η απουσία βρόχων,  $x_j^k \in [0,1]$  μπορεί να χρησιμοποιηθεί για είδη, τα οποία δεν είναι είδη εισόδου. Αυτό το χαρακτηριστικό έχει και κάποια υπολογιστικά πλεονεκτήματα ([31]).

Η τελευταία ομάδα μεταβλητών  $z_i^k$  που εισάγεται δείχνει εάν η αντίδραση  $i$  θα διεξαχθεί ( $z_i^k = 1$ ) ή όχι ( $z_i^k = 0$ ) στο πείραμα  $k$  σύμφωνα με τις προβλέψεις του μοντέλου. Έχει αποδειχθεί πως μία πραγματική μεταβλητή  $z_i^k \in [0,1]$

μπορεί να χρησιμοποιηθεί ισοδύναμα. Αυτή η αναδιατύπωση έχει κάποιες υπολογιστικά πλεονεκτήματα ([31]).

Για την περίπτωση που ένα είδος μετράται, η μέτρηση καθορίζεται ως  $x_j^{k,m}$ . Για τις μετρήσεις κατά Boole  $x_j^{k,m} \in \{0,1\}$ ; αλλιώς  $x_j^{k,m} \in [0,1]$  (υποθέτοντας μία κλίμακα, όπως προαναφέρθηκε). Η πρωταρχική αντικειμενική συνάρτηση σχηματίζεται με στόχο την ελαχιστοποίηση του σταθμισμένου σφάλματος μεταξύ των προβλέψεων του μοντέλου και των μετρήσεων  $\sum_{j,k} a_j^k |x_j^k - x_j^{k,m}|$ . Η απόλυτη τιμή αναδιατυπώνεται σαν  $x_j^{k,m} + (1 - 2x_j^{k,m})x_j^k$ . Μπορεί εύκολα να διατυπωθεί για τις δυαδικές μεταβλητές  $x_j^k$  και  $x_j^{k,m} \in \{0,1\}$ , αυτή η αναδιατύπωση είναι έγκυρη ([31]).

$$1. \quad x_j^k = 0$$

$$x_j^{k,m} + (1 - 2x_j^{k,m})x_j^k = x_j^{k,m} + (1 - 2x_j^{k,m})0 = x_j^{k,m} = |x_j^{k,m}| = |x_j^{k,m} - x_j^k|$$

$$2. \quad x_j^k = 1$$

$$x_j^{k,m} + (1 - 2x_j^{k,m})x_j^k = x_j^{k,m} + (1 - 2x_j^{k,m})1 = 1 - x_j^{k,m} = |1 - x_j^{k,m}| = |x_j^k - x_j^{k,m}|$$

Ας σημειωθεί, επίσης, ότι εναλλακτικές νόρμες, τέτοιες όπως τα σφάλματα ελαχίστων τετραγώνων, μπορούν να χρησιμοποιηθούν. Το προκύπτον πρόβλημα βελτιστοποίησης θα εξακολουθούσε να είναι ένα πρόβλημα ILP, μιας και η αντικειμενική συνάρτηση συμπεριλαμβάνει μόνο ακέραιες μεταβλητές. Για παράδειγμα, για την αντικειμενική συνάρτηση ελαχίστων τετραγώνων, η ακόλουθη γραμμική αναδιατύπωση είναι έγκυρη ([31]):

$$(x_j^k - x_j^{k,m})^2 = (x_j^k)^2 - (2x_j^k x_j^{k,m}) + (x_j^{k,m})^2 = (x_j^k) - (2x_j^k x_j^{k,m}) + (x_j^{k,m})^2$$

Ο δευτερεύων στόχος είναι η ελαχιστοποίηση το σταθμισμένο αριθμό των πιθανών αντιδράσεων  $\sum_i \beta_i y_i$ . Τυπικά, στη βελτιστοποίηση πολλών στόχων, εισάγονται οι έννοιες του *μετώπου-Pareto* ή αλλιώς *μη-κατώτερη λύση*, δηλαδή ένα σύνολο τιμών μεταβλητών απόφασης, τέτοιες ώστε εάν κάποιος προσπαθήσει να βελτιώσει ένα στόχο, ένας άλλος θα υποβαθμιστεί. Το σύνολο των σημείων Pareto σχηματίζει τη κατά Pareto βέλτιστη καμπύλη.

Εδώ, παρόλα αυτά, ο πρωταρχικός στόχος θεωρείται πολύ πιο σημαντικός από το δευτερεύων. Ως εκ τούτου, λαμβάνεται ένα μοναδικό σημείο μετώπου-Pareto, ελαχιστοποιώντας πρώτα τον πρωταρχικό στόχο και μετά το δευτερεύων στόχο, με την απαίτηση πως οι προηγούμενοι (πιο σημαντικοί) δεν επιδεινώνονται ([5]).

Το προτεινόμενο μοντέλο ILP μπορούν να συνοψιστούν στο παρακάτω σύστημα εξισώσεων ([31]):

$$\min_{X,y,Z} \sum_{k=1}^{n_e} \sum_{j \in M^{k,2}} a_j^k (x_j^{k,m} + (1 - 2x_j^{k,m}) x_j^k); \quad \sum_{i=1}^{n_r} \beta_i y_i \quad (1)$$

$$s.t. \quad \sum_{i=1}^{n_r} a_i^l y_i \leq b^l, \quad l = 1, \dots, n_c \quad (2)$$

$$z_i^k \leq y_i, \quad i = 1, \dots, n_r, \quad k = 1, \dots, n_e \quad (3)$$

$$z_i^k \leq x_j^k, \quad i = 1, \dots, n_r, \quad k = 1, \dots, n_e, j \in R_i \quad (4)$$

$$z_i^k \leq 1 - x_j^k, \quad i = 1, \dots, n_r, \quad k = 1, \dots, n_e, j \in I_i \quad (5)$$

$$z_i^k \geq y_i + \sum_{j \in R_i} (x_j^k - 1) - \sum_{j \in I_i} (x_j^k), \quad i = 1, \dots, n_r, \quad k = 1, \dots, n_e \quad (6)$$

$$x_j^k \geq z_i^k, \quad i = 1, \dots, n_r, \quad k = 1, \dots, n_e, \quad j \in P_i \quad (7)$$

$$x_j^k \leq \sum_{i=1, \dots, n_r: j \in P_i} z_i^k, \quad j = 1, \dots, n_s, \quad k = 1, \dots, n_e \quad (8)$$

$$x_j^k = 0, \quad k = 1, \dots, n_e, \quad j \in M^{k,0} \quad (9)$$

$$x_j^k = 1, \quad k = 1, \dots, n_e, \quad j \in M^{k,1} \quad (10)$$

$$X \in \{0,1\}^{n_e \times n_s}, \quad y \in \{0,1\}^{n_r}, \quad Z \in \{0,1\}^{n_e \times n_r} \quad (11)$$

όπου οι αντικειμενικές συναρτήσεις διαχωρίζονται από ερωτηματικό. Να σημειώσουμε πως για τα στοιχεία των μητρώων X και Z, ο δείκτης της σειράς (πείραμα) υποδεικνύεται σαν εκθέτης και ο δείκτης της στήλης (είδη και αντιδράσεις αντίστοιχα) υποδεικνύεται σαν δείκτης ([31]).

Στο σύστημα εξισώσεων (1) - (11) για τα διαχειριζόμενα είδη, μαζί με τις μεταβλητές απόφασης εισάγονται και οι περιορισμοί (9) και (10). Αυτό απλοποιεί τη συντομογραφία. Κατά την υλοποίηση, αυτές οι μεταβλητές αντικαθίστανται από σταθερές. Εναλλακτικά, ο προ-επεξεργαστής του λύτη βελτιστοποίησης μπορεί να χρησιμοποιηθεί για να αποκλείσει αυτές τις τετριμμένες μεταβλητές ([31]).

Στα παρακάτω, η αιτιολογία για τη μέθοδο είναι δεδομένη. Το πρώτο σύνολο περιορισμών, δηλαδή η εξίσωση (2) επιτρέπει στη μοντελοποίηση τον περιορισμό των συνδυασμών της θεωρούμενης συνδεσιμότητας. Για παράδειγμα, ας υποθέσουμε πως δύο αντιδρώντα  $R_1$ ,  $R_2$  σχηματίζουν ένα παράγωγο P, αλλά παρόλα αυτά δεν είναι γνωστό εάν απαιτούνται και τα δύο αντιδρώντα (AND) ή ένα από τα δύο (OR). Η πρόταση αυτή μπορεί να μοντελοποιηθεί με τρεις πιθανές αντιδράσεις ([31]):



$$r_1 : R_1 + R_2 \rightarrow P$$

$$r_2 : R_1 \rightarrow P$$

$$r_3 : R_2 \rightarrow P$$

Με τον πρόσθετο περιορισμό ότι η αντίδραση  $r_1$  αποκλείει τις αντιδράσεις  $r_2$  και  $r_3$ , οι οποίες μπορούν να μοντελοποιηθούν σαν δύο γραμμικές ανισότητες ([31]):

$$y_{r_1} + y_{r_2} \leq 1$$

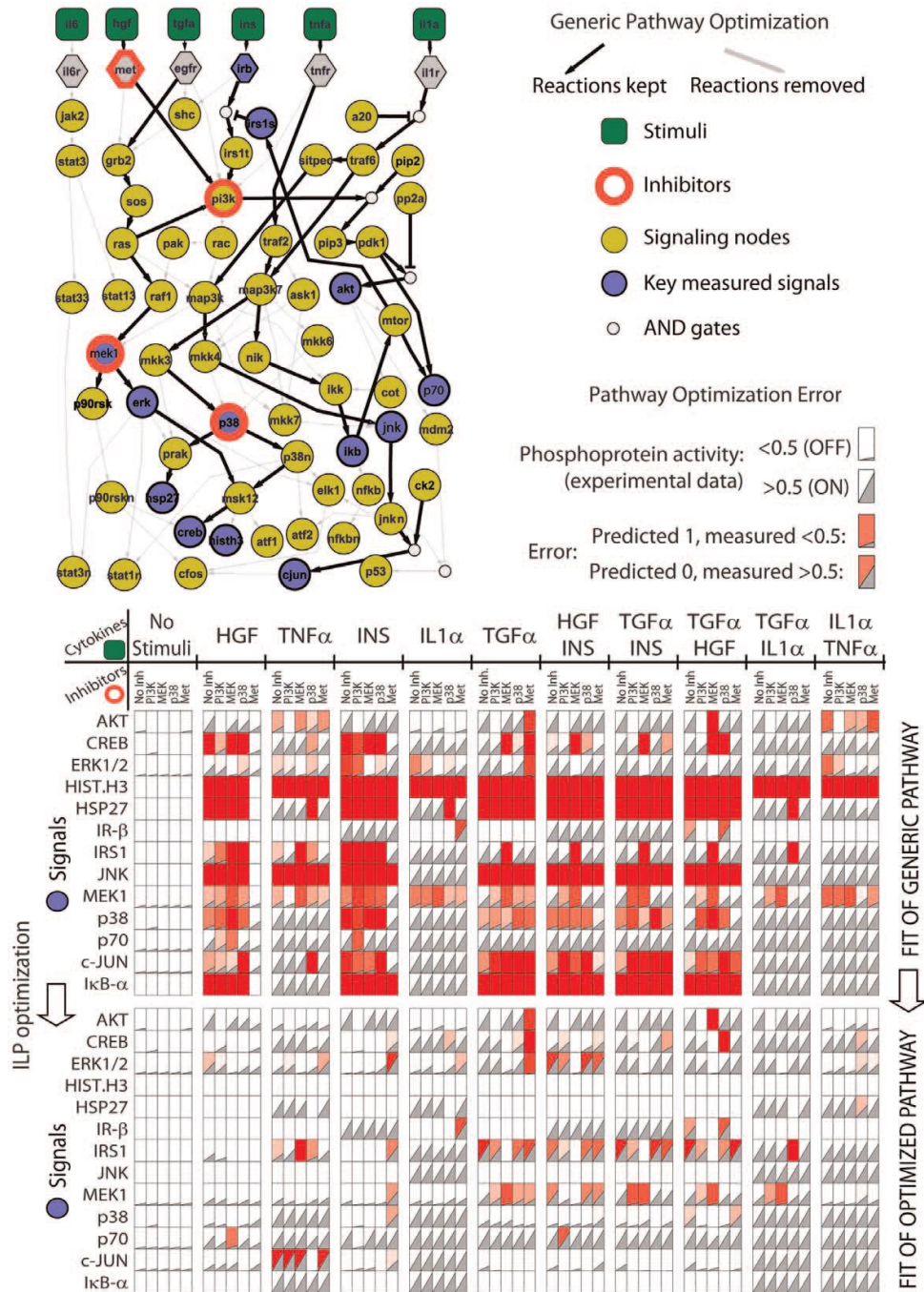
$$y_{r_1} + y_{r_3} \leq 1$$

Οι τρεις περιορισμοί υποδεικνύουν, ότι μία αντίδραση μπορεί μόνο να διεξαχθεί, εάν είναι πιθανή ( $y_i=1$ ). Αυτό μπορεί να παρατηρηθεί εύκολα, αφού  $y_i=0$ , δίνει  $z_i^k < 0$  και μαζί με τη συνθήκη  $z_i^k \in \{0,1\}$ , παίρνουμε  $z_i^k = 0$ . Παρόμοια, οι περιορισμοί (4) και (5) διασφαλίζουν αντίστοιχα, ότι μία αντίδραση μπορεί μόνο να διεξαχθεί, εάν είναι παρόντα όλα τα αντιδρώντα και κανένας αναστολέας. Εάν για παράδειγμα, ένα αντιδρών λείπει, επιβάλλεται η συνθήκη  $z_i^k = 0$  και οι άλλοι περιορισμοί είναι περιττοί. Από την άλλη μεριά, οι περιορισμοί της σχέσης (6) επιβάλλουν ότι εάν μία αντίδραση είναι δυνατή, όλα τα αντιδρώντα είναι παρόντα και δεν υπάρχει κανένας αναστολέας, τότε η αντίδραση θα διεξαχθεί ( $z_i^k = 1$ ) ([31]).

Οι περιορισμοί της σχέσης (7) διασφαλίζουν ότι ένα είδος θα σχηματισθεί, εάν συμβαίνει κάποια αντίδραση στην οποία υπάρχει κάποιο παράγωγο. Ας σημειωθεί πως πολλαπλές αντιδράσεις μπορούν να δώσουν το ίδιο είδος; μαθηματικά αυτό οδηγεί σε περιττούς περιορισμούς. Σε αντίθεση, οι περιορισμοί της σχέσης (8) οδηγούν στο ότι ένα είδος δεν θα είναι παρόν, εάν όλες οι αντιδράσεις στις οποίες δεν εμφανίζεται ως παράγωγο, δεν συμβαίνουν. Υπενθυμίζουμε πως τα διαχειριζόμενα είδη δεν θεωρούνται ως παράγωγα στις αντιδράσεις. Ας σημειωθεί, επίσης, ότι θα ήταν δυνατό να συνδυαστούν οι περιορισμοί της σχέσης (7) σε ένα μοναδικό περιορισμό για κάθε είδος, δηλαδή ([31]):

$$x_j^k \geq \sum_{i=1, \dots, n_r; j \in P_i} z_i^k / \sum_{i=1, \dots, n_r; j \in P_i} 1, \quad j=1, \dots, n_s, \quad k=1, \dots, n_e,$$

αλλά αυτό θα οδηγούσε σε ασθενέστερες συγκλίσεις (LP-relaxations). Ακόμα, η αναδιατύπωση του όρου  $x_j^k$  στα όρια  $[0, 1]$  δεν θα ήταν πλέον ακριβής.



**Εικόνα 69: Συγκεκριμένη κυτταρική τοπολογία με χρήση Ακέραιου Γραμμικού Προγραμματισμού ([31])**

Ο αλγόριθμος Ακέραιου Γραμμικού Προγραμματισμού (ILP) χρησιμοποιεί ένα υποσύνολο αντιδράσεων συμβολισμένες με μαύρα και γκρι βέλη σε ένα γενικό μονοπάτι για τη δημιουργία ενός HerG2 χάρτη μονοπατιού (μαύρα βέλη σε διάγραμμα μονοπατιού). Τα γκρι τρίγωνα δείχνουν τα επίπεδα ενεργοποίησης φωσφοπρωτεϊνών υπό κάποιο ερέθισμα (στήλες σε άνω και κάτω πάνελ) και κάποιους αναστολείς (υποστήλες στο επάνω και κάτω πάνελ). Το κόκκινο φόντο υποδεικνύει ένα σφάλμα μεταξύ πειραματικών και συναχθέντων αποκρίσεων μονοπατιών. Η γενική τοπολογία μπορεί δύσκολα να αντιπροσωπεύει τις HerG2 αποκρίσεις σηματοδότησης (κόκκινο φόντο στην κορυφή του πίνακα) και η

βελτιστοποίηση των μονοπατιών είναι κρίσιμη για τη λήψη της τοπολογίας των μονοπατιών που αιχμαλωτίζει τη HerG2 λειτουργία (ξεθωριασμένο κόκκινο φόντο στο κάτω πάνελ ([31]).

### **Υπολογιστική διαδικασία: προσαρμογή σφάλματος**

---

Υπολογίζεται το ποσοστιαίο σφάλμα από τη σχέση ([30]):

$$Error = \sum_{j=1}^{n_s} |x_j^{k,m} - x_j^k| / n_{s,m} \cdot 100\%$$

Ας σημειωθεί πως για τις δυαδικές μεταβλητές  $x_j^k$  και  $x_j^{k,m} \in [0,1]$  το ποσοστιαίο σφάλμα δεν μπορεί να είναι 0%, ακόμα και όταν δεν υπάρχει αναντιστοιχία μεταξύ του μοντέλου και των πειραματικών δεδομένων. Ένας άλλος τρόπος να ποσοτικοποιήσουμε τον καθορισμό του σφάλματος μετρώντας τον αριθμό των αναντιστοιχιών: οι περιπτώσεις, όπου η στρογγυλοποιημένη πειραματική τιμή (0 ή 1) δεν είναι η ίδια με την υπολογίσιμη τιμή, ή με άλλα λόγια, όταν το υπολογιστικό – πειραματικό σφάλμα είναι παραπάνω από 0,5 ([30]).

#### **2.7.3 Μέθοδος Μη-Γραμμικού Προγραμματισμού για την ποσοτική μοντελοποίηση μονοπατιών μεταγωγής σημάτων πρωτεϊνών**

Στην ενότητα αυτή παρουσιάζεται μία μέθοδος Μη-Γραμμικού Προγραμματισμού (Non Linear Programming – NLP) για την ποσοτική μοντελοποίηση μονοπατιών μεταγωγής σημάτων, βάσει των δεδομένων σηματοδότησης. Χρησιμοποιείται μία προσέγγιση ασαφούς λογικής για τη μοντελοποίηση των μηχανισμών μεταγωγής σήματος, η οποία συνδυάστηκε με μία μέθοδο NLP βελτιστοποίησης. Η προτεινόμενη μέθοδος επιτρέπει τη γρήγορη, σχετικά, βελτιστοποίηση των μονοπατιών σηματοδότησης σε υψηλής απόδοσης δεδομένων σηματοδότησης σε ένα ποσοτικό πλαίσιο ([28]).

Η μέθοδος NLP παρουσιάζει διάφορα πλεονεκτήματα και περιορισμούς στη βελτιστοποίηση μονοπατιών. Από την αρνητική σκοπιά, είναι ξεκάθαρο ότι η εξακρίβωση της παρουσίας ή της απουσίας κάποιας αντίδρασης στη γενική τοπολογία ή η μοναδική αναγνώριση των παραμέτρων της δεν είναι πιθανόν λόγω του σχετικά μικρού συνόλου δεδομένων. Αυτό συνεπάγεται ότι το πρόβλημα βελτιστοποίησης ενσωματώνει περισσότερες παραμέτρους από όσες είναι δυνατόν να περιορισθούν. Η μέθοδος που χρησιμοποιείται, βελτιστοποιώντας τις τιμές των παραμέτρων μοντέλου, ελαχιστοποιεί την προβλεπόμενη αναντιστοιχία των μετρήσεων. Ουσιαστικά, αντιδράσεις μπορούν να αφαιρεθούν ρυθμίζοντας την παράμετρο του κέρδους των αντίστοιχων αντιδράσεων στο μηδέν, παρόλα αυτά, δεν υπάρχει υποστήριξη για νέες συνδέσεις. Ένας άλλος περιορισμός είναι η μέτρηση ενιαίου χρόνου της δραστηριότητας σηματοδότησης. Οι μετρήσεις ενιαίου χρόνου μας

εμποδίζουν, επίσης, από την καταγραφή της δυναμικής των αντιδράσεων σηματοδότησης. Αν και μία δυναμική αναπαράσταση είναι πιο κοντά στην πραγματικότητα και δυνητικά μπορεί να αντιμετωπίζεται μέσα σε ένα λογικό πλαίσιο, τόσο το πειραματικό κόστος, όσο και ο αριθμός των απαιτούμενων παραμέτρων, καθιστούν δύσκολη τη μοντελοποίηση μεγάλων τοπολογιών. Από τη θετική σκοπιά, η μέθοδος μας είναι μία σημαντική εξέλιξη της λογικής του Boole, η οποία απευθύνεται επιτυχώς τόσο στην πρωτεϊνική συνδεσιμότητα όσο και στην ένταση/δραστηριότητα των αντιδράσεων σε μεγάλης κλίμακας μονοπάτια σηματοδότησης (120 είδη, 230 αντιδράσεις) ([28]).

Όταν συγκρίνεται με τη μοντελοποίηση κατά Boole, η προτεινόμενη προσέγγιση παρέχει μία ποσοτική όψη των μονοπατιών σηματοδότησης, υποστηρίζοντας συνεχείς τιμές για την ενεργοποίηση συμπεριλαμβανόμενων ειδών. Επί πλέον, κάθε αντίδραση μοντελοποιείται μέσω μιας σιγμοειδούς καμπύλης, η οποίας αναπαριστά στενά την πραγματική της μηχανική. Σαν αποτέλεσμα, η προτεινόμενη προσέγγιση δίνει χαμηλότερο σφάλμα προσαρμογής από την αντίστοιχη μοντελοποίηση κατά Boole. Σε σύγκριση με άλλα μοντέλα, ο προτεινόμενος αλγόριθμος απέδωσε εξίσου καλά με προηγούμενες προσεγγίσεις, που δείχνουν τη βελτιστοποίηση των μεσαίας κλίμακας μονοπατιών με τα δεδομένα σηματοδότησης. Ακόμα και αν οι δύο διαδικασίες ακολουθούν διαφορετικά διαγράμματα ροής, η τοπολογία των λύσεων είναι αρκετά όμοια και η προσαρμογή του σφάλματος είναι του ίδιου επιπέδου ([28]).

Η υπολογιστική απόδοση της μεθόδου NLP επιτρέπει την αναζήτηση των μονοπατιών μεγάλης κλίμακας. Απέδωσε σημαντικά καλύτερα απ' ό,τι η προσέγγιση κατά Boole σε ό,τι αφορά την προσαρμογή σφάλματος, μειώνοντάς το κατά 75% (μέσο σφάλμα προσαρμογής). Αν και αυξάνεται ο υπολογιστικός χρόνος που χρειάζεται η CPU, η λύση παρέμεινε υπολογιστικά εφικτή ([32]).

Συνολικά, η προτεινόμενη προσέγγιση αντιμετωπίζει επιτυχώς τη βελτιστοποίηση των δικτύων μεταγωγής σήματος μεγάλης και μεσαίας κλίμακας. Επέτρεψε τη γρήγορη βελτιστοποίηση των τοπολογιών σηματοδότησης, συνδυάζοντας την πολύπλευρη φύση των μοντέλων λογικής με τους παρουσιαζόμενους αλγορίθμους βελτιστοποίησης ([32]).

## **Μέθοδοι**

---

### ***Σύνθεση NLP μοντέλου***

Η προτεινόμενη μέθοδος NLP χτίζεται βασισμένη σε μία προϋπάρχουσα μέθοδο ILP και έτσι, χρησιμοποιεί την ίδια ονοματολογία ([32]).

**1.Ορισμοί:** Ένα μονοπάτι καθορίζεται σαν ένα σύνολο αντιδράσεων  $i=1,\dots,n_r$  και ειδών  $j=1,\dots,n_s$ . Κάθε αντίδραση έχει τρία αντίστοιχα σύνολα δεικτών, δηλαδή το σύνολο δεικτών των μορίων σηματοδότησης  $R_i$ , των αναστολέων  $I_i$

και των «παραγώγων»  $P_i$  (σαν «παράγωγο» μπορεί να ορισθεί, επίσης, το επίπεδο της φωσφορυλίωσης των πρωτεϊνών). Αυτά τα σύνολα, είναι όλα υποσύνολα του συνόλου των ειδών των δεικτών ( $R_i, I_i, P_i \subset \{1, \dots, n_s\}$ ) ([32]).

Ένα σύνολο εξελισσόμενων πειραμάτων παρουσιάζεται μιμούμενα τις συνθήκες του κάθε πραγματικού πειράματος. Τα πειράματα παίρνουν το δείκτη τους από τον εκθέτη  $k = 1, \dots, n_e$ . Σε κάθε πείραμα ένα υποσύνολο ειδών εισάγεται στο σύστημα και ένα άλλο υποσύνολο εξάγεται από το σύστημα, με παρόμοιο τρόπο με τα πραγματικά πειράματα, όπου ένας συνδυασμός ερεθισμάτων και αναστολέων εισάγονται στα κύτταρα. Η προβλεπόμενη τιμή ενεργοποίησης του κάθε είδους  $j$  στο πείραμα  $k$  αναπαρίσταται από τη σταθερά  $x_j^k \in [0,1]$ . Εάν είναι διαθέσιμη, η αντίστοιχη μετρούμενη τιμή αναπαρίσταται από  $x_j^{k,m} \in [0,1]$ . Η τελευταία ομάδα εισαχθέντων μεταβλητών,  $z_i^k \in [0,1]$ , αναπαριστά την ενεργοποίηση της αντίδρασης  $i$  στο πείραμα  $k$  ([32]).

**2. Αντικειμενική Συνάρτηση:** Η αντικειμενική συνάρτηση που θέλουμε να ελαχιστοποιηθεί είναι η παρακάτω ([32]):

$$\sum_{j,k} a_j^k |x_j^{k,m} - x_j^k| \quad (1)$$

Η οποία αναπαριστά την προβλεπόμενη σταθμισμένη μέτρηση αναντιστοιχίας (πειραματικών – υπολογιστικών τιμών);  $a_j^k \in [0,1]$  ρυθμίζουν τα ειδικά βάρη που ευνοούν την προσαρμογή των ειδικών κόμβων στο μονοπάτι. Στην παρούσα ενότητα, όλοι οι κόμβοι θεωρούνται εξίσου σημαντικοί (έχουν ίσα ειδικά βάρη  $a_j^k$ ) ([32]).

**3. Αντιδράσεις ενός αντιδρώντος – ενός προϊόντος:** Αντιδράσεις με μόνο ένα αντιδρών και ένα μόνο προϊόν μοντελοποιούνται με τη χρήση της ακόλουθης συνάρτησης μεταφοράς ([32]):

$$f(x) = a(p^n + 1) \frac{x^n}{x^n + p^n} \quad (2)$$

Η παραπάνω εξίσωση αναπαριστά μία κανονικοποιημένη συνάρτηση Hill. Η παράμετρος  $p$  καθορίζει το μεσαίο σημείο της καμπύλης (δηλ. η τιμή του  $x$  για την οποία η  $f(x)$  ισούται με την τιμή 0.5), το  $n$  είναι ο εκθέτης του Hill και καθορίζει την κλίση της καμπύλης, όπου το  $a$  είναι ο παράγοντας κλιμάκωσης. Η ενεργοποίηση της αντίδρασης  $i$  στο πείραμα  $k$  ισούται με:  $z_i^k = f(x_j^k)$ , όπου

$j \in R_i$ . Η τιμή ενεργοποίησης του κατάντη κόμβου ισούται με:  $x_j^k = z_i^k$ , όπου  $j \in P_i$ . Στην περίπτωση που το είδος  $j$  είναι αναστολέας, χρησιμοποιείται:  $z_i^k = f(1 - x_j^k)$ , όπου  $j \in R_i$  ([32]).

**4. Αντιδράσεις πολλαπλών αντιδρώντων – ενός μόνο παραγώγου (πύλες AND) :** Στην περίπτωση που περισσότερα από ένα αντιδρώντα χρειάζονται για τη διάδοση του σήματος στα κατάντη είδη, η δραστηριότητα της αντίδρασης  $i$  μοντελοποιείται σαν μία συνάρτηση του διγραμμικού παραγώγου του αντιδρώντος είδους ([32]):

$$z_i^k = f \left( \prod_{j \in R_i} x_j^k \times \prod_{j \in I_i} (1 - x_j^k) \right) \quad (3)$$

Η τιμή ενεργοποίησης του κατάντη κόμβου ισούται με:  $x_j^k = z_i^k, j \in P_i$ . Το διγραμμικό προϊόν επιλέγεται για την ικανοποίηση βασικών ιδιοτήτων, όπως είναι η συνέχεια, η διαφορισιμότητα και για την αναπαραγωγή της πύλης Boole AND για τις τιμές 0 και 1 των αντιδρώντων ειδών ([32]).

**5. Πολλαπλές αντιδράσεις που οδηγούν στο ίδιο προϊόν (πύλες OR):** Σε περίπτωση που περισσότερες από μία αντιδράσεις οδηγούν στο ίδιο προϊόν, η τιμή ενεργοποίησης των κατάντη ειδών δίνεται από την ακόλουθη διατύπωση ([32]):

$$x_j^k = b_{|T_j|}^k \quad (4)$$

Όπου,

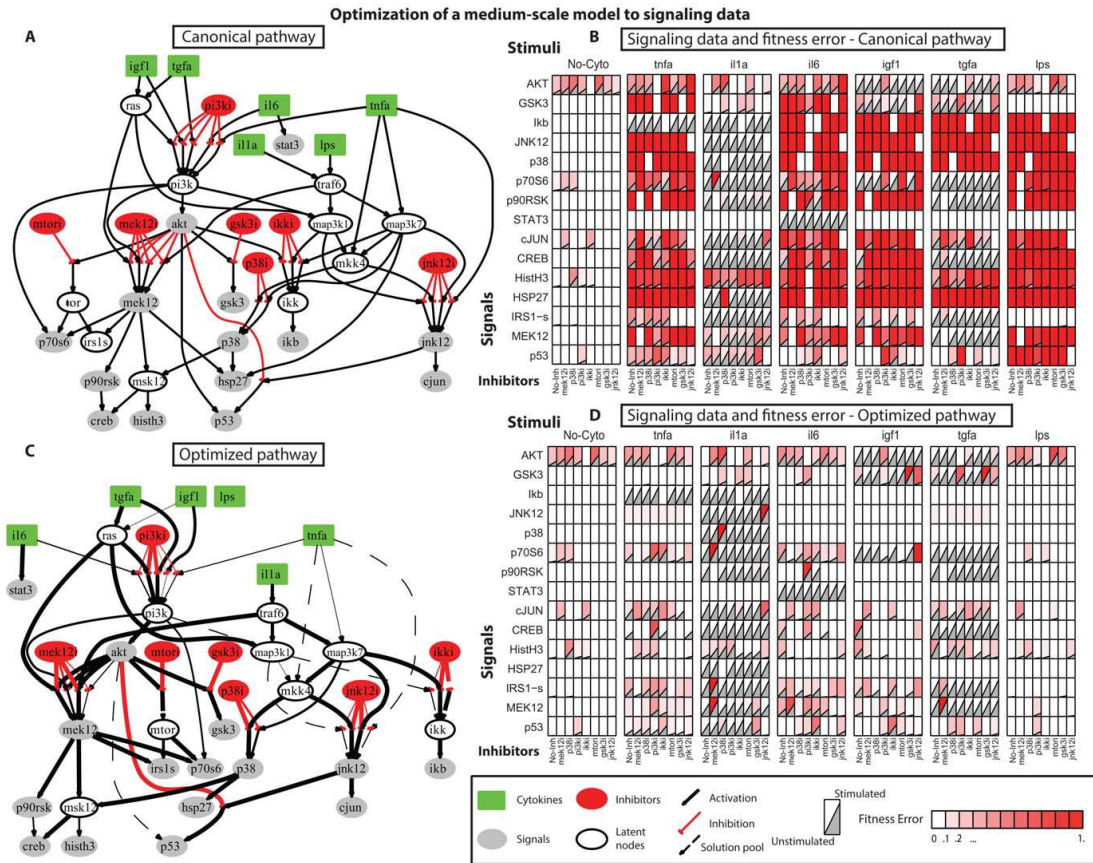
$$T_j = \{i \in \{1, \dots, n_r\} : j \in P_i\} \quad (5)$$

$T_j$  είναι το σύνολο όλων των αντιδράσεων που έχουν είδος  $j$  σαν προϊόν τους. Έστω ότι οι τιμές  $i_1, i_2, \dots, i_{|T_j|}$  υποδεικνύουν τα στοιχεία του  $T_j$ . Έπειτα, οι τιμές  $b_m^k$  υπολογίζονται αναδρομικά, ως ([32]):

$$b_m^k = b_{m-1}^k + z_{i_m}^k - b_{m-1}^k z_{i_m}^k ; 2 < m \leq |T_j| \quad (6)$$

$$b_2^k = z_{i_1}^k + z_{i_2}^k - z_{i_1}^k z_{i_2}^k \quad (7)$$

**6.Εκτέλεση:** Ο στόχος της μεθόδου NLP, όπως περιγράφηκε και παραπάνω, είναι η αναγνώριση των βέλτιστων τιμών για τις παραμέτρους  $\alpha, \rho$  και  $\eta$  κάθε αντίδρασης για την ελαχιστοποίηση της διαφοράς μεταξύ των προβλέψεων του μοντέλου και των πειραματικών δεδομένων ([32]).



**Εικόνα 70: Βελτιστοποίηση ενός μεσαίας κλίμακας μοντέλου με χρήση της NLP μεθόδου ([32])**

(A) Στην εικόνα αυτή παρουσιάζεται η αρχική τοπολογία του δικτύου. (B) Δεδομένα σηματοδότησης υπό τη συνδυαστική δράση 6 ερεθισμάτων (πράσινοι κόμβοι) και 7 αναστολέων (κόκκινοι κόμβοι) που εκθέτουν 15 σήματα (γκρι κόμβοι). Το κόκκινο φόντο αναπαριστά τη μέτρηση – πρόβλεψη αναντιστοιχίας της αρχικής τοπολογίας (46%) (μέσο σφάλμα προσαρμογής). Για την παραγωγή προβλέψεων ενός μοντέλου, οι αρχικές υποθέσεις όλων των παραμέτρων του μοντέλου χρησιμοποιήθηκαν ( $\alpha=1.0, \rho=0.5$ ). (C) Στην εικόνα αυτή παρουσιάζεται το δίκτυο, όπως προέκυψε μετά τη βελτιστοποίηση. Οι υπογραμμισμένες γραμμές αναφέρονται στο βέλτιστο δίκτυο, μετά την αφαίρεση περιττών/αντικρουόμενων αντιδράσεων. Οι διακεκομμένες γραμμές αναφέρονται στις αντιδράσεις που παρευρίσκονται στην οικογένεια λύσεων, οι οποίες αν και είναι περιττές, αναφέρονται μιας και μπορεί να έχουν βιολογική σημασία. Η αδιαφάνεια της κάθε ακμής αντιστοιχεί στη δραστηριότητα της αντίστοιχης αντίδρασης. (D) Τα δεδομένα σηματοδότησης και το εναπομείναν σφάλμα προσαρμογής (8%) (μέσο σφάλμα προσαρμογής). Το κόκκινο φόντο αναφέρεται στο σφάλμα προσαρμογής της παραγόμενης λύσης. Τα μειωμένες έντασης κόκκινα τριγώνια, σε σύγκριση με αυτά της εικόνας (B), συνεπάγεται το βέλτιστο μοντέλο, το οποίο προσαρμόζεται επιτυχώς στα δεδομένα σηματοδότησης (το μέσο σφάλμα προσαρμογής μειώθηκε από 46% σε 8%) ([32]).

### 3. Αναπτυσσόμενη Μέθοδος

Ανακεφαλαιώνοντας όλα όσα έχουν παρουσιασθεί εκτενώς παραπάνω, τα μονοπάτια σηματοδότησης είναι υψίστης σημασίας για την κατανόηση των κυτταρικών λειτουργιών και την πρόβλεψη της απόκρισης των σε ενδεχόμενες διαταραχές. Πρόσφατες εξελίξεις στον τομέα της «εξόρυξης κειμένου» (text mining) και η κατασκευή δικτύων πρωτεϊνικών αλληλεπιδράσεων (**Protein – Protein Interaction – PPI networks**) έχει οδηγήσει στη σύσταση μεγάλων διεθνών βάσεων δεδομένων μονοπατιών σηματοδότησης, δείχνοντας πως οι πρωτεΐνες αλληλεπιδρούν η μία με την άλλη. Παρόλα αυτά, η συλλογή και η οπτικοποίηση της συνδεσιμότητας των πρωτεϊνών στα δίκτυα σηματοδότησης είναι απλά το πρώτο βήμα για την κατανόηση των κυτταρικών μηχανισμών σηματοδότησης. Η ανάλυση και η μοντελοποίηση αυτών των δικτύων είτε σε επίπεδο συνδεσιμότητας, είτε σε επίπεδο των μηχανισμών μεταγωγής σημάτων μεταξύ των κόμβων είναι ένα κρίσιμο επόμενο βήμα για την κατασκευή λειτουργικών μοντέλων, προβλέψιμων της κυτταρικής βιολογίας ([34]).

Μία ποικιλία μεθόδων έχουν προταθεί για αυτή τη δουλειά, κάθε μία υιοθετεί μία διαφορετική προοπτική στη φύση των συμπεριλαμβανόμενων αντιδράσεων και εστιάζει σε διάφορες ιδιότητες του δικτύου σηματοδότησης. Δύο ευρείες κατηγορίες ανάλυσης δικτύων μπορούν να διακριθούν ([34]):

- (i) **Τοπολογική ανάλυση** των δικτύων σηματοδότησης που εξαγει την εικόνα της λειτουργίας των κυτάρων ερευνώντας τα δομικά χαρακτηριστικά των δικτύων σηματοδότησης (δηλ. βρόχοι ανάδρασης, ισχυρά συνεκτικές συνιστώσες).
- (ii) **Αναγνώριση δικτύου**, για την αναγνώριση της δομής των δικτύων, (δηλ. συνδεσιμότητα των ειδών σηματοδότησης) ή των παραμέτρων αντίδρασης που καθορίζουν τους μηχανισμούς της μεταγωγής σημάτων από τον ένα κόμβο στον επόμενο.

Τυπικά υιοθετείται μία μαθηματική μέθοδος για τη μοντελοποίηση του τρόπου διεξαγωγής της μεταγωγής των σημάτων και ένα εκτελέσιμο μοντέλο δημιουργείται συνδυάζοντας αυτή τη μέθοδο με προηγούμενη γνώση δικτύων (prior knowledge network – PKN), οποία χρησιμοποιείται ουσιαστικά σαν μία «μήτρα». Προσομοιάζοντας το μοντέλο υπό διαφορετικές παραμέτρους κόμβων και αντιδράσεων, συμπεράσματα μπορούν να εξαχθούν για τη σημασία κάθε κόμβου και αντίδρασης στη διάδοση του σήματος. Μεταξύ των πιο ευρέως διαδεδομένων μεθόδων βρίσκονται, όπως έχει προαναφερθεί, οι διάφορες αρχές της λογικής μοντελοποίησης και των Συνήθων Διαφορικών Εξισώσεων (Ordinary Differential Equations – ODEs). Σε συγκεκριμένες περιπτώσεις, το αρχικό μοντέλο έχει εκπαιδευτεί σε δεδομένα σηματοδότησης μέσω μίας προσέγγισης βελτιστοποίησης για τον υπολογισμό των τιμών των παραμέτρων μοντέλου, οι οποίες προσεγγίζουν τα δεδομένα καλύτερα ή χρησιμοποιείται μία αναλυτική προσέγγιση ευαισθησίας για τον υπολογισμό της επίδρασης των παραμέτρων μοντέλου στη συνολική απόκριση του μοντέλου. Η ενσωμάτωση των δεδομένων σηματοδότησης επιτρέπει τη



δημιουργία κυτταρικό – κεντρικών, ιστό – κεντρικών ή ασθενειών – κεντρικό μονοπατιών ([34]).

### **3.1 Εισαγωγικά**

Η επιλογή της προσέγγισης της μοντελοποίησης και κατά συνέπεια της διαδικασίας βελτιστοποίησης σχετίζεται στενά με τη διαθεσιμότητα των υπαρχόντων δεδομένων και βιολογικών ερευνών. Για παράδειγμα, εάν είναι διαθέσιμα δεδομένα εξέλιξης χρόνου και η δυναμική των αντιδράσεων σηματοδότησης είναι ενδιαφέρουσες, μία προσέγγιση μέσω συνηθών διαφορικών εξισώσεων μπορεί να είναι κατάλληλη, ειδικά εάν το εξεταζόμενο δίκτυο σηματοδότησης είναι μικρό σε μέγεθος. Για το σκοπό αυτό, έχει δημοσιευθεί ανά περιόδους, σημαντική δουλειά στην εκτίμηση παραμέτρων των μοντέλων βασισμένων σε συνήθεις διαφορικές εξισώσεις, χρησιμοποιώντας ένα ευρύ φάσμα μεθόδων, συμπεριλαμβανομένων γενικών τελολογικών μεθόδων βελτιστοποίησης (αλγόριθμοι βασισμένοι στην κλίση, στοχαστικοί αλγόριθμοι, γεωμετρικός προγραμματισμός, εκτίμηση δυναμικής ροής και άλλοι). Παρόλα αυτά, τα δίκτυα σηματοδότησης μεγάλης κλίμακας, δεν μπορούν εύκολα να αντιμετωπιστούν μέσα στο πλαίσιο της μοντελοποίησης με Συνήθεις Διαφορικές Εξισώσεις, λόγω της μεγάλης χρονικής – υπολογιστικής διάρκειας (που χρειάζεται η CPU για την προσπέλαση των δεδομένων) και την έλλειψη κατάλληλων περιορισμών σταθερών σύνδεσης και διάστασης. Εάν δεδομένα για μεγάλης κλίμακας δίκτυα είναι διαθέσιμα, τότε μπορεί να χρησιμοποιηθεί μία πιο λογικό – κεντρική μοντελοποίηση (λογική κατά Boole, προσομοιωμένη σε μία «ψευδό – σταθερή κατάσταση») για την αναγνώριση της δομής των μονοπατιών σηματοδότησης ([33]).

Στα πλαίσια της παρούσας διπλωματικής, ξεφεύγουμε από τα παραπάνω προτυπα και επιλέχθηκε μία αλγοριθμική προσέγγιση, η οποία ξεφεύγει από τα πλαίσια της βελτιστοποίησης δικτύων, όπως αυτή περιγράφηκε στην παραπάνω ενότητα. Όπως θα φανεί εκτενώς και πιο κάτω, η παρούσα μοντελοποίηση ξεκινάει με την αναγνώριση των εκάστοτε δεδομένων δικτύου, συνεχίζει με τη στατιστική ανάλυση του, με την εύρεση βασικών χαρακτηριστικών δικτύου (βάσει των αλγορίθμων που εκτέθηκαν στην ενότητα 2.5) και καταλήγει στην εύρεση των πιθανών μονοπατιών από κάθε κόμβο προς κάθε άλλο κόμβο του δικτύου, εάν αυτά υπάρχουν. Τέλος, η κατά τρόπο βελτιστοποίηση που πραγματοποιείται ξεφεύγει από τα πλαίσια της συμβατικής βελτιστοποίησης, υπό την έννοια της υπάρξεως κάποιας αντικειμενικής συνάρτησης, περιορισμών, μαθηματικών μοντέλων και συναρτήσεων, βρόχων επαναλήψεων ή υπολογιστικού κόστους.

Η μοντελοποίηση εστιάζει σε αλγοριθμικές μεθόδους και προγραμματιστικές διαχειρίσεις δεδομένων.

### 3.2 Μελέτη προβλήματος

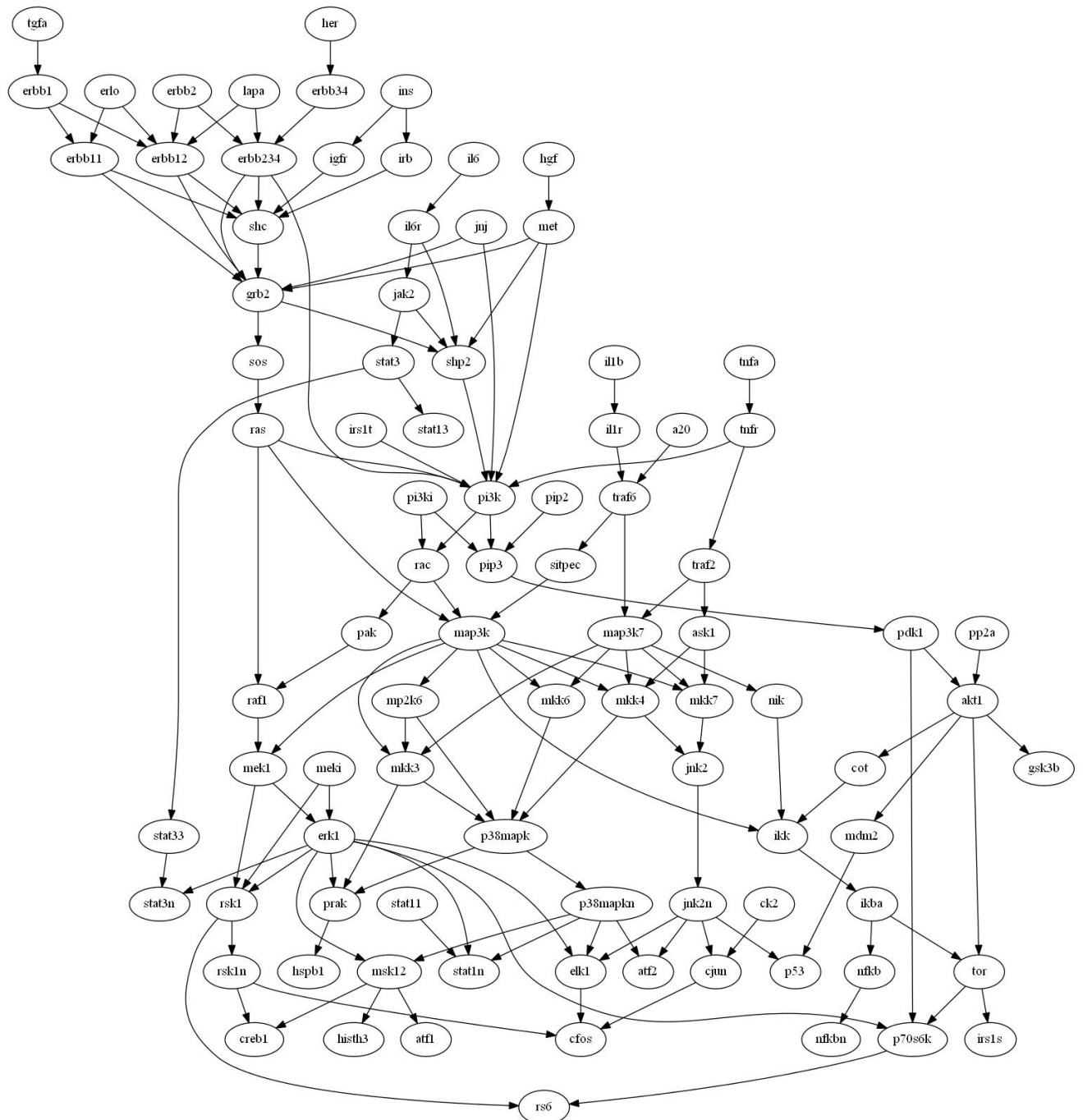
Στην ενότητα αυτή παρουσιάζουμε αναλυτικά το πρόβλημα που εξετάζεται στην παρούσα διπλωματική. Πιο συγκεκριμένα:

Έχουμε μία αρχική τοπολογία δικτύου, βάσει της διεθνούς βάσεως δεδομένων και βιβλιογραφίας. Ουσιαστικά, αναφερόμαστε σε όλες τις γνωστές αντιδράσεις και που έχουν καταγραφεί από πειραματικές διαδικασίες σε εργαστήρια ανά τον κόσμο. Οπτικοποιώντας αυτές τις αντιδράσεις με το κατάλληλο λογισμικό παίρνουμε μία πλήρη και σαφή απεικόνισή του εξεταζόμενου δικτύου. Στην παρακάτω εικόνα 71 παρουσιάζεται ένα παράδειγμα τέτοιας απεικόνισης.

Στην εικόνα αυτή παρουσιάζεται ουσιαστικά ένα δίκτυο αντιδράσεων, όπως αυτές έχουν προκύψει από προγενέστερες μετρήσεις. Στην εικόνα αυτή μπορούμε να διακρίνουμε καθαρά τις αλληλεπιδράσεις μεταξύ πρωτεϊνών και τα μονοπάτια που ενώνουν πιθανές «πηγές» σήματος (κόμβοι κορυφής) με πιθανούς «στόχους» (απόληξη σήματος). Τώρα, έστω δεδομένης τοπολογίας δικτύου (εικόνα 71), λαμβάνουμε ένα σύνολο πειραματικών μετρήσεων σαν και αυτό το οποίο παρουσιάζεται στον πίνακα της εικόνας 72. Ο συγκεκριμένος πίνακας – παράδειγμα, αποτελείται από 6 γραμμές και 16 στήλες. Κάθε γραμμή αντιστοιχεί και αντιπροσωπεύει κάθε μία από τις ουσίες – ερεθίσματα, τις οποίες «έριξαν» οι πειραματιστές στις εξεταζόμενες ανά περίπτωση κυτταροσειρές. Αντίστοιχα, κάθε στήλη αντιπροσωπεύει την κάθε μετρούμενη πρωτεΐνη. Ας σημειωθεί σε αυτό το σημείο πως οι μετρήσεις αυτές – πειραματικά δεδομένα (experimental data) έχουν προκύψει από την περιγραφόμενη, σε παραπάνω ενότητα, μέθοδο μέτρησης πρωτεϊνών ELISA.

Οι τιμές 0 και 1 που παρουσιάζονται στα κελιά του πίνακα, ανακαλώντας και τα όσα περιγράφηκαν στην παράγραφο 2.5.5 για τη μοντελοποίηση βάσει της λογικής κατά Boole, αντιστοιχούν στις απλές λογικές εκφράσεις «δεν αντέδρασε» ή «αντέδρασε». Επεξηγηματικά, για την πρώτη γραμμή βλέπουμε πως ρίχνοντας σαν ερέθισμα il6, τότε οι πρωτεΐνες π.χ. hspb1 ή mp2k6 δεν αντιδρούν, ενώ οι πρωτεΐνες ikba και nfkb αντιδρούν.

*Πρέπει σε αυτό το σημείο να επισημάνουμε πως τα πρόσημα στις τιμές -1 και 1, έχουν βιολογική σημασία και υποδεικνύουν τη μείωση ή την αύξηση των αντίστοιχων ουσιών μετά τη ρήψη του ερεθίσματος. Ουσιαστικά, αποτελούν μία ιδιαίτερα χρήσιμη και σημαντική πηγή πληροφοριών για τον πειραματιστή. Παρόλα αυτά, στην παρούσα μοντελοποίηση δεν θα μας απασχολήσει αυτό, καθώς αντιμετωπίζουμε και τις δύο τιμές σαν ένα διακόπτη ON. Υπενθυμίζουμε πως μας ενδιαφέρει η μεταγωγή του σήματος και ο καθορισμός της πορείας της και όχι η ποσοτικοποίηση του.*



Εικόνα 71: Παράδειγμα απεικόνισης τοπολογίας δικτύου

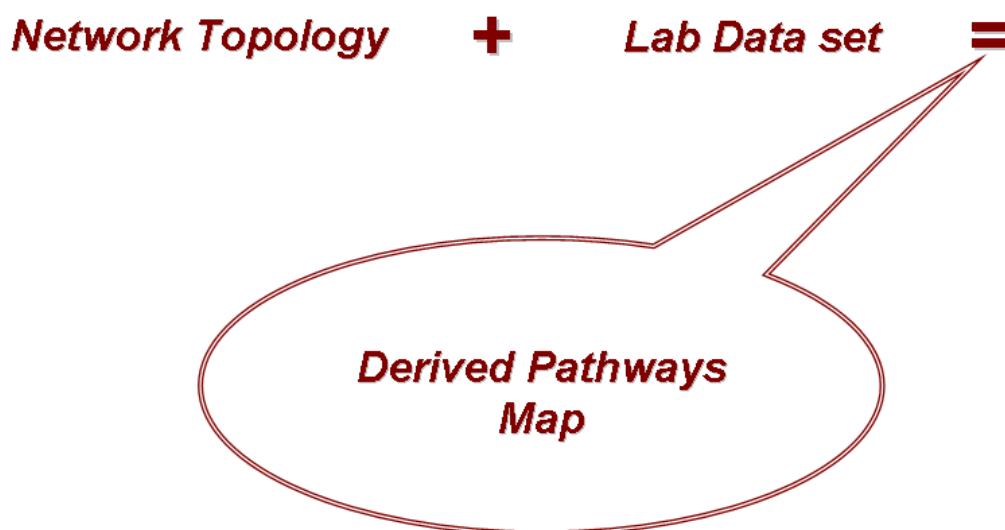
	<i>hs pb1</i>	<i>akt1</i>	<i>p70s6k</i>	<i>shp2</i>	<i>jnk2</i>	<i>ikba</i>	<i>gsk3b</i>	<i>p38mapk</i>	<i>nfk b</i>	<i>mp2k6</i>	<i>tor</i>	<i>mek1</i>	<i>erk1</i>	<i>rsk1</i>	<i>creb1</i>	<i>rs6</i>
<i>il6</i>	0	0	0	0	0	-1	0	0	1	0	0	0	0	0	0	0
<i>tnfa</i>	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0
<i>il1b</i>	1	0	0	0	0	-1	0	1	1	0	0	1	0	0	1	0
<i>tgfa</i>	0	1	1	1	0	0	1	1	1	1	0	1	1	1	1	1
<i>ins</i>	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
<i>dme</i>	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Εικόνα 72: Παράδειγμα πίνακα εξάρτησης δεδομένων

Στην παραπάνω ενότητα 2.7 μιλήσαμε για βελτιστοποίηση δικτύων και εύρεση βέλτιστων μονοπατιών μεταγωγής σήματος με χρήση μεθόδων γραμμικού – μη γραμμικού ακέραιου προγραμματισμού. Αυτό που είναι αναγκαίο να κατανοήσει ο αναγνώστης, είναι πως η έννοια της βελτιστοποίησης υπεισέρχεται στο βέλτιστο «πάντρεμα» αρχικής τοπολογίας (σαν αυτή που παρουσιάστηκε στην εικόνα 71) και πειραματικών δεδομένων (σαν και αυτά που παρουσιάστηκαν στην εικόνα 72). Αυτός ακριβώς είναι και ο σκοπός μας στην παρούσα διπλωματική.

Καταλήγοντας, λοιπόν, έχοντας καθορίσει τη φύση των υπαρχόντων δεδομένων και το βασικό σκοπό θα προσπαθήσουμε στις παρακάτω ενότητες να παρουσιάσουμε μία διαφορετική μέθοδο ενός τέτοιου «παντρέματος», η οποία θα διαφοροποιείται από τις ήδη υπάρχουσες μεθόδους και θα βασίζεται μόνο σε αλγοριθμικές προσεγγίσεις.

Στην παρακάτω εικόνα απεικονίζεται συνοπτικά το πραγματευόμενο πρόβλημα.



Εικόνα 73: Συνοπτική παρουσίαση εξεταζόμενου προβλήματος

### 3.3 Εφαρμοζόμενη μέθοδος

Αρχικά, είναι χρήσιμο να αναφερθεί πως οι κώδικες γράφτηκαν στο προγραμματιστικό κομμάτι του λογισμικού Matlab. Επιλέχθηκε το συγκεκριμένο λογισμικό, λόγω εξοικείωσης του παρόντος φοιτητή με το συγκεκριμένο προγραμματιστικό περιβάλλον και κυρίως λόγω των δυνατοτήτων του λογισμικού. Πιο συγκεκριμένα, το συγκεκριμένο λογισμικό παρέχει πέραν της δυνατότητας προγραμματισμού, τη δυνατότητα εκμετάλλευσης πολλαπλών αναπτυσσόμενων εργαλείων στο πεδίο της βιοπληροφορικής (έτοιμα toolboxes). Εργαλεία που βοηθούν στην οπτικοποίηση, στην παρουσίαση, στη σχεδίαση ή και στη διαχείριση δικτύων και μονοπατιών. Ενδεικτικά αναφέρονται κάποια τέτοια εργαλεία: CellNetAnalyzer, SBMLToolbox, MAPK\_dynsys\_identifier κ.λπ. Αξίζει να σημειωθεί πως η διεθνής κοινότητα της βιοπληροφορικής έχει αναπτύξει πολλά έτοιμα τέτοια εργαλεία, τα περισσότερα από τα οποία είναι συμβατά με το λογισμικό της Matlab. Εξάλλου, το συγκεκριμένο λογισμικό χρησιμοποιείται ευρέως από την κοινότητα, λόγω αυτών ακριβώς των διευκολύνσεων που παρέχει.

Γενικά, αναφέρουμε πως η κοινότητα της βιοπληροφορικής χρησιμοποιεί κυρίως, πέραν της Matlab, την προγραμματιστική γλώσσα της C,C++ και της Python. Αυτό δεν σημαίνει πως υπάρχει κάποιος κανόνας, απλά αναφέρεται η συνήθης κατάσταση. Πολλοί από τους κώδικες των βασικών αλγορίθμων δικτύων (όπως παρουσιάζονται στην παράγραφο 2.5.2) είναι ήδη γραμμένοι σε γλώσσα C. Αυτό που έπρεπε να γίνει σε πρώτο βήμα, λοιπόν, είναι να μεταφερθούν οι αλγόριθμοι αυτοί στο περιβάλλον της Matlab.

Κατά τη διάρκεια της πραγμάτωσης του παρόντος, εξετάσθηκαν πολλά και διαφορετικά δεδομένα. Η ανάπτυξη των αλγορίθμων έγινε σταδιακά και παράλληλα με την εξέταση των διαφορετικών τύπου δεδομένων. Για το λόγο αυτό οι αλγόριθμοι συντάσσονταν κάθε φορά από την αρχή, ώστε να προσαρμόζονται στον διαφορετικό τύπο δεδομένων. Έτσι, αναπτύχθηκαν συνολικά τέσσερις τελικοί και χρηστικοί αλγόριθμοι. Πολλά κομμάτια του ενός επικαλύπτουν τους υπόλοιπους, πάραυτα οι διαφορές σε συγκεκριμένα σημεία είναι εμφανείς και συνάδουν, όπως αναφέρθηκε ήδη, στη διαφορετικότητα των δεδομένων. Για λόγους πληρότητας στο τελικό παράρτημα θα παρουσιασθούν όλοι οι αλγόριθμοι, μιας και αποτελούν αποτέλεσμα πολύμηνης, κοπιαστικής, πνευματικής εργασίας του παρόντος φοιτητή.

## Παρουσίαση βημάτων μεθόδου

---

Οι αναπτυσσόμενοι κώδικες συνοψίζονται στα επόμενα βήματα:

### ➤ Ανάγνωση δεδομένων

Στην αρχή ο αλγόριθμος δέχεται τα δεδομένα από το χρήστη. Ως δεδομένα μπορούν να θεωρηθούν:

- αρχεία που περιέχουν τις συνολικές αλληλεπιδράσεις του δικτύου, ουσιαστικά την τοπολογία του (Network Generic Topology – *εικόνα 71*).
- αρχεία που θέλουμε να προσομοιάσουμε στο δίκτυο μας. Πρόκειται ουσιαστικά για τα πειραματικά δεδομένα (experimental data sets), βάσει των οποίων θέλουμε να κατασκευάσουμε το τελικό δίκτυο με τα μονοπάτια μεταγωγής σήματος. Η μορφή με την οποία δέχεται ο αλγόριθμος τα δεδομένα αυτά, μπορεί να είναι της μορφής που παρουσιάστηκε παραπάνω στην *εικόνα 72* ή μπορεί να δίνονται σε μορφή αλληλεπιδράσεων, όπως παρουσιάζεται στην παρακάτω *εικόνα*.

TGFA	1	AKT1
TGFA	1	MAP2K1
TGFA	1	GSK3B
TGFA	1	EGFR
TGFA	1	CREB1
TNF	1	HSPB1
TNF	1	TP53
TNF	1	MAPK14
TNF	1	NFKB1
IL6	1	STAT3
IL6	1	AKT1
IL6	1	GSK3B
IL1A	1	NFKB1
IL1A	1	MAPK14
IL1A	1	MAPK9

Ο παραπάνω πίνακας δείχνει, ουσιαστικά, πως θέλουμε για παράδειγμα, να πάμε από τον κόμβο TGFA στον κόμβο AKT1. Το ένα ενδιάμεσα συμβολίζει, ουσιαστικά, αυτό που αναφέρθηκε και παραπάνω πως υπάρχει αντίδραση μεταξύ των δύο αυτών κόμβων.

- Αρχεία που περιλαμβάνουν τα ερεθίσματα τα οποία χρησιμοποίησαν οι πειραματιστές κατά την πειραματική διαδικασία (stimuli data).

- Αρχεία με τους κόμβους των οποίων τα σήματα μετρήθηκαν κατά τη διάρκεια της πειραματικής διαδικασίας (Signal Nodes). Ουσιαστικά πρόκειται για εκείνα τα σήματα που μετρήθηκαν από τους πειραματιστές, τα οποία αποτελούν εξαρχής και το λόγο πραγματοποίησης των πειραμάτων αυτών. Στο σημείο αυτό οφείλουμε να σημειώσουμε, πως ένα βιολογικό πείραμα για τη μέτρηση πρωτεϊνών με την τεχνική ELISA, είναι αρκετά ακριβό. Για το λόγο αυτό είναι και ιδιαίτερα σημαντική η μοντελοποίηση δικτύων. Γιατί γλυτώνει τους πειραματιστές από ανεπιθύμητα έξοδα.

### ➤ Στατιστική Ανάλυση Δικτύων

Σε επόμενο βήμα ο αλγόριθμος παρουσιάζει μία πρώτη στατιστική εικόνα του δικτύου, σχετικά με το συνολικό αριθμό των κόμβων, των αντιδρώντων, των παραγώγων, των συνολικών αντιδράσεων, των σημάτων, των ερεθισμάτων, των μετρούμενων σημάτων, των πηγών και των στόχων του δικτύου.

Τα στοιχεία αυτά παρουσιάζονται στην οθόνη, ώστε ο χρήστης να μπορεί να αποκτήσει μία πρώτη εικόνα του εξεταζόμενου δικτύου.

### ➤ Αναπαράσταση Δικτύου

Το πρώτο βήμα για την μοντελοποίηση του δικτύου, είναι η αναπαράσταση του με τη χρήση μητρώου γειννίαςσης (Adjacency Matrix Representation). Όπως είδαμε και στην ενότητα 2.5.2, η αναπαράσταση αυτή είναι ιδιαίτερα σημαντική, ώστε να μπορούμε να έχουμε μία μητρική – αριθμητική αναπαράσταση του δικτύου.

Πιο συγκεκριμένα, έστω  $n$  ο συνολικός αριθμός των κόμβων του δικτύου, όπως προέκυψε από την παραπάνω στατιστική ανάλυση του δικτύου. Έτσι, δημιουργούμε ένα μητρώο, έστω  $adjacency(n \times n)$ , με όλα του τα στοιχεία μηδενικά στην αρχή.

Έστω τώρα, πως η τοπολογία του δικτύου που αναφέρθηκε παραπάνω δίνεται στην μορφή:

$$\begin{array}{c}
 R_1 \xrightarrow{1} P_i \\
 R_2 \xrightarrow{1} P_2 \\
 \hline
 R_i \xrightarrow{1} P_i
 \end{array}$$

όπου  $R_i$  είναι το αντιδρών της  $i$ -οστής αντίδρασης και  $P_i$  το αντίστοιχο παράγωγο και  $i=1,2,\dots,N$  με  $N$  ο συνολικός αριθμός αντιδράσεων στο δίκτυο.

Έστω, τώρα, πως έχουμε δημιουργήσει ένα μητρώο  $nodes(n \times 1)$  στο οποίο αποθηκεύουμε τον αύξοντα αριθμό του κάθε κόμβου. Για τη δημιουργία, λοιπόν του μητρώου γειτνίασης, πρέπει να:

$$adjacency(nodes(R_i), nodes(P_i)) = 1$$

Με αυτό τον απλό τρόπο μπορούμε να αποτυπώσουμε όλες τις αντιδράσεις του δικτύου σε ένα μητρώο με τη μορφή 1.

*Να σημειωθεί πως για μεγάλης κλίμακας δίκτυα ( $N > 5000$ ), η δημιουργία του μητρώου γειτνίασης μπορεί να πάρει πολλή ώρα υπολογιστικού χρόνου. Για το λόγο αυτό ο αρμόδιος προγραμματιστής θα πρέπει να είναι σίγουρος για την ορθότητα του κώδικά του, ώστε να μην χάνει πολύτιμο χρόνο. «Τρέχοντάς» τον μία φορά, μπορεί μετά να αποθηκεύσει το παραγόμενο μητρώο και να το ανακαλεί όποτε το χρειάζεται.*

#### ➤ **Ανάλυση Δικτύου με χρήση Θεμελιωδών Αλγορίθμων**

Στο σημείο αυτό ο αλγόριθμος αναλύει το εκάστοτε δίκτυο χρησιμοποιώντας τους θεμελιώδεις αλγορίθμους που παρουσιάστηκαν στην ενότητα 2.5.2.

Έτσι ο αλγόριθμος βρίσκει τις Ισχυρά Συνεκτικές Συνιστώσες (Strongly Connected Components) του δικτύου με χρήση του αλγορίθμου του Tarzan. Επίσης, υπολογίζει το ζευγνύον δέντρο (spanning tree) και το ελάχιστο ζευγνύον δέντρο (minimum spanning tree) του δικτύου. Επίσης, υπολογίζεται η μεταβατική κλειστότητα του δικτύου σύμφωνα με τον κώδικα που παρατέθηκε στην παραπάνω ενότητα.

Έτσι ο χρήστης μπορεί να αποκτήσει μία ακόμα εικόνα για τη δομή του δικτύου και τη συνεκτικότητα των στοιχείων του.

#### ➤ **Δημιουργία μονοπατιών με χρήση της μεθόδου του Warshall**

Το κομμάτι αυτό αποτελεί ένα από τα πιο σημαντικά του αλγορίθμου καθώς είναι αυτό το οποίο δημιουργεί τα μονοπάτια επικοινωνίας μεταξύ των κόμβων. Πιο συγκεκριμένα, ξεκινώντας από κάθε κόμβο σαν έναρξη και για κάθε άλλο κόμβο, μπορούμε να βρίσκουμε την ενδιάμεση διαδρομή, δηλαδή τους ενδιάμεσους κόμβους που συνδέουν την αρχή και το τέλος. Το τελικό αποτέλεσμα είναι ένα μητρώο με τον αύξοντα αριθμό των κόμβων που μεσολαβούν μεταξύ των επιθυμητών κόμβων.

Έτσι πλέον ο χρήστης έχει τις επιθυμητές διαδρομές ανάμεσα σε όλους τους κόμβους, βάσει πάντα της τοπολογίας του δικτύου.



Να σημειωθεί πως για μεγάλης κλίμακας δίκτυα ( $N > 5000$ ), η δημιουργία του μητρώου σύντομων διαδρομών μπορεί να πάρει πολλή ώρα υπολογιστικού χρόνου. Για το λόγο αυτό ο αρμόδιος προγραμματιστής θα πρέπει να είναι σίγουρος για την ορθότητα του κώδικά του, ώστε να μην χάνει πολύτιμο χρόνο. «Τρέχοντάς» τον μία φορά, μπορεί μετά να αποθηκεύσει το παραγόμενο μητρώο και να το ανακαλεί όποτε το χρειάζεται.

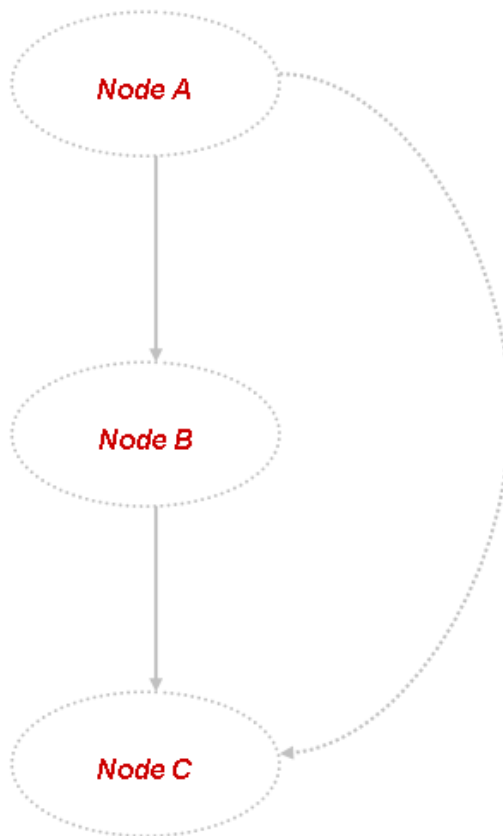
### ➤ **Έλεγχος Αρχείου Πειραματικών Δεδομένων**

Όπως αναφέρθηκε και επεξηγήθηκε και παραπάνω, ο βασικός στόχος είναι η προσαρμογή των πειραματικών δεδομένων σε μία αρχική βασική τοπολογία. Για να γίνει αυτό, όμως, πρέπει πριν αρχίσουμε να χτίζουμε το τελικό δίκτυο μονοπατιών που αντιστοιχεί κατά το βέλτιστο τρόπο στις μετρήσεις μας, να επαληθεύσουμε εάν όλες οι καταγραμμένες μετρήσεις υπάρχουν σαν κόμβοι του εξεταζόμενου δικτύου. Γιατί εάν κάποια δεν υπάρχει, η τελική τοπολογία δεν θα συμπίπτει με την αρχική τοπολογία. Άρα, θα υπάρχει σφάλμα ταιριάσματος (matching error).

Όσες μετρήσεις δεν υπάρχουν, ο αλγόριθμος τις αποκλείει πρώτα από το αρχείο πειραματικών δεδομένων και μετά συνεχίζει τη ανάλυση.

### ➤ **Εύρεση Αντικειμενικών Μονοπατιών**

Στο κομμάτι αυτό πραγματοποιείται η εύρεση των πρώτων στοιχείων του τελικού δικτύου. Με τον όρο **Αντικειμενικό Μονοπάτι** ονομάζουμε το μονοπάτι εκείνο το οποίο συνδέει τους δύο κόμβους εξάρτησης (dependency file) μέσω ενδιάμεσων κόμβων που δεν αποτελούν όμως **Κόμβους Συγκρούσεων (Conflict Nodes)**.



Εικόνα 74: Παράδειγμα Κόμβου Σύγκρουσης

Στο σημείο αυτό πρέπει να ορίσουμε τι είναι ένας Κόμβος Συγκρούσεων. Ως Κόμβος Συγκρούσεων, ορίζεται ο κόμβος αυτός, από τον οποίο δεν θέλουμε να περάσουμε και θέλουμε να αποφύγουμε, εάν μπορούμε. Ουσιαστικά και όλο το κομμάτι της βελτιστοποίησης που αναφέρθηκε στην ενότητα 2.7 γίνεται με ακριβώς αυτόν το σκοπό: **να γίνει το καλύτερο ταίριασμα μεταξύ πειραματικών δεδομένων και τοπολογίας, ώστε να έχουμε τις λιγότερες δυνατές συγκρούσεις στο δίκτυό μας**. Κόμβοι σύγκρουσης μπορούν να θεωρηθούν οι κόμβοι σήματος (Signal Nodes) που αναφέρθηκαν παραπάνω. Στην παραπάνω εικόνα παρουσιάζεται ένα τέτοιο παράδειγμα σύγκρουσης.

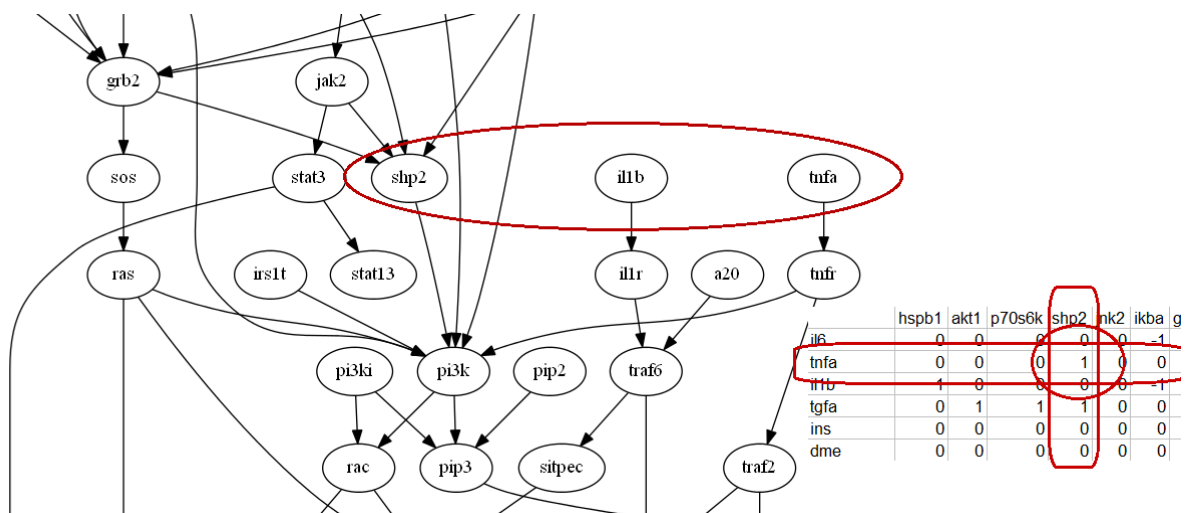
Πιο συγκεκριμένα, έστω πως θέλουμε να πάμε από τον κόμβο A στον κόμβο C. Οι μη διακεκομμένες γραμμές φανερώνουν την πορεία από τον A στο C μέσω του κόμβου B. Η διακεκομμένη γραμμή φανερώνει μία εναλλακτική πορεία προς τον κόμβο C, η οποία μπορεί να είναι είτε απευθείας, είτε μέσω ενδιάμεσων άλλων κόμβων που δεν μας ενδιαφέρουν στην παρούσα φάση. Εάν θεωρήσουμε, τώρα, πως ο κόμβος B αποτελεί έναν κόμβο σύγκρουσης, τότε υπάρχει μία σύγκρουση στο δίκτυο μας, μιας και θέλουμε να πάμε από τον κόμβο A στον κόμβο C, αλλά παρόλα αυτά δεν θέλουμε να περάσουμε από τον κόμβο B. Παρακάτω θα δούμε πως αντιμετωπίζονται στην παρούσα μοντελοποίηση αυτού του είδους οι συγκρούσεις.

Στην συγκεκριμένη φάση του αλγορίθμου, συνεπώς, ευρίσκονται τα αντικειμενικά μονοπάτια που ενώνουν κατευθείαν τους επιθυμητούς κόμβους, χωρίς τη μεσολάβηση Κόμβων Σύγκρουσης.

➤ **Εύρεση Αδύνατων Μονοπατιών**

Στο επόμενο βήμα, ελέγχουμε εάν κάποια από τα πειραματικά δεδομένα είναι αδύνατον να αποτυπωθούν στο τελικό δίκτυο, μιας και βάσει της υπάρχουσας τοπολογίας οι δύο κόμβοι (Κόμβος Έναρξης – Κόμβος Πέρατος) δεν επικοινωνούν καθόλου (δεν υπάρχουν ενδιάμεσοι κόμβοι να τους ενώνουν).

Πιο συγκεκριμένα, στο παράδειγμα δικτύου της εικόνας 71, σε συνδυασμό με τα πειραματικά δεδομένα της εικόνας 72, παρατηρούμε πως η επιθυμητή εξάρτηση *tnfa* → *shp2* είναι αδύνατον να επιτευχθεί μέσω καμίας δυνατής ακολουθίας κόμβων. Στην παρακάτω εικόνα παρουσιάζεται πιο μεγεθυμένο το συγκεκριμένο κομμάτι της τοπολογίας, ώστε να μπορέσει ο αναγνώστης να το κατανοήσει καλύτερα.



Εικόνα 75: Παράδειγμα αδυναμίας ένωσης κόμβων μέσω διαδρομής

Στην παραπάνω εικόνα φαίνεται ξεκάθαρα πως αν και τα πειραματικά δεδομένα αναφέρουν σύνδεση μεταξύ των δύο κόμβων, η τοπολογία δεν επιβεβαιώνει κάτι τέτοιο. Γι' αυτό και άλλωστε οι δύο κόμβοι βρίσκονται και στο ίδιο επίπεδο (παράλληλη σειρά).

Στο σημείο αυτό, λοιπόν, ο αλγόριθμος βρίσκει τα αδύνατα μονοπάτια και τα απομονώνει. Κρατάει μόνο τα μονοπάτια που υπάρχουν εν δυνάμει στην τοπολογία του δικτύου.

### ➤ Εύρεση Εναλλακτικών Μονοπατιών

Μέχρι τώρα, έχουμε ήδη βρει τα αντικειμενικά μονοπάτια και τα μη δυνατά μονοπάτια. Παρόλα αυτά, η διαδικασία εύρεσης μονοπατιών δεν έχει ολοκληρωθεί. Τα αντικειμενικά μονοπάτια που έχουν βρεθεί μέσω της μεθόδου του Warshall για κάποια από τα δεδομένα μας περιέχουν κάποιους Κόμβους Σύγκρουσης. Παρόλα αυτά, πρέπει να ξεκαθαρίσουμε πως υπάρχει περίπτωση να επιτύχουμε την επιθυμητή διαδρομή για τα δεδομένα μας, μέσω κάποιας άλλης εναλλακτικής διαδρομής. Από τον αλγόριθμο του Warshall έχει ήδη δημιουργηθεί ένα μητρώο, έστω *distance*, το οποίο περιέχει τις αποστάσεις όλων των κόμβων του δικτύου προς τους υπόλοιπους κόμβους. Στη παρούσα φάση, επειδή η φύση του εξεταζόμενου προβλήματος δεν έχει διαστασιολογική υπόσταση, θεωρούμε την απόσταση από κάθε κόμβο στον αμέσως επόμενο μοναδιαία. Έτσι στο παράδειγμα της εικόνας 74, η απόσταση από τον κόμβο A στον κόμβο B ισούται με μονάδα, ενώ η απόσταση από τον κόμβο A στον κόμβο C ισούται με δύο.

Έχοντας, λοιπόν, το παραπάνω μητρώο, μπορούμε μέσω των θεμελιωδών αλγορίθμων **Αναζήτησης κατά Πλάτος (Depth First Search)** και **Αναζήτησης κατά Βάθος (Breadth First Search)** που αναφέρθηκαν στην ενότητα 2.5.2ν να προσπελάσουμε το δίκτυό μας και να βρίσκουμε κάθε φορά εναλλακτικούς τρόπους να φτάσουμε από τον κόμβο έναρξης στον επιθυμητό κόμβο στόχο. Με διαδοχικούς αλγορίθμους αναζήτησης μπορούμε συνεπώς να βρίσκουμε εναλλακτικές διαδρομές και κάθε φορά να ελέγχουμε εάν σε κάποια από αυτές τις διαδρομές υπάρχει κάποιος Κόμβος Σύγκρουσης, ανάλογα με το πώς τους έχουμε ορίσει. Με αυτόν τον τρόπο εξασφαλίζουμε τη σύνδεση επιπλέον επιθυμητών κόμβων.

*Να σημειωθεί σε αυτό το σημείο πως στην τελική παρουσίαση του παραγόμενου δικτύου συμπεριλαμβάνονται όλα τα εναλλακτικά μονοπάτια για την ίδια διαδρομή, ώστε το εξαγόμενο δίκτυο να περιλαμβάνει όσο το δυνατόν περισσότερη πληροφορία σχετικά με το αρχικό δίκτυο και τοπολογία.*

Εάν τώρα και σε αυτήν την περίπτωση για κάποιο συγκεκριμένο συνδυασμό κόμβων έναρξης – κόμβων πέρατος είναι πιθανών να μην υπάρχει ξανά κάποιο μονοπάτι ανάμεσά τους. Στην περίπτωση αυτή επανερχόμαστε στην παραπάνω περίπτωση των αδύνατων μονοπατιών, όπως φαίνεται και στην εικόνα 75.

➤ **Εύρεση - Παρουσίαση Κόμβων Σύγκρουσης**

Έχοντας βρει και παρουσιάσει όλα τα πιθανά μονοπάτια που οδηγούν από τους επιθυμητούς κόμβους έναρξης στους επιθυμητούς κόμβους πέρατος, χωρίς τη μεσολάβηση κάποιων κόμβων σύγκρουσης, οδηγούμαστε στο τελικό στάδιο της προσέγγισής μας, η οποία αφορά την εύρεση και την παρουσίαση των περιπτώσεων αυτών στις οποίες μπορούμε να φτάσουμε στους επιθυμητούς κόμβους πέρατος της σηματοδότησης, μόνο μέσω κάποιου Κόμβου Σύγκρουσης (δεν υπάρχει άλλη εναλλακτική λύση).

Κατά συνέπεια, στο σημείο αυτό ο αλγόριθμος επιτρέπει στο χρήστη να αποκτήσει μία πλήρη εικόνα για το συνολικό αριθμό κόμβων σύγκρουσης που υπάρχει στο δίκτυο και τη συγκεκριμένη τοπολογία τους.

➤ **Ελαχιστοποίηση Συγκρούσεων**

Στο σημείο αυτό συμπυκνώνεται όλη η ουσία της διπλωματικής αυτής. Το πλάνο από την αρχή ήταν η εύρεση μίας προσέγγισης πιο αλγοριθμικής – προγραμματιστικής, η οποία θα στόχευε στην ελαχιστοποίηση των συγκρούσεων σε ένα δίκτυο, χωρίς όμως τη χρήση κάποιας διαδικασίας βελτιστοποίησης.

*Πρέπει να σημειώσουμε πως αυτό που πραγματώνεται είναι η δημιουργία ενός βέλτιστου τελικού δικτύου, το οποίο δεν θα έχει προκύψει από την αποκοπή των περιπτώων αντιδράσεων, αλλά από την εξ' αρχής δημιουργία του δικτύου με τη σταδιακή εισαγωγή και πρόσθεση αντιδράσεων.*

Η πρωτοτυπία του αλγορίθμου αυτού (εάν θα μπορούσαμε με επιφύλαξη να χρησιμοποιήσουμε τη λέξη αυτή) είναι **ότι εξετάζει στο κάθε μονοπάτι μεταξύ των κόμβων του κάθε ζευγαριού των πειραματικών δεδομένων (dependencies), πόσες συγκρούσεις (conflicts) υπάρχουν, αλλά και πόσοι κόμβοι εξάρτησης (dependencies) εξυπηρετούνται. Αν ο αριθμός των συγκρούσεων (conflicts) είναι μεγαλύτερος, το μονοπάτι, καθώς και το συγκεκριμένο ζευγάρι εξάρτησης (dependency pair), απορρίπτονται από το τελικό δίκτυο. Αν ο αριθμός των συγκρούσεων (conflicts) είναι ίσος ή μικρότερος, τότε θεωρούμε πως το συγκεκριμένο μονοπάτι πρέπει να συμπεριληφθεί στο τελικό δίκτυο, έτσι ώστε να έχουμε και όσο δυνατόν περισσότερη πληροφορία της αρχικής τοπολογίας στο τελικό δίκτυο.**

Με μία πρώτη ματιά, η συγκεκριμένη μέθοδος φαίνεται ιδιαίτερα απλή και ίσως κάπως αναξιόπιστη, μιας και δεν βασίζεται στην οργάνωση και στην εκτέλεση κάποιου μαθηματικού μοντέλου μοντελοποίησης, αλλά ουσιαστικά στην κοινή λογική. Παρόλα αυτά, όπως θα φανεί και στην παρουσίαση των αποτελεσμάτων στην παρακάτω ενότητα, είναι ιδιαίτερα αποτελεσματική, αξιόπιστη και υπολογιστικά γρήγορη, ειδικά για μικρής κλίμακας δίκτυα.

Στην παρακάτω ενότητα θα δούμε επακριβώς την εφαρμογή της μεθόδου αυτής και την ορθότητα των αποτελεσμάτων της.

### ➤ Εξαγωγή – Παρουσίαση Αποτελεσμάτων

Σαν τελευταίο βήμα ο αλγόριθμος εξάγει τα αποτελέσματα (πρόκειται για ένα μητρώο που περιέχει όλες τις αντιδράσεις του τελικού δικτύου). Τα αποτελέσματα αυτά με κατάλληλη επεξεργασία από τον κώδικα, εξάγονται και σε αρχείο μορφής .dot.

Έπειτα, ακολουθεί η οπτικοποίησή των με τη χρήση των λογισμικών Graph Visualization Software (GraphViz) και Cytoscape. Επίσης, για την πιο ευκρινή και όμορφη παρουσίαση των χρησιμοποιήθηκε το λογισμικό Adobe Illustrator CS6.

## 3.4 Σύγκριση Μεθόδων

Κρίνεται ιδιαίτερα αναγκαίο στην ενότητα αυτή να πραγματοποιήσουμε μία σύντομη σύγκριση των δύο μεθόδων: της βελτιστοποίησης μέσω χρήσης του μοντέλου Γραμμικού Ακέραιου Προγραμματισμού και της αλγοριθμικής προσέγγισης που πραγματεύεται η παρούσα διπλωματική εργασία.

Σε ότι αφορά το μοντέλο προγραμματισμού ILP, η μοντελοποίηση βασίζεται στην πραγματική επίδραση των συμβάντων φωσφορυλίωσης, προσεκτικά εξαπλωμένη στο δίκτυο σηματοδότησης. Από υπολογιστικής άποψης, η προσέγγιση με ILP παρουσιάζει μία πιο γρήγορη και πιο αποδοτική βελτιστοποίηση μονοπατιών σε σχέση με τους σύγχρονους αλγορίθμους και μπορεί να αναγνωρίσει τις βασικές επιδράσεις των φαρμάκων, καθώς και άγνωστες επιδράσεις στους τομείς των μονοπατιών που περιορίζονται μεταξύ των ενεργών υποδοχέων και των μετρούμενων φωσφορυλιωμένων πρωτεϊνών.

Αυτή η προσέγγιση δημιουργείται σύμφωνα με τη μοντελοποίηση κατά Boole, όπως περιγράφηκε εκτενώς σε παραπάνω ενότητα. Η ILP μοντελοποίηση λύνεται με τη χρήση τυποποιημένων εμπορικών πακέτων λογισμικού με εγγυημένη αποδοτικότητα (εντός μίας ορισμένης από το χρήστη μικρής, αριθμητικής ανοχής). Για την αξιολόγηση των επιδράσεων των φαρμάκων, τα κύτταρα υπόκεινται στο ίδιο ερέθισμα υπό την παρουσία διάφορων φαρμάκων και ανιχνεύουμε τις μεταλλάξεις των ιδίων γεγονότων φωσφορυλίωσης. Μετά, επαναεφαρμόζεται το μοντέλο ILP, χωρίς κάποια εκ των προτέρων υπόθεση του φάρμακου-στόχου και παρακολουθούνται οι αλλαγές στην τοπολογία του μονοπατιού με ή χωρίς την παρουσία φαρμάκων.

Συνεπώς, η χρήση του μοντέλου ILP χρησιμοποιείται ευρέως με ιδιαίτερα αποτελεσματικά και σημαντικά αποτελέσματα στη μοντελοποίηση και στη δημιουργία βιολογικών δικτύων. Μιας και πρόκειται για γραμμική μέθοδο προγραμματισμού, ο χρόνος περάτωσης της είναι σχετικά γρήγορος και μιας και τα περισσότερα φαινόμενα στο φυσικό κόσμο διέπονται από γραμμικότητα

συμπεριφοράς, η επιλογή της για την μοντελοποίηση των βιολογικών φαινομένων καθίσταται επιβλητική.

Ένα ακόμα πλεονέκτημα της μεθόδου αυτής είναι, επίσης, το ότι επειδή βασίζεται σε μαθηματικά μοντέλα εξισώσεων, με μεταβλητές, περιορισμούς και αντικειμενικές συναρτήσεις μπορεί να χρησιμοποιηθεί σε όλα τα είδη δικτύων, ανεξαρτήτως μεγέθους (αριθμός βιολογικών αντιδράσεων) με την ίδια αποτελεσματικότητα.

Παρόλα αυτά, ακριβώς για τους λόγους που εκτέθηκαν στην παραπάνω παράγραφο αποτελεί, όπως και να το κάνουμε, μία ιδιαίτερα δαπανηρή, από άποψη υπολογιστικού χρόνου και υπολογιστικών δυνατοτήτων, μέθοδο.

Από την άλλη μεριά, η παρούσα μέθοδος φαίνεται ιδιαίτερα απλή στην σύλληψη και υλοποίησή της, παρόλα αυτά η διαμόρφωση της και ο αλγοριθμικός προγραμματισμός της ήταν σημαντικά επίμονος και δύσκολος. Βασίζεται αρκετά στην ευχέρεια του εκάστοτε προγραμματιστή και στην διεκπεραίωση θεμελιωδών αλγορίθμων δικτύων.

Για τη μοντελοποίηση μεγάλης κλίμακας δικτύων, ίσως δεν είναι ακόμα αρκετά ενδεικνυόμενη, μιας και δεν έχει δοκιμαστεί σε αρκετά πειραματικά δεδομένα για να μπορούμε ακόμα να αποφανθούμε. Ας μην ξεχνάμε πως για μεγάλο όγκο δεδομένων η εκτέλεση πολλών από των θεμελιωδών αλγορίθμων δικτύων που αναφέρθηκε στην ενότητα 2.5.2, απαιτεί πολύ χρόνο. Παρόλα αυτά, για μικρής κλίμακας δίκτυα, η εκτέλεση του είναι σημαντικά γρήγορη (< 10'). Έτσι γλυτώνει το χρήστη από σημαντικό υπολογιστικό χρόνο.

Με μία πρώτη ματιά η ορθότητα των αποτελεσμάτων της για μικρής κλίμακας δίκτυα είναι αρκετά ικανοποιητική, όπως μπορεί να παρατηρήσει και ο αναγνώστης καλύτερα στην επόμενη ενότητα. Η αλήθεια είναι πως δεν έχει ακόμα δοκιμαστεί σε πολλά δεδομένα, κυρίως μεγάλου μεγέθους, αποτελεί, όμως, μία πρώτη αξιολογη προσπάθεια. Τέλος, ένα ακόμα σημαντικό μειονέκτημα της μεθόδου, είναι πως ακριβώς λόγω έλλειψης του μαθηματικού υποβάθρου της, δεν υπάρχει κάποιος παράγοντας να μας υποδεικνύει το σφάλμα ταιριάσματος μεταξύ των πειραματικών δεδομένων και της αρχικής δεδομένης τοπολογίας.

Μία πρόταση θα μπορούσε να είναι η χρήση της εξεταζόμενης μεθόδου στα μικρής κλίμακας δίκτυα και η χρήση του μοντέλου ILP στα μεγάλα μεγέθους δίκτυα.

## 4. Αποτελέσματα

Στην παρούσα ενότητα θα γίνει η παρουσίαση των εξαγόμενων αποτελεσμάτων, όπως προέκυψαν από τον αναπτυσσόμενο αλγόριθμο. Η οπτικοποίησή των, όπως αναφέρθηκε και παραπάνω θα πραγματοποιηθεί με τα λογισμικά λογισμικών Graph Visualization Software (GraphViz) και Cytoscape. Επίσης, για την πιο ευκρινή και όμορφη παρουσίαση των χρησιμοποιήθηκε το λογισμικό Adobe Illustrator CS6.

Στην αρχή θα παρατίθεται η αρχική τοπολογία του δικτύου και κάποια βασικά σχόλια επί αυτής και έπειτα θα παρουσιάζεται το εξαγόμενο τελικό μοντέλο του, με τις βέλτιστες διαδρομές μεταγωγής των σημάτων.

### 4.1 Παρουσίαση αποτελεσμάτων

Στην επόμενη εικόνα παρουσιάζεται ένα πρώτο δίκτυο αντιδράσεων (αρχική τοπολογία).

Το συγκεκριμένο δίκτυο μπορούμε να πούμε πως είναι ένα δίκτυο, σχετικά μικρής κλίμακας. Ακολουθούν κάποια πρώτα στατιστικά στοιχεία, όπως προκύπτουν από τον αλγόριθμο

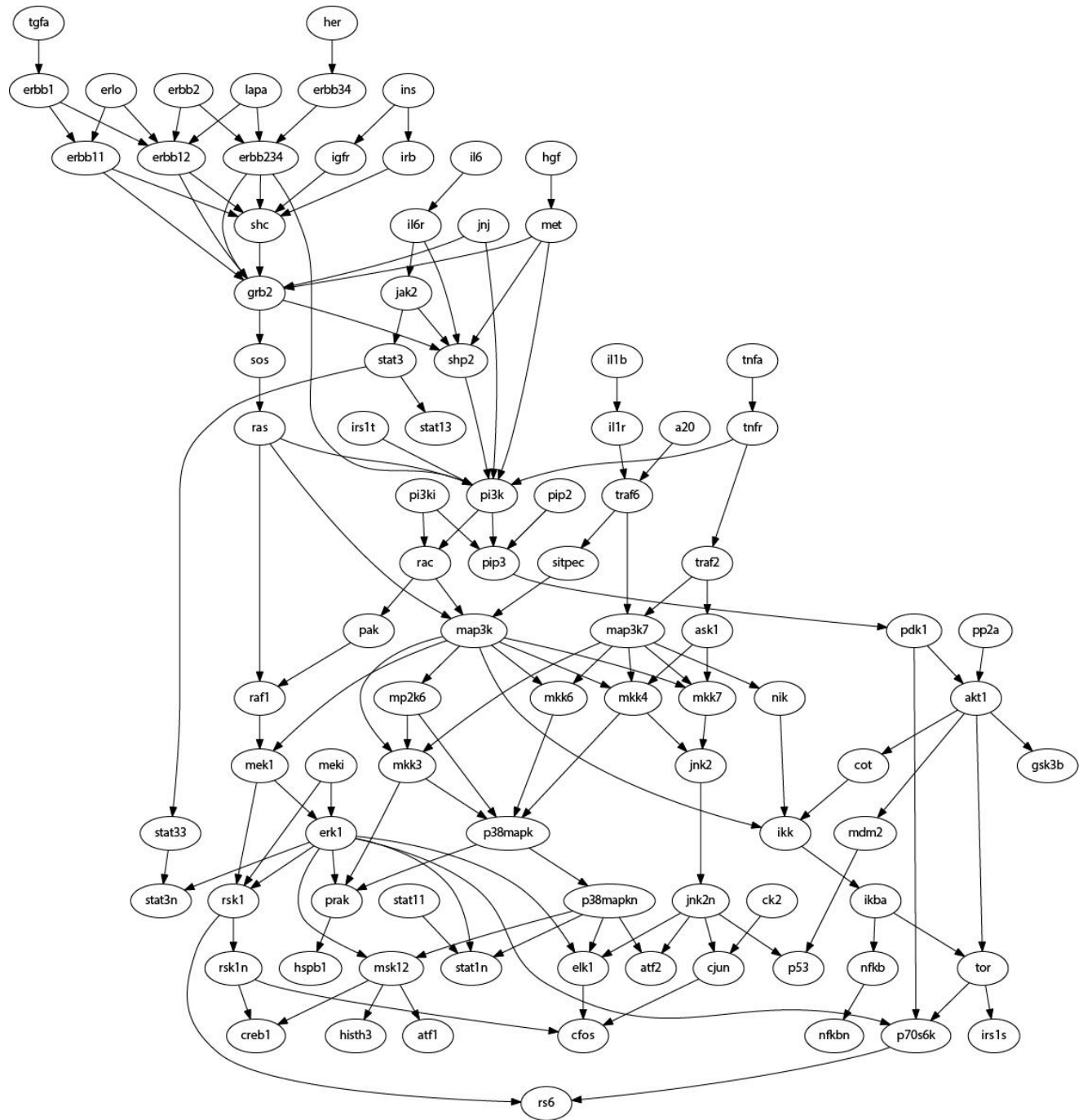
#### Graph Statistical Analysis

---

the total number of reactions is:	139
the total number of reactors is:	76
the total number of produced is:	71
the total number of nodes in graph is:	90
the total number of signals in graph is:	22
the total number of stimuli in graph is:	5
the total number of sources node in graph is:	4
the total number of targets in graph is:	14

---

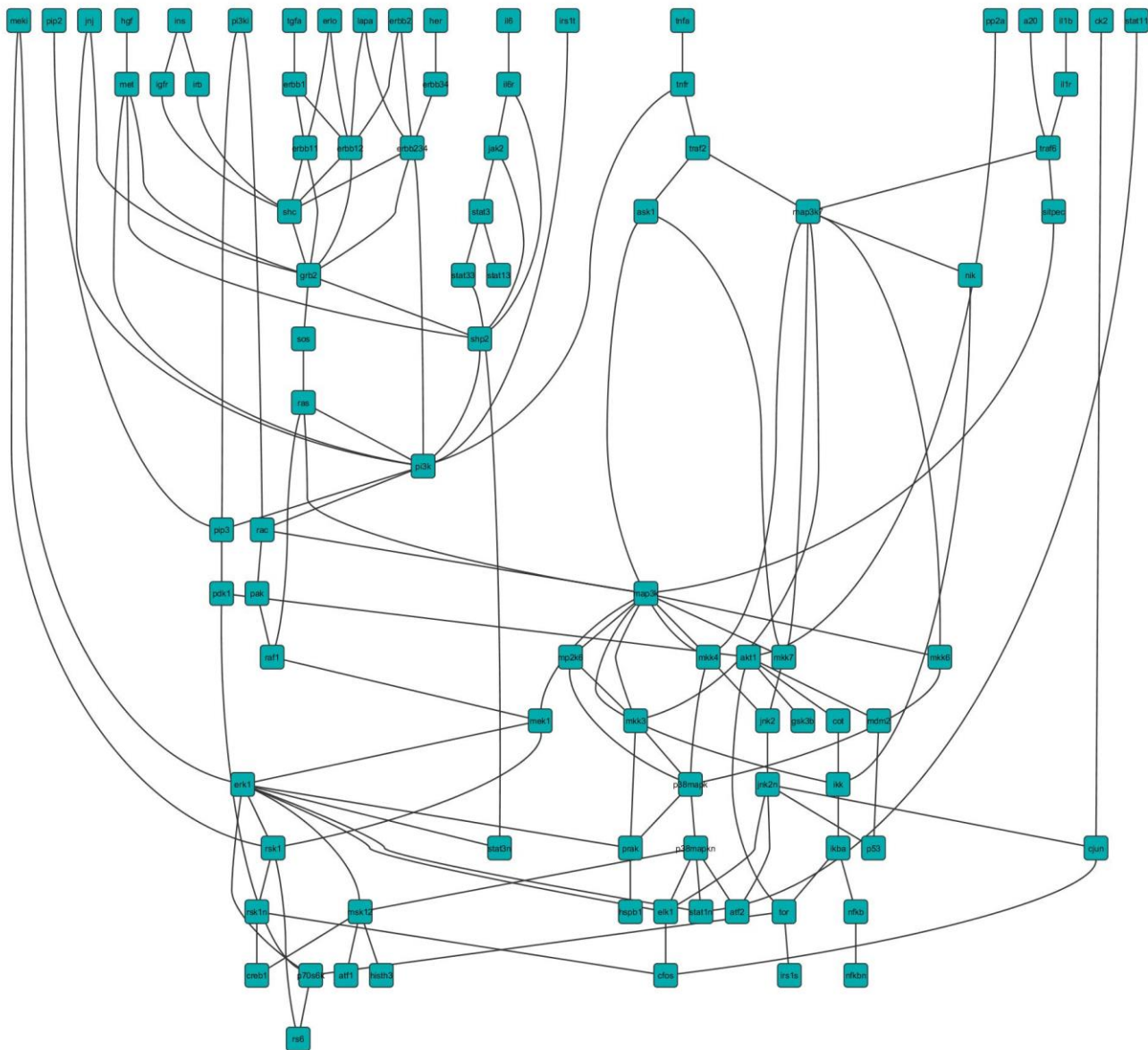




Εικόνα 76: Τοπολογία δικτύου σχετικά μικρής κλίμακας

Πιο συγκεκριμένα το παραπάνω δίκτυο είναι ένα προσανατολισμένο δίκτυο. Αυτό σημαίνει πως η φορά κάθε βέλους υποδηλώνει ουσιαστικά τη φορά κάθε αντίδρασης. Αυτό είναι λογικό μιας και στα βιολογικά φαινόμενα η φορά των βελών έχει μεγάλη σημασία για την διεκπεραίωση της κάθε αντίδρασης και της ερμηνείας τους.

Με τη βοήθεια του λογισμικού Cytoscape παρακάτω παρουσιάζεται μία ακόμα απεικόνιση του εξεταζόμενου δικτύου.



**Εικόνα 77: Απεικόνιση δικτύου με το λογισμικό Cytoscape**

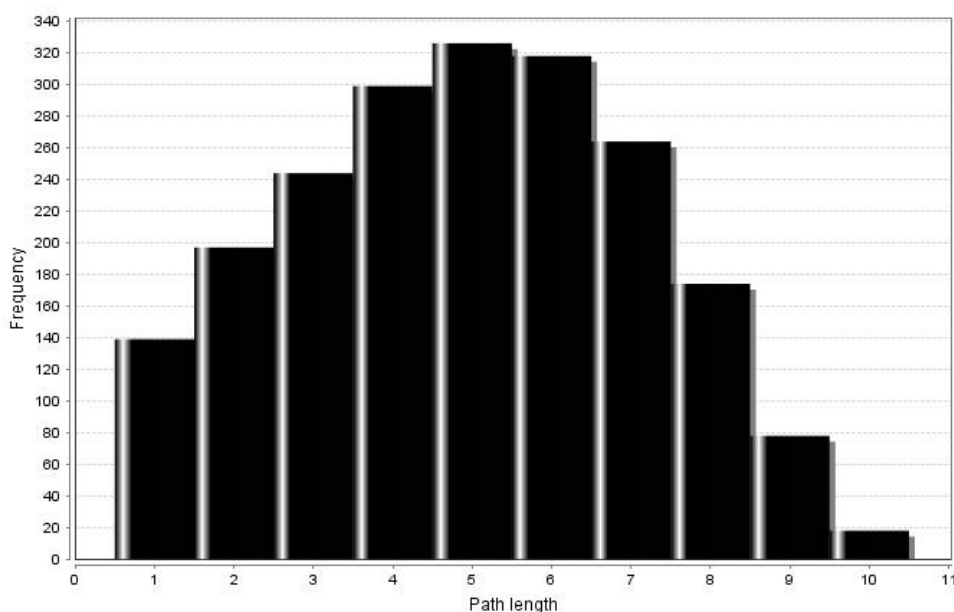
Στην παρακάτω εικόνα παρουσιάζονται κάποια ακόμα στατιστικά του δικτύου, όπως προέκυψαν από την ανάλυσή του μέσω του λογισμικού Cytoscape.

Clustering coefficient : <b>0.026</b>	Number of nodes : <b>90</b>
Connected components : <b>1</b>	Network density : <b>0.0</b>
Network diameter : <b>10</b>	Isolated nodes : <b>0</b>
Network radius : <b>1</b>	Number of self-loops : <b>0</b>
Shortest paths : <b>2057 (25%)</b>	Multi-edge node pairs : <b>0</b>
Characteristic path length : <b>4.920</b>	Analysis time (sec) : <b>0.811</b>
Avg. number of neighbors : <b>3.089</b>	

**Εικόνα 78: Στατιστικά στοιχεία δικτύου**

Στο συγκεκριμένο δίκτυο βλέπουμε πως υπάρχουν 90 κόμβοι, 139 αντιδράσεις μεταξύ πρωτεϊνών, 76 αντιδρώντες ουσίες και 71 παραγόμενες. Από τα πειραματικά δεδομένα βλέπουμε πως κατά τη διάρκεια των μετρήσεων, οι πειραματιστές χρησιμοποίησαν 5 ερεθίσματα και μέτρησαν 22 κόμβους σηματοδότησης.

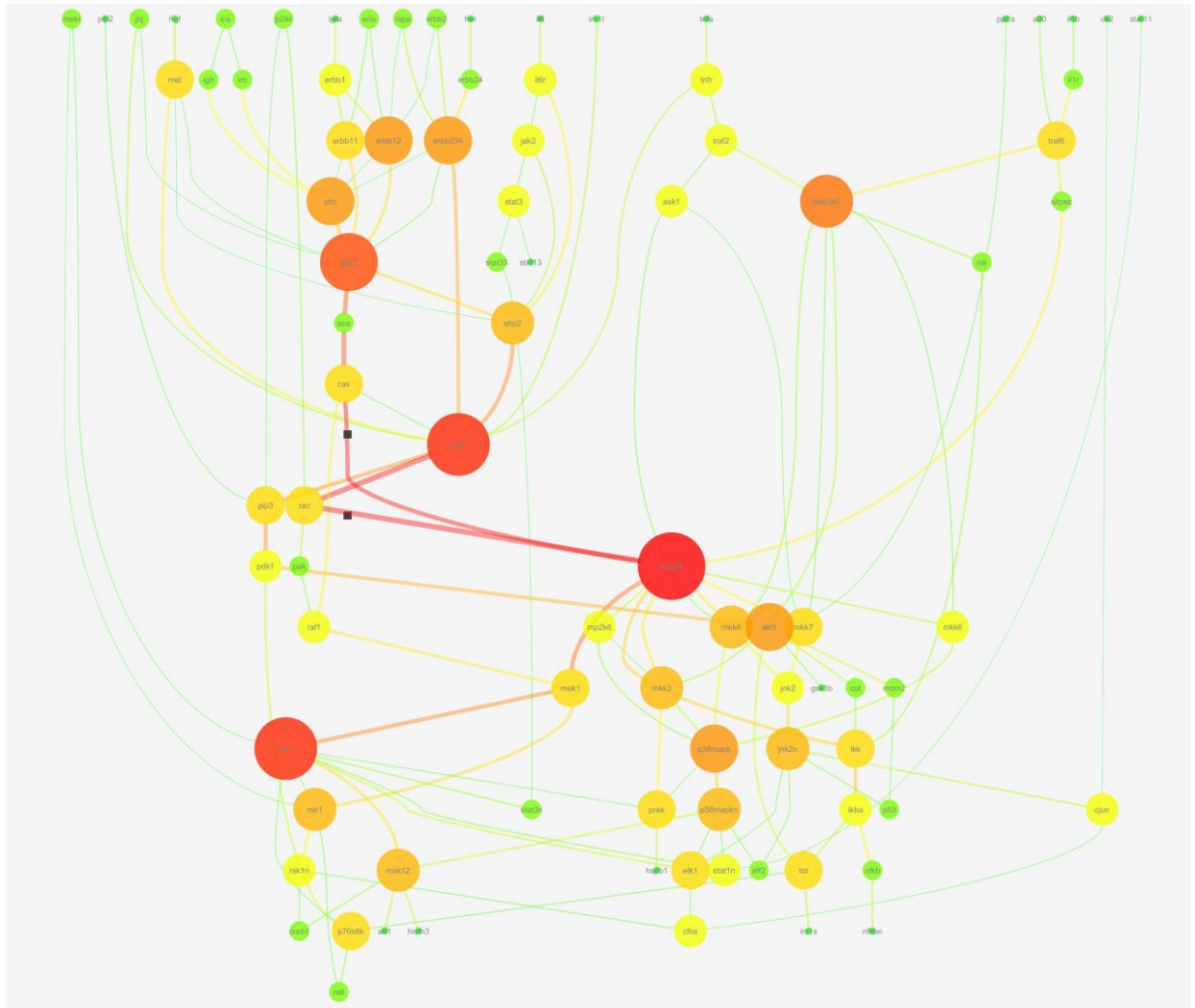
Στην παρακάτω εικόνα μπορούμε επίσης να απεικονισθεί η κατανομή των συντομότερων διαδρομών στο δίκτυο, μεταξύ των κόμβων του. Πιο συγκεκριμένα, απεικονίζεται η συχνότητα εμφάνισης της απόστασης κάθε διαδρομής μεταξύ όλων των πιθανών συνδυασμών διαδρομών μεταξύ κόμβων συναρτήσει της απόστασης κάθε μονοπατιού.



Εικόνα 79: Ιστόγραμμα Shortest Paths

Προχωρώντας την ανάλυση του δικτύου μας ένα βήμα παρακάτω, παρουσιάζουμε μία διαφορετική απεικόνιση του δικτύου μας στην οποία φαίνεται ξεκάθαρα η κατανομή της συγκέντρωσης βελών. Ουσιαστικά φαίνεται ξεκάθαρα ποιοι κόμβοι είναι πιο συχνά παράγωγα στις διάφορες αντιδράσεις του δικτύου.

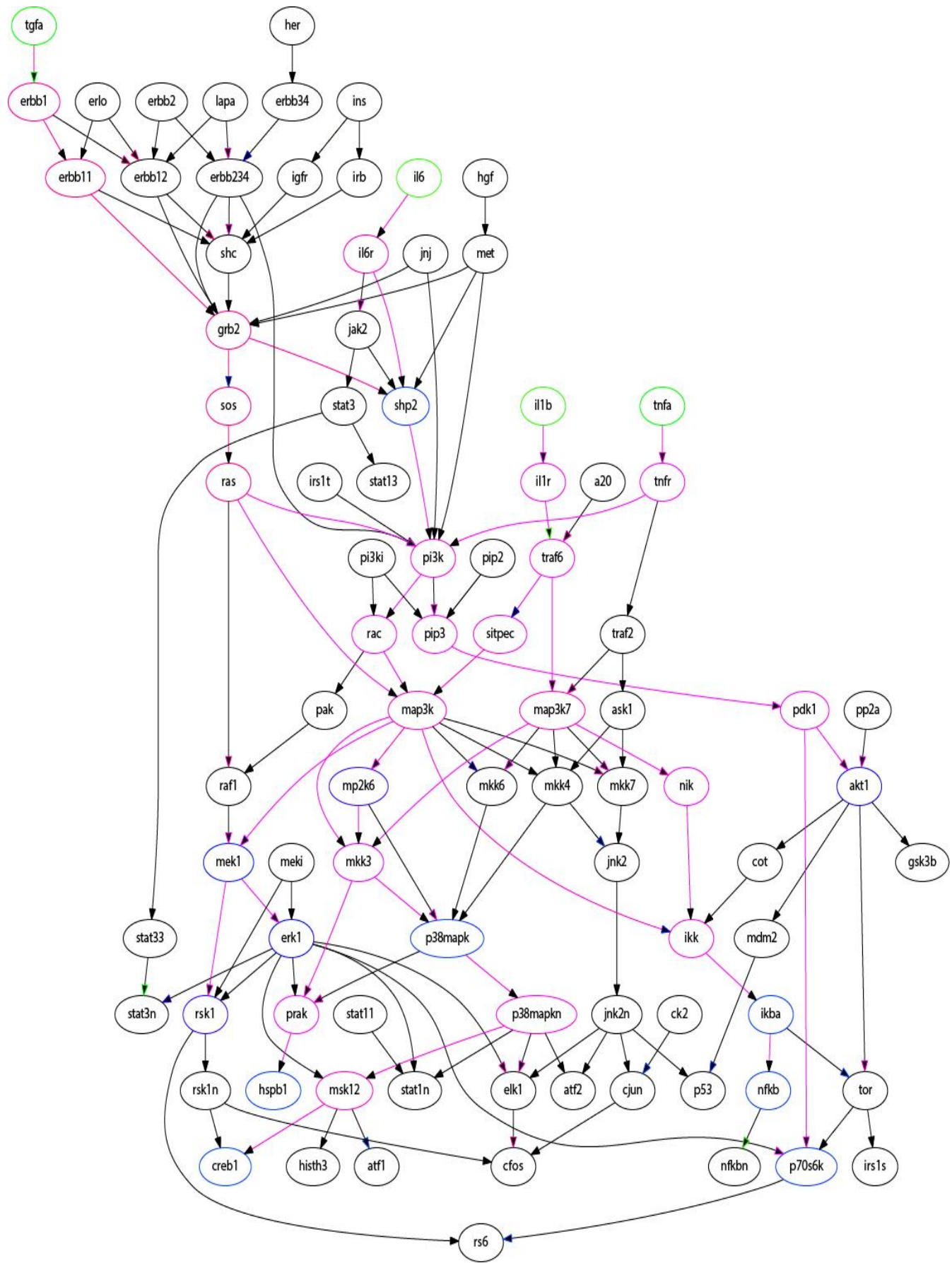
Οι κόκκινοι κόμβοι και οι πορτοκαλί παρουσιάζουν τη μεγαλύτερη συγκέντρωση ακμών, ενώ οι πράσινοι και οι κίτρινοι τη μικρότερη. Όπως είναι λογικό παρατηρούμε τις μεγαλύτερες συγκεντρώσεις στο μεσαίο επίπεδο του δικτύου, ενώ στην οροφή του και στο κατώτερο επίπεδό του κατανέμονται οι μικρότερες συγκεντρώσεις.



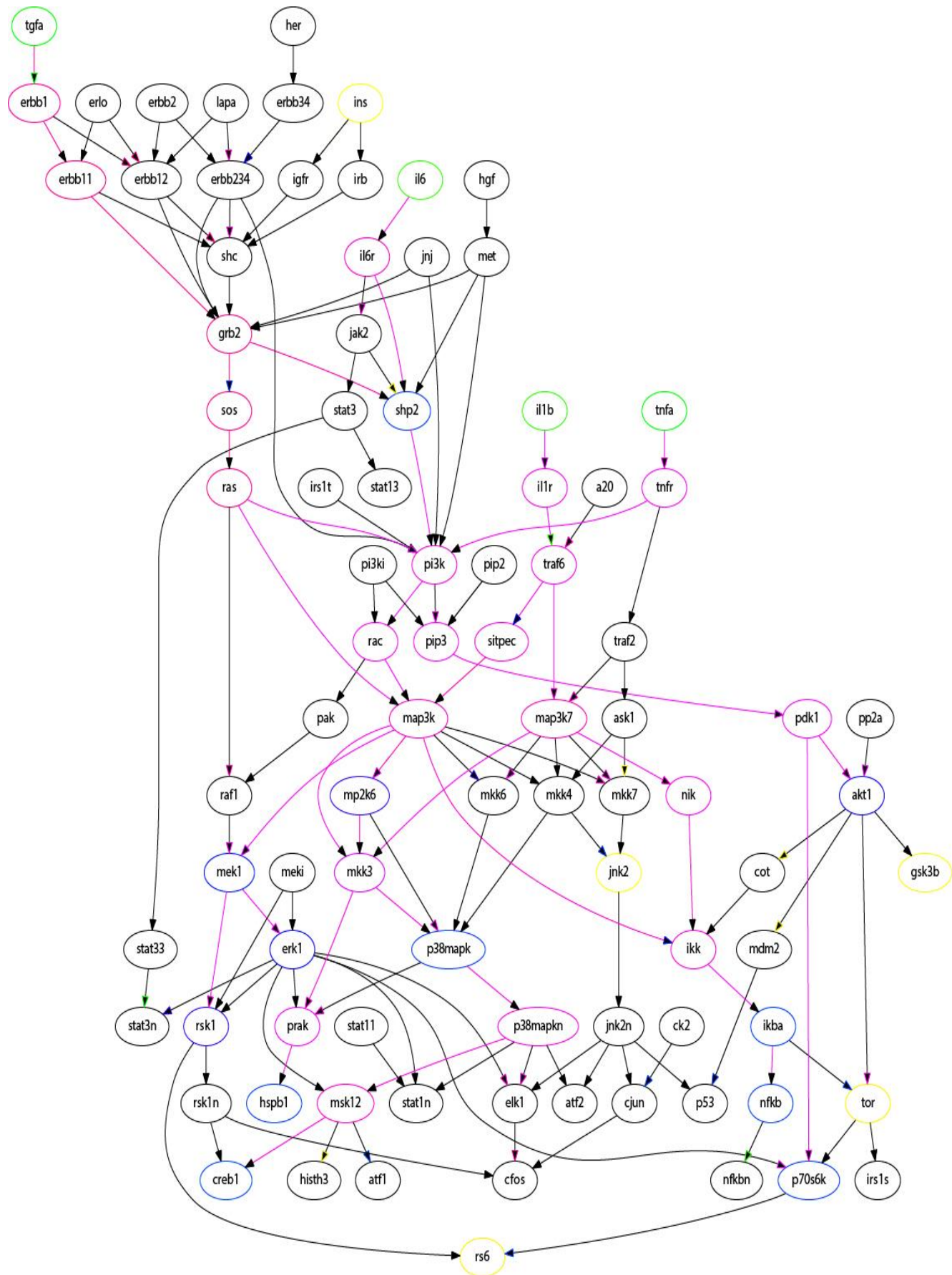
**Εικόνα 80: Απεικόνιση κατανομής ακμών στο δίκτυο**

Έχοντας αναλύσει πλήρως το δίκτυο μας, αφήνουμε τον αλγόριθμο να τρέξει και να παράγει τα τελικά αποτελέσματα του βέλτιστου δικτύου. Τα αποτελέσματα που θα ακολουθήσουν προκύπτουν αφού συνδυάσουμε την αρχική τοπολογία με τα πειραματικά δεδομένα που παρουσιάζονται στην εικόνα 72.

Παρακάτω παρουσιάζεται το τελικό δίκτυο που προκύπτει, στην αρχή μέσα στην αρχική τοπολογία και ύστερα μόνο του.



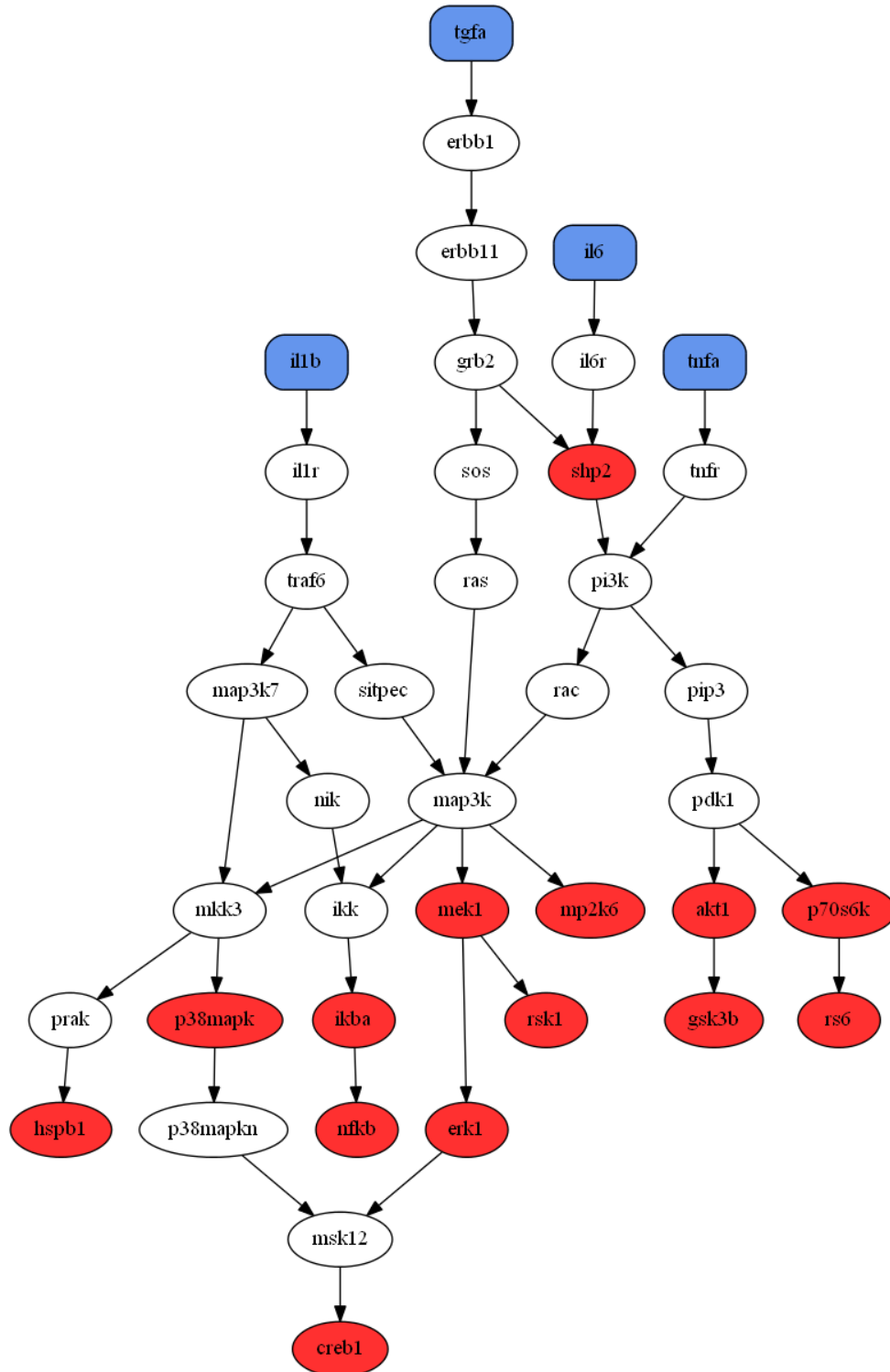
Εικόνα 81: Εξαγόμενο τελικό δίκτυο



Εικόνα 82: Εξαγόμενο τελικό δίκτυο 2

Στα παραπάνω δίκτυα με πράσινο περίγραμμα εμφανίζονται οι κόμβοι από τους οποίους ξεκινάει η σηματοδότηση (ερεθίσματα), οι κόμβοι με το μπλε περίγραμμα είναι οι κόμβοι στους οποίους μετράμε το σήμα, ενώ με κίτρινο χρώμα παρουσιάζονται οι κόμβοι που ενώ ανήκουν στα πειραματικά δεδομένα, εξαιρέθηκαν μετά τη βελτιστοποίηση από το τελικό δίκτυο.

Παρακάτω παρουσιάζεται το τελικό δίκτυο που προέκυψε μόνο του.



Εικόνα 83: Τελικό εξαγόμενο δίκτυο 3

όπου με κόκκινο χρώμα φαίνονται πάλι οι κόμβοι στους οποίους μετρήθηκε το σήμα, ενώ με μπλε τα χρησιμοποιούμενα ερεθίσματα από τους πειραματιστές.



## 4.2 Σχολιασμός Αποτελεσμάτων

Στην ενότητα αυτή θα σχολιάσουμε τα παραπάνω αποτελέσματα σε σχέση πάντα με την αρχική τοπολογία και με τα πειραματικά δεδομένα της εικόνας 72, τα οποία για λόγους πληρότητας παρουσιάζονται ξανά και στην παρακάτω εικόνα.

	hsqb1	akt1	p70s6k	shp2	jnk2	ikba	gsk3b	p38mapk	nfbk	mp2k6	tor	mek1	erk1	rsk1	creb1	rs6
il6	0	0	0	0	0	-1	0	0	1	0	0	0	0	0	0	0
tnfa	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0
il1b	1	0	0	0	0	-1	0	1	1	0	0	1	0	0	1	0
tgfa	0	1	1	1	0	0	1	1	1	1	0	1	1	1	1	1
ins	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
dme	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Εικόνα 84: Πειραματικά Δεδομένα

### Κόμβοι που αποκλείστηκαν από το τελικό δίκτυο

Αρχικά, είναι αναγκαίο να σχολιασθεί η απουσία ορισμένων κόμβων, οι οποίοι όπως φαίνεται και στην εικόνα 82 παρουσιάζονται με κίτρινο περίγραμμα. Πιο συγκεκριμένα όπως είναι δυνατόν να διακριθεί και από τα παραπάνω πειραματικά δεδομένα, τα ερεθίσματα *ins* και *dme* δεν προκαλούν την εμφάνιση μεταβολής σε κανένα σήμα, οπότε δεν υπάρχει λόγος να εμφανίζονται και στο τελικό δίκτυο. Επιπροσθέτως, ο κόμβος *dme* δεν εμφανίζεται καν στην αρχική τοπολογία του δικτύου. Εκ των πραγμάτων, άρα δεν θα μπορούσε να εμφανίζεται και στην τελική.

Για τους κόμβους σημάτων, τώρα:

Αρχικά παρατηρώντας τα παραπάνω πειραματικά δεδομένα φαίνεται πως το σήμα *gnk2* δεν αντέδρασε σε κανένα ερέθισμα από αυτά τα οποία «έριξαν» οι πειραματιστές. Άρα, συνεπάγεται πως δεν υπάρχει λόγος να τοποθετηθεί στο τελικό δίκτυο. Εξάλλου, όπως είναι ευκρινές και από την τελική τοπολογία στην εικόνα 82, ο συγκεκριμένος κόμβος δεν εξυπηρετεί στη μετάδοση καμίας άλλης πληροφορίας, μιας και η προσπέλαση του οδηγεί σε υποσύστημα του δικτύου που δεν συμπεριλαμβάνεται στα πειραματικά δεδομένα.

Σε ό,τι αφορά το σήμα *gsk3b* τώρα. Από τα πειραματικά δεδομένα είναι επίσης ευκρινές πως ο συγκεκριμένος κόμβος συμμετέχει σε **μία μόνο** πληροφορία εξάρτησης (dependency) και πιο συγκεκριμένα φαίνεται πως πρέπει να μεταδοθεί πληροφορία από τον κόμβο *tgfa* στον κόμβο *gsk3b*. Για να γίνει αυτό και βάσει του αλγοριθμικού σκεπτικού που παρουσιάστηκε στην παραπάνω ενότητα 3 και το οποίο ουσιαστικά αντικαθιστά τη διαδικασία της βελτιστοποίησης με τη λογική του ακέραιου γραμμικού προγραμματισμού, για φτάσουμε την πληροφορία αυτή στον αντίστοιχο κόμβο πρέπει να περάσουμε από τα ενδιάμεσα σήματα *shp2* και *akt1*. Ουσιαστικά, δηλαδή, στο άθροισμα

των συγκρούσεων (conflicts) είμαστε ήδη στο -2 για την εξυπηρέτηση μόνο μίας σχέσης εξάρτησης των πειραματικών δεδομένων (dependency). Άρα, το τελικό άθροισμα είναι:

$$\mathbf{Final\_sum = -2 (conflicts) + 1 (dependency) < 0}$$

Άρα, σωστά ο αλγόριθμος «έπραξε» και απέκλισε το συγκεκριμένο κόμβο μίας και η επιλογή του τελικά δεν μας συμφέρει.

Εξετάζοντας τώρα το σήμα **tor** μπορούμε εύκολα να διακρίνουμε από τα πειραματικά δεδομένα πως το συγκεκριμένο σήμα δεν αντέδρασε σε κανένα ερέθισμα από αυτά τα οποία «έριξαν» οι πειραματιστές. Κατά συνέπεια, δεν υπάρχει λόγος αυτός να τοποθετηθεί στο τελικό δίκτυο.

Τέλος, έχουμε τον κόμβο σήματος **rs6**. Ο κόμβος αυτός σύμφωνα με τα πειραματικά δεδομένα υπάρχει σε μία μόνο σχέση εξάρτησης. Πιο συγκεκριμένα, επιθυμούμε να μεταφέρουμε την πληροφορία από τον κόμβο **tgfa** στον κόμβο **rs6**. Για να γίνει αυτό βάσει της συγκεκριμένης τοπολογίας και του τελικού εξαγόμενου δικτύου είναι εύκολο να διακρίνουμε πως για να φτάσουμε στο συγκριμένο κόμβο, πρέπει να περάσουμε από το σήμα **tor**, το οποίο όπως αναφέρθηκε μόλις στην παραπάνω παράγραφο δεν συμμετέχει καθόλου στο τελικό δίκτυο. Ας υποθέσουμε τώρα, πως η αλγοριθμική διαδικασία μας είναι ελλιπής και πως ο κόμβος **tor** θα έπρεπε να είχε τοποθετηθεί στο τελικό δίκτυο. Για να φτάσουμε τώρα στο κόμβο **rs6**, για τη συγκεκριμένη μετάδοση πληροφορίας, θα έπρεπε να περάσουμε από τα σήματα **shp2**, **akt1**, **tor** και **p70s6k**. Για να εξυπηρετήσουμε, δηλαδή, μία **μόνο σχέση εξάρτησης (dependency)** θα έπρεπε να πληρώσουμε και να περάσουμε από **τέσσερις κόμβους σύγκρουσης (conflicts)**. Σύμφωνα με τη μέθοδο μας, θα έχουμε:

$$\mathbf{Final\_sum = -4 (conflicts) + 1 (dependency) < 0}$$

το οποίο σημαίνει πως και πάλι η τοποθέτηση του συγκεκριμένου κόμβου στο τελικό δίκτυο θα ήταν ακατάλληλη.

Θα μπορούσε να σκεφτεί κάποιος, τώρα, πως θα μπορούσαμε να φτάσουμε στον κόμβο **rs6** ακολουθώντας κάποια άλλη εναλλακτική διαδρομή. Το σκεπτικό αυτό είναι σίγουρα σωστό. Και προφανώς παρατηρώντας την αρχική τοπολογία σε συνδυασμό με το εξαγόμενο τελικό μονοπάτι στην εικόνα 82, είναι ευκρινές πως μπορούμε να φτάσουμε τον επιθυμητό κόμβο ακολουθώντας την εναλλακτική διαδρομή:

$$grb2 \rightarrow sos \rightarrow ras \rightarrow map3k \rightarrow mek1 \rightarrow rsk1 \rightarrow rs6$$

Ακολουθώντας ξανά την αλγοριθμική διαδικασία που περιγράφηκε ξανά και ξανά παραπάνω, παρατηρούμε πως για να εξυπηρετηθεί τη **συγκεκριμένη σχέση εξάρτησης (dependency)**, **πρέπει αν περάσουμε από δύο ενδιάμεσα σήματα (conflicts)** και σύμφωνα με την αλγοριθμική διαδικασία:

$$\mathbf{Final\_sum = -2 (conflicts) + 1 (dependency) < 0}$$

Άρα, τελικά ο κόμβος αυτός όντως δεν έπρεπε να τοποθετηθεί στην τελική εξαγόμενη τοπολογία.

### Κόμβοι που εμφανίζονται στο τελικό δίκτυο

Όπως φάνηκε και στα παραπάνω σχόλια η επιλογή των αποκλεισμένων κόμβων τελικώς ήταν αρκετά ορθή. Αλλά τι γίνεται τώρα με αυτούς που επιλέχθηκαν τελικώς για το εξαγόμενο δίκτυο.

Αρχικά οφείλουμε να παρατηρήσουμε πως στο τελικό δίκτυο, κάποιοι κόμβοι σήματα είναι καταληκτικοί, άρα δεν χρειάζεται να σχολιασθεί η ύπαρξη τους μιας και το ότι δεν είναι ενδιάμεσοι, ουσιαστικά δεν τους καθιστά και κόμβους σύγκρουσης (conflicts), απλά κόμβους κατάληξης του σήματος. Θα σχολιάσουμε, συνεπώς, με την ύπαρξη των κόμβων **p38mapk**, **shp2**, **mek1** και **ikba**.

Ας ξεκινήσουμε με το θεμελιώδης σημασίας κόμβο **shp2**. Ο συγκεκριμένος κόμβος, λοιπόν, εμφανίζεται σε δύο σχέσεις εξάρτησης. Από τον κόμβο όμως **tgfa** στον κόμβο **shp2**, όπως έχει ήδη αναφερθεί και στο κεφάλαιο 3, δεν μπορεί να πραγματοποιηθεί μετάδοση σήματος μιας και στην αρχική τοπολογία οι δύο κόμβοι βρίσκονται στην ίδια ευθεία και όπως είναι προφανές δεν υπάρχει καμία σχέση σύνδεσης μεταξύ τους. Η δεύτερη σχέση εξάρτησης, ξεκινάει από τον κόμβο **tgfa**, όπου βλέπουμε πως τελικά το σήμα καταφτάνει. Παρόλα αυτά, θα πρέπει να τονίσουμε πως σε ότι αφορά τον κόμβο **tgfa** σαν αρχικό ερέθισμα και κόμβο έναρξης της σηματοδότησης, το **shp2** δεν λειτουργεί σαν ενδιάμεσος κόμβος σύγκρουσης (conflict), μιας και υπάρχει πάντα και μια εναλλακτική διαδρομή, μέσω των κόμβων **sos** και **ras**. Άρα, η εμφάνιση του συγκεκριμένου κόμβου δεν μας ενοχλεί σε ότι αφορά το ερέθισμα **tgfa** μιας και λειτουργεί σαν καταληκτικός κόμβος της πληροφορίας και η εμφάνισή του από συγκεκριμένης άποψη, απλά εμπλουτίζει την πληροφορία του δικτύου. Άρα, εξυπηρετείται ήδη μία σχέση εξάρτησης πειραματικών δεδομένων (  $final\_sum = +1$  dependency).

Ας δούμε τώρα γιατί θεωρούμε την παρουσία του συγκεκριμένου κόμβου καταλυτική. Από την εικόνα 82 είναι εύκολο να διακρίνουμε πως η ύπαρξη του συγκεκριμένου κόμβου είναι ο μόνος τρόπος επικοινωνίας του ερεθίσματος **il6** με το υπόλοιπο δίκτυο και συνεπώς με τις σχέσεις εξάρτησης του συγκεκριμένου ερεθίσματος. Χωρίς την ύπαρξη του καμία από τις σχέσεις εξάρτησης του **il6** δεν θα μπορούσε να ικανοποιηθεί στην τελική απεικόνιση του δικτύου, κάτι που σημαίνει πως θα χάναμε εξαιρετικά χρήσιμη πληροφορία. Χάρη στην παρουσία του κόμβου **shp2**, λοιπόν, εξυπηρετούνται, βάσει του παραπάνω πίνακα πειραματικών δεδομένων επιπλέον δύο σχέσεις εξάρτησης (  $final\_sum = +1$  dependency (tgfa) +2 dependencies (il6) ).

Συνοψίζοντας, για τον κόμβο **shp2** η αλγοριθμική μας διαδικασία δίνει:

$$final\_sum = +3 (dependencies) -1 (conflict) > 0$$

Ας σχολιάσουμε, τώρα, την ύπαρξη του κόμβου **p38mapk**. Ο συγκεκριμένος κόμβος λειτουργεί σαν καταληκτικός κόμβος πληροφορίας που ξεκινάει από τους κόμβους **il1b** και **tgfa**. Άρα, η παρουσία του εξυπηρετεί, ήδη, δύο σχέσεις εξάρτησης πειραματικών δεδομένων (dependencies) (  $final\_sum = +2$  (dependencies) ) . Παρόλα αυτά, ο συγκεκριμένος κόμβος αποτελεί και κόμβο σύγκρουσης (conflict) κατά τη μετάδοση σήματος από τους ίδιους κόμβους στον κόμβο **creb1** (  $final\_sum = +2$  (dependencies) -1 (conflict) ). Παρόλα αυτά, η ύπαρξη του εξυπηρετεί κατ' επέκταση και τις δύο σχέσεις εξάρτησης αυτές, οπότε σύμφωνα με την αλγοριθμική διαδικασία μας:

$$final\_sum = +4 (dependencies) -1 (conflict) > 0$$

Άρα, η εκλογή του συγκεκριμένου κόμβου φαίνεται τελικά σωστή.

Ας περάσουμε τώρα στον κόμβο **ikba**. Η εμφάνιση του συγκεκριμένου κόμβου εξυπηρετεί δύο σχέσεις εξάρτησης σαν καταληκτικός κόμβος σηματοδότησης, αλλά πιο σημαντικά λειτουργεί σαν το μοναδικό κόμβο σύνδεσης των ερεθισμάτων με τον κόμβο **nfkf**. Άρα, ουσιαστικά εξυπηρετεί συνολικά, +2 dependencies (προς ikba) +4 dependencies (προς nfkb). Άρα, σύμφωνα ξανά με την αλγοριθμική διαδικασία για το συγκεκριμένο κόμβο θα έχουμε:

$$final\_sum = +6 (dependencies) -1 (conflict) > 0$$

Άρα, η εκλογή και του συγκεκριμένου κόμβου φαίνεται τελικά σωστή.

Τέλος, για τον κόμβο **mek1**. Η εμφάνιση του εξυπηρετεί δύο σχέσεις εξάρτησης σαν καταληκτικός κόμβος σηματοδότησης, αλλά επίσης βοηθάει την πληροφορία να μεταδοθεί και στον κόμβο **rsk1** από τον κόμβο **tgfa**. Συνολικά, δηλαδή εξυπηρετούνται τρεις σχέσεις εξάρτησης. Άρα, πάλι σύμφωνα της γνωστής αλγοριθμικής διαδικασίας η εκλογή του συγκεκριμένου κόμβου αποδεικνύεται σωστή.

### Ποσοστό επιτυχίας της μεθόδου

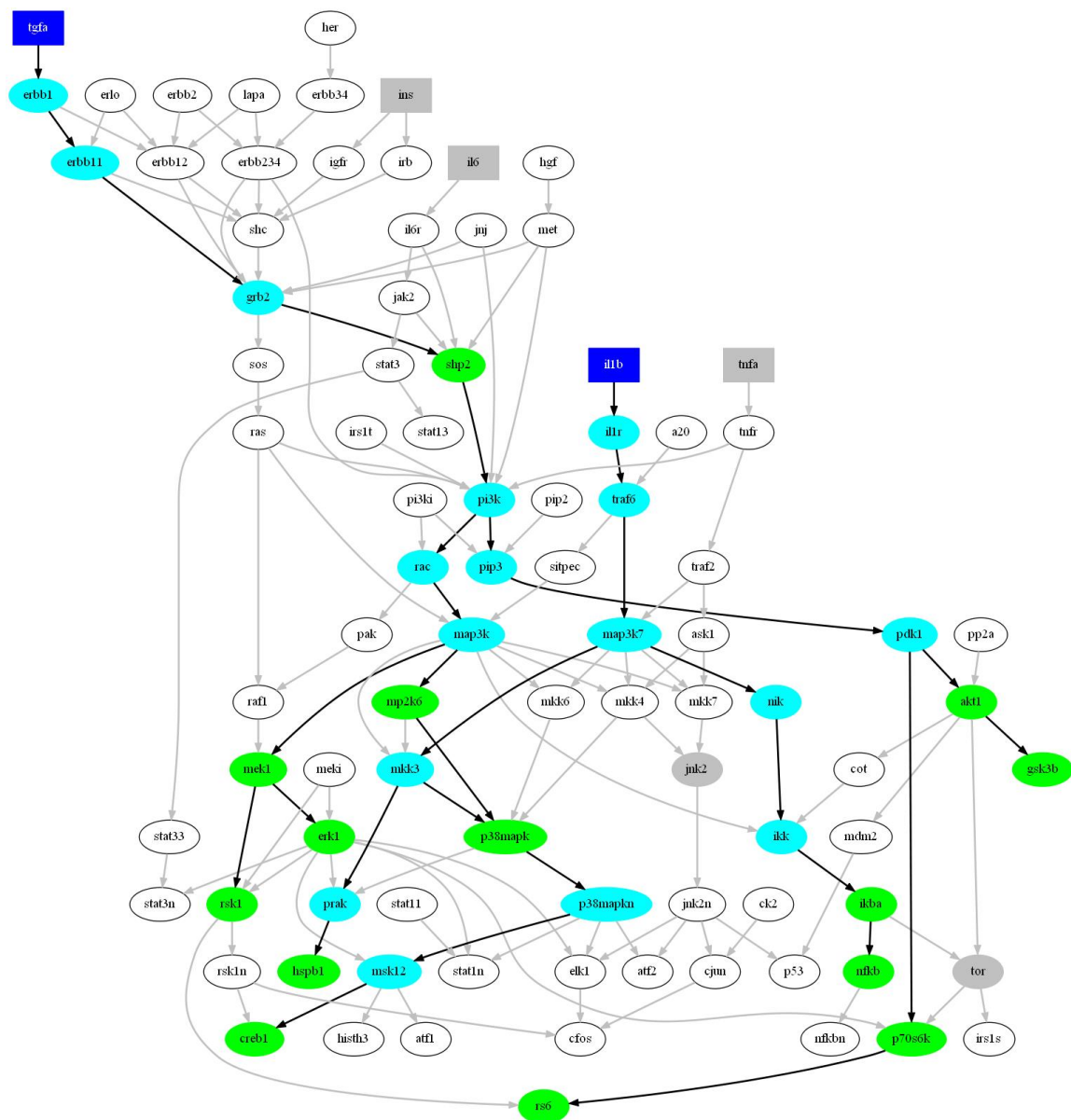
Βλέπουμε, συνεπώς, πως τα αποτελέσματα της μεθόδου μας είναι αρκετά σωστά και λογικά και ανταποκρίνονται και στην αρχική τοπολογία, αλλά και στις σχέσεις εξάρτησης μεταξύ των πειραματικών δεδομένων. Επίσης, πρέπει να σχολιασθεί το γεγονός, ότι από το τελικό δίκτυο και βάσει της **συγκεκριμένης αρχικής τοπολογίας** και των **συγκεκριμένων πειραματικών δεδομένων, μόνο δύο σχέσεις εξάρτησης (dependencies) έμειναν ανεκπλήρωτες**. Το ποσοστό της μεθόδου για το δεδομένο πρόβλημα και δεδομένα, φαίνεται αρκετά ικανοποιητικό:

$$success\_rate = \frac{19dependencies}{21dependencies} \approx 90.48\%$$

### 4.3 Παρουσίαση Αποτελεσμάτων με χρήση της μεθόδου του Ακέραιου Γραμμικού Προγραμματισμού (ILP)

Στην ενότητα αυτή θα παρουσιάσουμε τα αποτελέσματα που προέκυψαν για τα ίδια δεδομένα με χρήση της μεθόδου βελτιστοποίησης Ακέραιου Γραμμικού Προγραμματισμού.

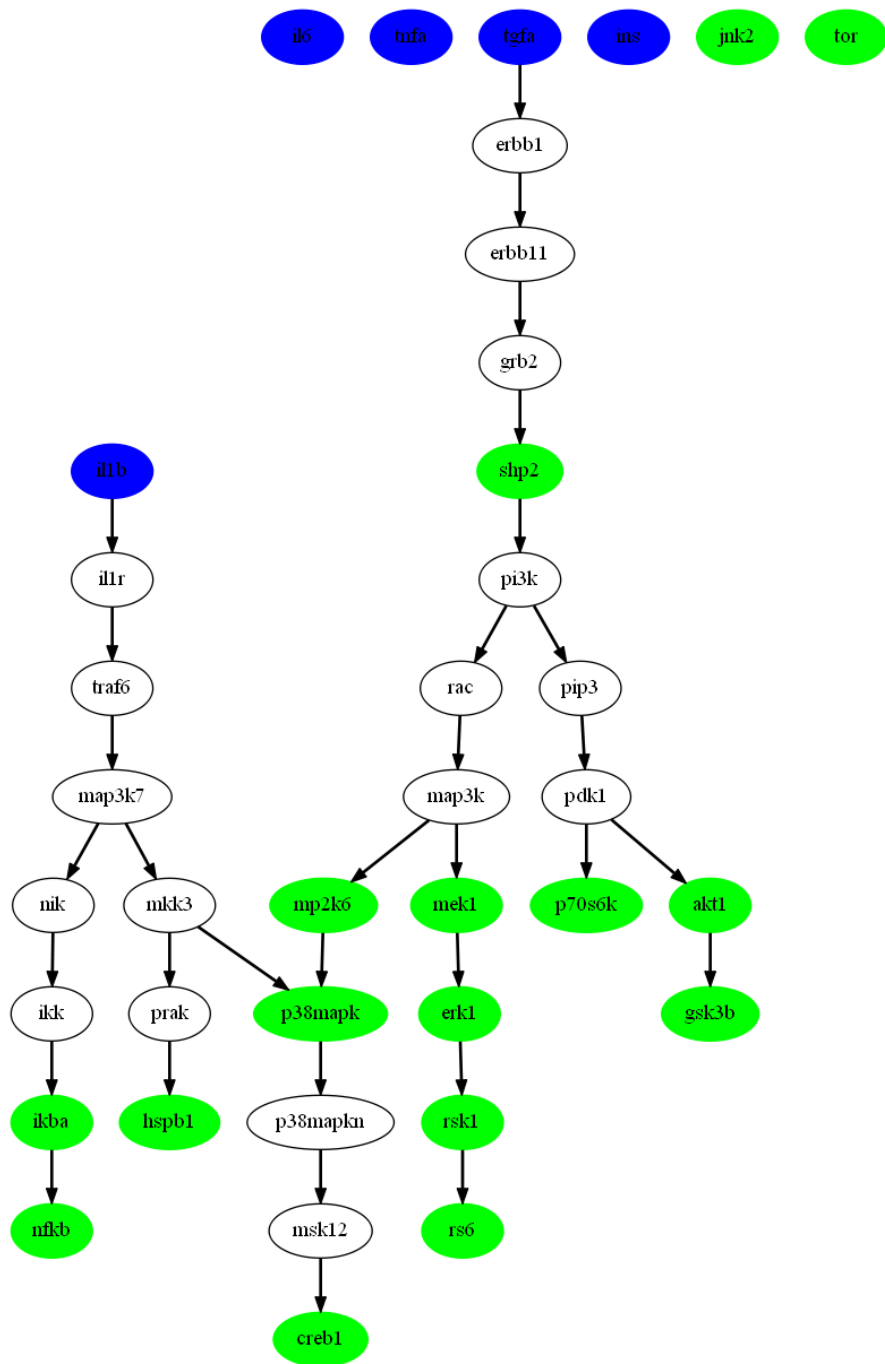
Η αρχική τοπολογία είναι η ίδια (συγκεκριμένα αυτή που απεικονίζεται στην εικόνα 76). Η μέθοδος «έτρεξε» και για τη μεγιστοποίηση, αλλά και για την ελαχιστοποίηση της αντικειμενικής συνάρτησης. Στην παρακάτω εικόνα παρουσιάζεται το τελικό εξαγόμενο δίκτυο σε μία πρώτη απεικόνιση.



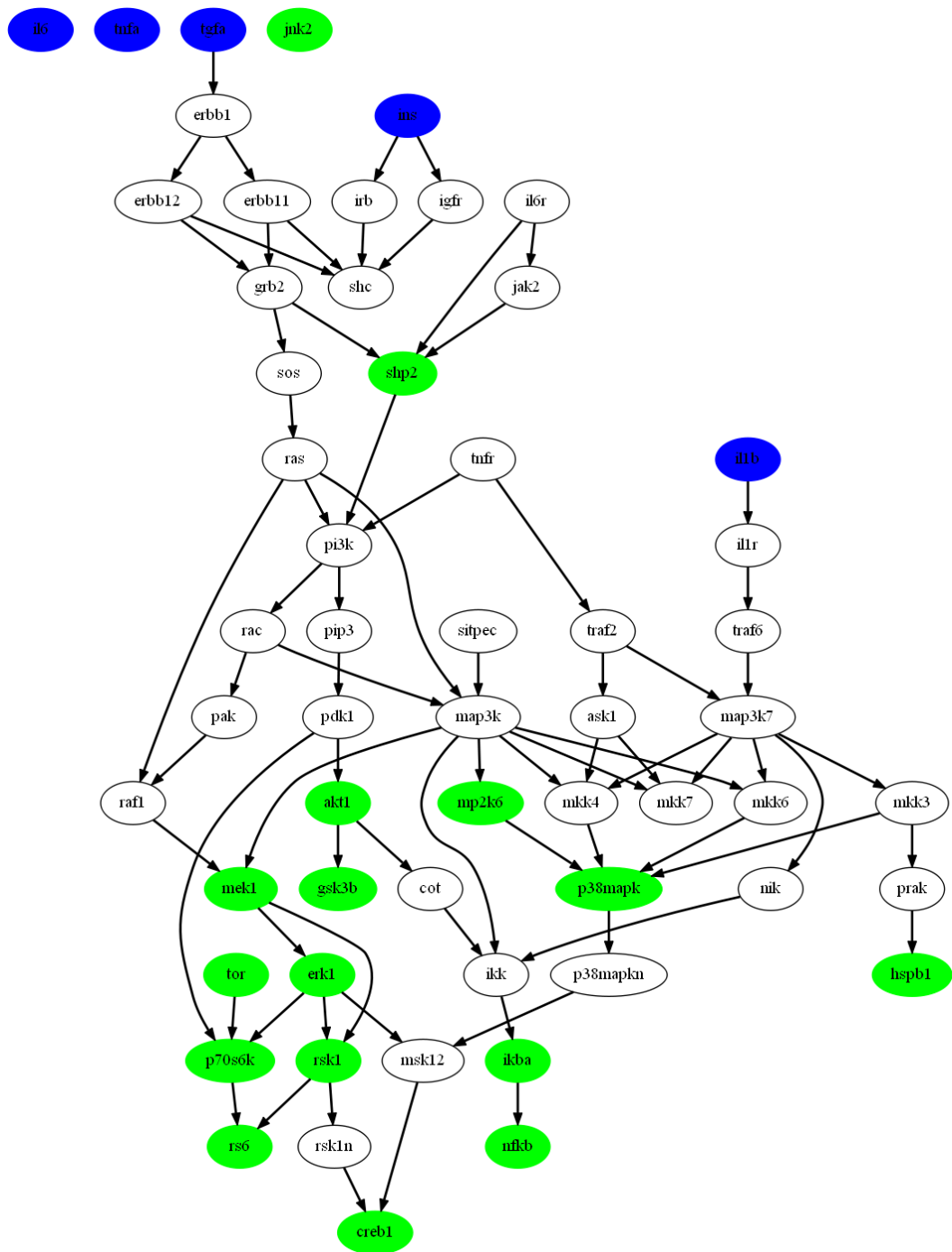
Εικόνα 85: Απεικόνιση τελικού δικτύου με χρήση μεθόδου ILP

Στην παραπάνω εικόνα παρουσιάζονται με πράσινο χρώμα οι κόμβοι μετρούμενα σήματα που συμπεριλήφθηκαν στο τελικό δίκτυο, με γαλάζιο χρώμα οι απλοί κόμβοι και με σκούρο μπλε χρώμα τα ερεθίσματα που έμειναν στο τελικό δίκτυο. Αντίθετα με γκρι χρώμα παρουσιάζονται τα ερεθίσματα και τα σήματα που έμειναν εκτός τελικού δικτύου.

Στην επόμενη εικόνα παρουσιάζεται το τελικό δίκτυο το οποίο προέκυψε με τη χρήση της μεθόδου ILP μία φορά για την περίπτωση ελαχιστοποίησης της αντικειμενικής συνάρτησης και μία για την περίπτωση μεγιστοποίησης της αντικειμενικής συνάρτησης.



Εικόνα 86: Τελικό δίκτυο με ελαχιστοποίηση της αντικειμενικής συνάρτησης



Εικόνα 87: Τελικό δίκτυο με μεγιστοποίηση της αντικειμενικής συνάρτησης

Και στις δύο παραπάνω εικόνες με μπλε χρώμα παρουσιάζονται τα ερεθίσματα, τα οποία χρησιμοποιήθηκαν από τον κάθε πειραματιστή κατά την πειραματική διαδικασία, ενώ με πράσινο χρώμα τα μετρούμενα σήματα.



#### 4.4 Σύγκριση μεθόδων και αποτελεσμάτων

Στην ενότητα αυτή θα παρουσιάσουμε μία πρώτη σύγκριση μεταξύ των αποτελεσμάτων που αποδίδουν οι δύο μέθοδοι, αλλά και μία σύγκριση μεταξύ των δύο μεθόδων.

Ως σύγκριση μεταξύ των αποτελεσμάτων εννοείται ουσιαστικά η ομοιότητα που παρουσιάζουν τα τελικά εξαγόμενα δίκτυα με τη χρήση των διαφορετικών μεθόδων.

Για να πραγματοποιηθεί, η σύγκριση αυτή μεταξύ των αποτελεσμάτων, χρησιμοποιήθηκαν δύο μέθοδοι: Μία πρώτη εκτίμηση, η οποία σε καμία περίπτωση δεν μπορεί να θεωρηθεί επιστημονικά και στατιστικά ακριβής, αλλά πάρα αυτά μπορεί να παρέχει μία πρώτη εικόνα, είναι η χρήση του πηλίκου του αριθμού των τελικών αντιδράσεων της μίας μεθόδου με τον αριθμό των τελικών αντιδράσεων της άλλης μεθόδου.

Η μέθοδος που εξετάζεται στην παρούσα διπλωματική δίνει τελικό δίκτυο 43 αντιδράσεων, ενώ η χρήση του Ακέραιου Γραμμικού Προγραμματισμού δίνει τελικό δίκτυο 33 αντιδράσεων. Σύμφωνα, λοιπόν, με τη μέθοδο αυτή, η ομοιότητα των δύο λύσεων είναι της τάξης του 76,7442%. Αυτό σημαίνει πως οι δύο λύσεις ταυτίζονται κατά ένα ποσοστό περίπου 77% ή ότι το 77% των αντιδράσεων της μίας λύσης υπάρχουν και στην άλλη λύση.

Γενικά, όσο πιο κοντά στη μονάδα είναι αυτός ο αριθμός, τόσο πιο καλή θεωρείται ότι είναι και η ομοιότητα των αποτελεσμάτων. Αυτό συμβαίνει γιατί η προσέγγιση αυτή θεωρεί ότι ο ίδιος αριθμός αντιδράσεων υπονοεί και ίδιες αντιδράσεις, κάτι το οποίο είναι άστοχο μιας και δεν είναι υποχρεωτικό το ότι αν οι δύο λύσεις δίνουν ίδιο αριθμό τελικών αντιδράσεων, αυτές θα είναι και ίδιες αντιδράσεις μεταξύ τους.

Για το λόγο αυτό είναι πιο έγκυρη η χρήση της δεύτερης μεθόδου, η οποία είναι ο Δείκτης Jaccard ή αλλιώς Συντελεστής Ομοιότητας κατά Jaccard (συντελεστής που επινοήθηκε αρχικά από τον Paul Jaccard). Ο συντελεστής αυτός είναι ουσιαστικά ένα στατιστικό μέγεθος το οποίο χρησιμοποιείται για τη σύγκριση της ομοιότητας και της ποικιλομορφίας συνόλων δειγμάτων. Ορίζεται σαν το μέγεθος της τομής των δύο συνόλων διαιρούμενης με την ένωση των δύο συνόλων, δηλαδή:

$$J(A, B) = \frac{|A \cap B|}{|A \cup B|}$$

Σύμφωνα, λοιπόν, με αυτό το μέτρο της τομής των δύο συνόλων ισούται με 30, ενώ το μέτρο της ένωσης των δύο συνόλων ισούται με 43. Από τον τύπο τον παραπάνω, λοιπόν, ο συντελεστής ισούται με  $J(A,B) = 65.2174\%$ , δηλαδή τα εξαγόμενα δίκτυα που προκύπτουν από τις δύο μεθόδους συμπίπτουν μεταξύ τους κατά ένα ποσοστό 65%.

Για τη σύγκριση των μεθόδων ισχύουν όσα έχουν ειπωθεί και στην ενότητα 3.4.

#### 4.5 Σχολιασμός Παραρτήματος

Όπως αναφέρθηκε και παραπάνω στην ενότητα 3, κατά τη διάρκεια της παρούσας διπλωματικής αναπτύχθηκαν συνολικά τέσσερις αλγόριθμοι, οι οποίοι δοκίμαζαν διάφορες μεθόδους σε διάφορους συνδυασμούς πειραματικών δεδομένων και τοπολογιών δικτύων. Στο Παράρτημα πίσω, παρατίθεται ένας τελικός αλγόριθμος, ο οποίος συνοψίζει την πειραματική μέθοδο που περιγράφηκε παραπάνω στην ενότητα 3.

Ανατρέχοντας ο ενδιαφερόμενος αναγνώστης και διαβάζοντας τον αλγόριθμο **Main Programm**, παρατηρεί πως είναι δομημένος βάσει της μεθοδολογίας που περιγράφηκε και στο κεφάλαιο 3, σε βασικά βήματα.

Επίσης, παρουσιάζεται και η υπορουτίνα **Path\_Construction**, η οποία χρησιμοποιείται ουσιαστικά για τη σταδιακή δόμηση των μονοπατιών που ξεκινούν από κάθε κόμβο και καταλήγουν στους άλλους κόμβους του δικτύου.

---

## 5. Ανακεφαλαίωση - Συμπεράσματα

Στις μέρες μας για την εξέλιξη της έρευνας απαιτείται η συνεργασία επιστημόνων από διαφορετικά πεδία. Χαρακτηριστικό παράδειγμα άρτιας συνεργασίας μεταξύ δύο επιστημών αποτελεί η **Βιοϊατρική Τεχνολογία**, όπου η έρευνα στην ιατρική προωθείται με τη βοήθεια της μηχανολογίας. Η παρούσα επιστήμη διαιρείται σε αρκετούς κλάδους. Ένας από τους κλάδους αυτούς είναι και αυτός της **Συστημικής Βιολογίας**.

Ένας τομέας του συγκεκριμένου επιστημονικού κλάδου, είναι ο τομέας της **Βιοπληροφορικής**. Γενικά, ο συγκεκριμένος ερευνητικός τομέας, τα τελευταία χρόνια, κατέχει ένα ιδιαίτερα σημαντικό και ενδιαφέρον ρόλο στη διεθνή επιστημονική κοινότητα με πολλές προοπτικές εξέλιξης και πολλά πεδία εφαρμογής. Ο συνδυασμός των θεωρητικών γνώσεων της ιατρικής με τις πρακτικές εφαρμογές των γνώσεων της μηχανολογίας τον καθιστούν ιδιαίτερα χρήσιμο για την εξαγωγή κατάλληλων και χρήσιμων αποτελεσμάτων στην ανάπτυξη φαρμάκων και στην καταπολέμηση ασθενειών και παθήσεων.

Ένα τέτοιο πεδίο εφαρμογής είναι και η Μοντελοποίηση της Κυτταρικής Μεταγωγής Σήματος. Ο όρος **κυτταρική μεταγωγή σήματος (Cellular Signal Transduction)** αναφέρεται ουσιαστικά στη διαδικασία **«μετάφρασης»** μίας περιβαλλοντικής διαταραχής σε μία κυτταρική απόκριση. Ο τρόπος με τον οποίο τα κύτταρα αντιλαμβάνονται αυτή τη διαταραχή, οφείλεται σε ειδικευμένα πρωτεϊνικά μόρια, στην επιφάνειά τους, που ονομάζονται **υποδοχείς (receptors)**. Ερεθίσματα από το περιβάλλον του κυττάρου προσκολλώνται στους υποδοχείς του, οι οποίοι κατά συνέπεια ενεργοποιούνται αρχίζοντας, έτσι, τη **διαδικασία σηματοδότησης**. Η μεταγωγή του σήματος είναι, ουσιαστικά, η διαδικασία της διάδοσής του από τη μία πρωτεΐνη στην επόμενη, ξεκινώντας από το επίπεδο των υποδοχέων φτάνοντας μέσα στον πυρήνα. Γενικά, η μελέτη της μεταγωγής σήματος μπορεί να δια φωτίσει τον τρόπο με τον οποίο τα κύτταρα αποκρίνονται στους παράγοντες του βιοχημικού τους μικροπεριβάλλοντος.

Τα γεγονότα της σηματοδότησης των κυττάρων αναπαρίστανται τυπικά σε δίκτυα, τα οποία αποκαλούνται **δίκτυα σηματοδότησης, μονοπάτια σηματοδότησης ή γράφοι αλληλεπίδρασης**. Με τον όρο **μονοπάτια μεταγωγής σημάτων μεταξύ κυττάρων (cell signaling transduction pathways)**, αναφερόμαστε σε μία ομάδα πρωτεϊνών, οι οποίες μεταφέρουν μεταγωγικά σήματα (το οποίο σημαίνει πως μετατρέπουν σήματα από το εξωτερικό περιβάλλον του κυττάρου σε διαφορετικά σήματα στο εσωτερικό του κυττάρου).

Η **μετάδοση σημάτων μεταξύ κυττάρων** είναι μέρος ενός πιο σύνθετου συστήματος επικοινωνίας, το οποίο διέπει βασικές κυτταρικές δραστηριότητες και προσανατολίζει κυτταρικές πράξεις. Η δυνατότητα των κυττάρων να αντιλαμβάνονται και να αντιδρούν ανάλογα στο μικρό – περιβάλλον τους, είναι θεμελιώδης αρχή της ανάπτυξης και ανάπτυξης των ιστών και της ανοσίας και ομοιοσύστασης τους. Η πλήρης και σωστή κατανόηση της μεταγωγής σημάτων μεταξύ κυττάρων, μπορεί να οδηγήσει σε αποτελεσματική

αντιμετώπιση ασθενειών και, θεωρητικά, σε δημιουργία τεχνητών ιστών και οργάνων.

Παραδοσιακά, η επιστήμη της βιολογίας έχει εστιάσει στη μελέτη των επιμέρους τμημάτων των μονοπατιών μεταγωγής σημάτων μεταξύ κυττάρων. Η έρευνα στον τομέα της συστημικής βιολογίας μας βοηθάει να κατανοήσουμε καλύτερα, την υποκείμενη δομή των δικτύων σηματοδότησης κυττάρων και πως οι αλλαγές σε αυτά τα δίκτυα μπορεί να επηρεάσουν τη μετάδοση και τη ροή της πληροφορίας. Τέτοιου τύπου δίκτυα είναι πολύπλοκα συστήματα σε ό,τι αφορά την οργάνωσή τους και μπορούν να επιδείξουν ένα σημαντικό αριθμό αναδυόμενων ιδιοτήτων, όπως η βιοσταθερότητα και η υπερευαίσθητοποίηση. Η ανάλυση αυτών των δικτύων απαιτεί ένα συνδυασμό πειραματικών και θεωρητικών προσεγγίσεων, συμπεριλαμβανομένων της ανάπτυξης και της ανάλυσης της προσομοίωσης και της μοντελοποίησης των αντίστοιχων βιολογικών συστημάτων.

Η μοντελοποίηση των δικτύων σηματοδότησης αναφέρεται στη χρήση του μαθηματικού φορμαλισμού στο να περιγραφεί το πώς το σήμα διαδίδεται από τον ένα κόμβο στον άλλο, καταλήγοντας σε εκτελέσιμα μοντέλα. Γενικά, η μοντελοποίηση αυτού του τύπου των δικτύων πραγματοποιείται με χρήση τριών γενικών μεθόδων: με τη χρήση μοντέλου **Συνήθων Διαφορικών Εξισώσεων (ODEs)**, βάσει της **λογικής κατά Boole (Boolean Model Analysis)** και βάσει της **θεμελιώδους θεωρίας δικτύων (Elementary Graph Algorithms)**.

Στην παρούσα διπλωματική ακολουθήθηκε η μοντελοποίηση βάσει του τρίτου μοντέλου. Η **θεωρία των γράφων** αποτελεί ένα βασικό κλάδο των συνδυαστικών μαθηματικών και έχει μελετηθεί εντατικά για πολύ καιρό τώρα. Πολλές σημαντικές και χρήσιμες ιδιότητες των γράφων έχουν αποδειχτεί, αλλά πολλά δύσκολα προβλήματα μένουν να ακόμα να επιλυθούν. Η συμπεριφορά των πιο πολύπλοκων συστημάτων, από το κύτταρο μέχρι το διαδίκτυο, αναδύεται από την ενορχηστρωμένη ενεργοποίηση πολλών μερών, τα οποία αντιδρούν το ένα με το άλλο μέσω ζευγών αλληλεπίδρασης. Σε ένα εξαιρετικά αφηρημένο επίπεδο, τα μέρη μπορεί να μειωθούν σε μία σειρά κόμβων οι οποίοι συνδέονται ο ένας με τον άλλον με γραμμές, με την κάθε γραμμή να αντιπροσωπεύει τις αντιδράσεις μεταξύ των δύο μερών. Οι κόμβοι αυτοί και οι γραμμές αυτές σχηματίζουν ένα **δίκτυο (network)** ή σε μία πιο επίσημη μαθηματική γλώσσα, ένα **γράφο (graph)**. Η θεωρία των δικτύων στη βιολογία προσφέρει μία ποσοτική περιγραφή των δικτύων, η οποία χαρακτηρίζει ποικίλα βιολογικά συστήματα.

Μία πλούσια πηγή κατανόησης των μηχανισμών αυτών των βιολογικών συστημάτων παρέχουν τα πειραματικά δεδομένα. Τα δεδομένα αυτά προκύπτουν από ποικίλες και διαφορετικές πειραματικές μετρήσεις, οι οποίες αποσκοπούν κυρίως στη μέτρηση πρωτεϊνών. Αυτό συμβαίνει γιατί η μελέτη σε γονιδιακό επίπεδο δεν αρκεί. Η λειτουργία ενός κυττάρου είναι πολύ πολύπλοκη και δεν μπορεί να ερμηνευτεί μόνο μέσω της έκφρασης του γονιδιώματος. Σύμφωνα με το κεντρικό δόγμα της βιολογίας, οι πρωτεΐνες είναι οι τελικοί αποδέκτες της γενετικής πληροφορίας, η οποία ξεκινά από το γονιδίωμα και ως εκ τούτου αποτελούν, από βιολογικής άποψης τουλάχιστον,

τον **φαινότυπο** κάθε συστήματος. Η αρμόδια επιστήμη είναι αυτή της **Πρωτεϊνωματικής (Proteomics)**.

Κάθε τύπος κυττάρων έχει συγκεκριμένους μηχανισμούς μεταγωγής σήματος και διάφορες ασθένειες προκύπτουν από μεταλλάξεις στα μονοπάτια σηματοδότησης. Μικρό-μόρια αναστολείς έχουν αναδειχθεί ως νέες φαρμακευτικές παρεμβάσεις που στοχεύουν να μπλοκάρουν ορισμένα μονοπάτια, σε μία προσπάθεια να αναστραφεί η ανωμαλία του φαινοτύπου των ασθενών κυττάρων. Παρόλο που ενώσεις έχουν σχεδιαστεί για να αναστέλλουν διάφορα μόρια, λίγα είναι γνωστά για τον τρόπο δράσης τους σε ένα «λειτουργικό» δίκτυο σηματοδότησης. Συνδυάζοντας υψηλής απόδοσης μετρήσεις πρωτεϊνών-σημάτων και υπολογιστικές τεχνικές επιτυγχάνεται η αξιολόγηση των επιδράσεων των φαρμάκων στα κύτταρα. Βάσει της απόκρισης σηματοδότησης που προκύπτει από τις μετρήσεις και ενός συνδυασμού πιθανών αντιδράσεων (γενετικά μονοπάτια), δημιουργείται ένα συγκεκριμένο μονοπάτι κυτταρικού τύπου χρησιμοποιώντας μία αποδοτική φόρμουλα βελτιστοποίησης, γνωστή και ως **Ακέραιος Γραμμικός Προγραμματισμός (Integer Linear Programming)**. Αυτή η προσέγγιση δημιουργείται σύμφωνα με τη μοντελοποίηση κατά Boole και λύνεται με τη χρήση τυποποιημένων εμπορικών πακέτων λογισμικού με εγγυημένη αποδοτικότητα. Εν γένει, η μέθοδος αυτή βασίζεται στην πραγματική επίδραση των συμβάντων φωσφορυλίωσης, προσεκτικά εξαπλωμένη στο δίκτυο σηματοδότησης, ενώ από υπολογιστικής άποψης παρουσιάζει μία πιο γρήγορη και πιο αποδοτική βελτιστοποίηση μονοπατιών σε σχέση με τους σύγχρονους αλγορίθμους.

Παρόλα αυτά, στην παρούσα διπλωματική ασχοληθήκαμε με μία διαφορετική προσέγγιση για τη μοντελοποίηση των βιολογικών δικτύων, η οποία ξεφεύγει από τα πλαίσια συμβατικής βελτιστοποίησης, υπό την έννοια της υπέρβασης κάποιας αντικειμενικής συνάρτησης, περιορισμών, μαθηματικών μοντέλων και συναρτήσεων, βρόχων επαναλήψεων ή υπολογιστικού κόστους. Η μοντελοποίηση εστιάζει σε αλγοριθμικές μεθόδους και προγραμματιστικές διαχειρίσεις δεδομένων. Ο αλγόριθμος που αναπτύχθηκε, **εξετάζει στο κάθε μονοπάτι μεταξύ των κόμβων του κάθε ζευγαριού των πειραματικών δεδομένων (dependencies), πόσες συγκρούσεις (conflicts) υπάρχουν, αλλά και πόσοι κόμβοι εξάρτησης (dependencies) εξυπηρετούνται. Αν ο αριθμός των συγκρούσεων (conflicts) είναι μεγαλύτερος, το μονοπάτι, καθώς και το συγκεκριμένο ζευγάρι εξάρτησης (dependency pair), απορρίπτονται από το τελικό δίκτυο. Αν ο αριθμός των συγκρούσεων (conflicts) είναι ίσος ή μικρότερος, τότε θεωρεί πως το συγκεκριμένο μονοπάτι πρέπει να συμπεριληφθεί στο τελικό δίκτυο, έτσι ώστε να υπάρχει και όσο δυνατόν περισσότερη πληροφορία της αρχικής τοπολογίας.**

Χρησιμοποιώντας τη μέθοδο αυτή προέκυψαν δίκτυα πιο συνοπτικά, τα οποία περιλαμβάνουν τα κατά δύναμιν βέλτιστα μονοπάτια, ελαχιστοποιώντας τους πιθανούς κόμβους σύγκρουσης (conflicts). Ταυτόχρονα μεγιστοποιήθηκε ο αριθμός των ικανοποιούμενων σχέσεων εξάρτησης (dependencies), ώστε να ανταποκρίνεται το τελικό δίκτυο, όσο το δυνατόν περισσότερο, στην αρχική

τοπολογία. Τέλος, περιλαμβάνονται εναλλακτικές διαδρομές, ώστε να υπάρχει η μέγιστη εν δυνάμει πληροφορία στο δίκτυο.

Τα αποτελέσματα, όπως παρουσιάζονται, φαίνονται αρκετά λογικά και ανταποκρίνονται στους περιορισμούς και τις προδιαγραφές του προβλήματος. Κατ' επέκταση, η μέθοδος είναι αρκετά λειτουργική και μπορεί να ανταπεξέλθει στις απαιτήσεις της συγκεκριμένης μελέτης, κυρίως όμως στην περίπτωση δικτύων μικρής κλίμακας. Θα πρέπει να συνεχιστεί η αλγοριθμική της ανάπτυξη, ώστε να μπορέσει να τελειοποιηθεί και να φτάσει το όριο των δυνατοτήτων της, όπως καθώς και να δοκιμασθεί σε περισσότερα δεδομένα, κυρίως δίκτυα μεσαίας και μεγάλης κλίμακας. Συνεπώς, σταδιακά θα υπάρχει μία γενίκευση της ισχύος της και μία δομημένη σύνταξη της, η οποία θα μπορέσει να την καθιερώσει σαν μία εναλλακτική μεν, ευρέως χρησιμοποιούμενη δε, μέθοδο.

Τέλος, στόχος υπήρξε εξ αρχής η ανάπτυξη και η προαγωγή της έρευνας στο συγκεκριμένο επιστημονικό πεδίο. Για το λόγο αυτό θα ήταν ιδιαίτερα ενδιαφέρουσα, σε επίπεδο ενδεχομένως κάποιας άλλης διπλωματικής εργασίας ή κάποιας διδακτορικής διατριβής, η περαιτέρω έρευνα πάνω στη χρήση ή και ανάπτυξη μεθόδων, οι οποίες θα μπορούσαν να συνδυάζουν τη δομημένη σκέψη και το μαθηματικό υπόβαθρο των κλασικών μεθόδων βελτιστοποίησης, με την ελευθερία σκέψης και κινήσεων και την ταχύτητα, που παρέχουν εναλλακτικές αλγοριθμικές μέθοδοι, οι οποίες βασίζονται κυρίως στο προσωπικό χαρακτήρα του εκάστοτε προγραμματιστή και στη διαφορετική αντίληψη με την οποία αντιλαμβάνεται αυτός το εξεταζόμενο πρόβλημα.

---

**Main Programm**

```
%%%%%%%%%%  
%  
%           Code4           %  
%           %  
%%%%%%%%%%  
  
clc;  
clear all;  
  
%%%%%%%%%%  
% Prologue %  
%%%%%%%%%%  
%  
% In this code we do some things differently than before like: we have a  
% different "dependencies" type file and we attempt to make a "conflict  
% minimize" algorithm.  
%  
%%%%%%%%%%  
  
           %%%%%%%%%%%  
           % Data reading %  
%%%%%%%%%%  
%  
% Dependencies  
%%%%%%%%%%  
  
[num,textdata]=xlsread('hepg2_phosphos.xls');  
  
dependmtr=num;  
dependencies={};  
k=1;  
  
for i=1:size(num,1)  
for j=1:size(num,2)  
  
if num(i,j)~=0  
  
dependencies{k,1}=textdata{i+1,1};  
dependencies{k,2}=textdata{1,j+1};  
k=k+1;  
  
end  
  
end  
end  
  
uniquesource=unique(dependencies(:,1));  
uniquetarget=unique(dependencies(:,2));  
  
%  
% Network  
%%%%%%%%%%  
  
[num,textdata2]=xlsread('Network_generic.xls');  
clear num;  
network=textdata2;  
clear textdata2;
```



```

%
% Signals
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

k=1;
signals={};

for i=1:size(textdata,2)

if isempty(textdata{1,i})==0
signals{k}=textdata{1,i};
k=k+1;
end

end

signals=signals';

clear('i','j','k','textdata','num');

%
% Stimuli
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

[num,textdata4]=xlsread('stimuli.xls');
clear num;
stimuli=textdata4;
clear textdata4;

%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Counting the reactions - making the array %
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

% Sorting 1 %
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

reactors=network(:,1);
reactors=sort(reactors);

match1=reactors{1};
react={};
count1=1;
k=1;

for i=2:length(reactors)
B=reactors{i};
matching = strcmp(match1, B);

if matching==1
count1=count1+1;
else
react{k,1}=match1;
react{k,2}=count1;
match1=B;
k=k+1;
count1=1;
end

end

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Last Array Element %
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

react{k,1}=match1;
react{k,2}=count1;

```

```

%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
produced=network(:,3);
produced=sort(produced);

match2=produced{1};
prod={ };
count2=1;
k=1;

for i=2:length(produced)
    B=produced{i};
    matching = strcmp(match2, B);

    if matching==1
        count2=count2+1;
    else
        prod{k,1}=match2;
        prod{k,2}=count2;
        match2=B;
        k=k+1;
        count2=1;
    end
end

%
% Last Array Element %
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

prod{k,1}=match2;
prod{k,2}=count2;

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Finding Edge - Vertex number %
%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%
%
%
% Total Number of Nodes %
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

nodes=union(react(:,1),prod(:,1));

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Graph Statistical Analysis %
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

clc;

fprintf('-----\n');
fprintf('\n');
fprintf(' Graph Statistical Analysis \n');
fprintf('\n');

```

```

fprintf('-----\n');

fprintf('-----\n');

fprintf('\n');
disp(['the total number of reactions is:' num2str(length(reactors))]);
fprintf('\n');
disp(['the total number of reactors is:' num2str(length(react))]);
fprintf('\n');
disp(['the total number of produced is:' num2str(length(prod))]);
fprintf('\n');
disp(['the total number of nodes in graph is:' num2str(length(nodes))]);
fprintf('\n');
disp(['the total number of signals in graph is:' num2str(length(signals))]);
fprintf('\n');
disp(['the total number of stimuli in graph is:' num2str(length(stimuli))]);
fprintf('\n');
disp(['the total number of sources node in graph is:' num2str(length(uniqesource))]);
fprintf('\n');
disp(['the total number of targets in graph is:' num2str(length(uniquetarget))]);
fprintf('\n');

fprintf('-----\n');

pause;
clc;

%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%                               %
%               Adjacency Matrix %
%               Representation    %
%                               %
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%
% Making the Adjacency Matrix
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

adjmtrx=zeros(length(nodes));

for i=1:length(network)

    r=network(i,1);
    p=network(i,3);

    for j=1:length(nodes)

        matching1=strcmp(nodes,r);
        matching2=strcmp(nodes,p);

    end

    pos1=find(matching1==1);
    pos2=find(matching2==1);

    adjmtrx(pos1,pos2)=1;

end

% Adjmxtr verification
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

dokim=strcmp(nodes,'a20');
ps1=find(dokim==1);
dokim=strcmp(nodes,'traf6');

```

```

ps2=find(dokim==1);
adjmtrx(ps1,ps2)

%Delete the unnecessary variables
clear('ans','ps1','ps2','dokim','matching1','matching2','pos1','pos2','r');
clear('k','match1','match2','count1','count2','B','p','i','j','ps','matching');
clear('textdata3');

%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%
%           Elementary Graph Algorithms           %
%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%
%
% Strongly Connected Components :
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%
%
%finds the strongly connected components of the graph represented
%by matrix G using Tarjan's algorithm.
%A strongly connected component is a maximal group of nodes that are
%mutually reachable without violating the edge directions.
%Input G is an N-by-N sparse matrix that represents a graph.
%Nonzero entries in matrix G indicate the presence of an edge
%

G=sparse(adjmtrx);

disp('the number of components found in the graph is:')
[S, C] = graphconncomp(G);
disp(S)
disp('and to which component each node belongs')
disp(C);

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Spanning tree :
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%
%
%A spanning tree must touch all the nodes and must be acyclic.
%G is an N-by-N sparse matrix whose lower triangle represents an undirected
%graph. Nonzero entries in matrix G indicate the presence of an edge
%
%

TF = graphisspantree(G); %returns logical 1 (true) if G is a spanning tree,
                        %and logical 0 (false)

if TF==1
    disp('the graph is spanning tree')
else
    disp('the graph is not a spanning tree')
end

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Minimum Spanning Tree :
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%
%
%finds an acyclic subset of edges that connects all the nodes in the
%undirected graph G and for which the total weight is minimized.

```

```

%Weights of the edges are all n%the N-by-N sparse matrix G.
%Output Tree is a spanning tree represented by a sparse matrix
%
%

[Tree, pred] = graphminspantree(G);

clc;

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Transitive closure NOT Warshall's algorithm :
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%
%

path = adjmtrx;
ind = length(adjmtrx);
counter=0;

for i=1:ind
for j=1:ind
if path(i,j)==1
for k=1:ind
if path(j,k)==1
path(i,k) = 1;
end
end
end
end

end

% Nodes communication

tran_clos={};
k=1;
for i=1:length(adjmtrx)
tran_clos{k,1}=nodes{i};
col=2;
for j=1:length(adjmtrx)
if (i~=j) && (path(i,j)==1)
tran_clos{k,col}=nodes{j};
col=col+1;
end
end
k=k+1;
end

end

%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Floyd - Warshall Algorithm with Path Reconstruction %
%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

react_new={};
prod_new={};
db_new=[];

for i=1:length(network)

react_new{i}=network{i,1};
prod_new{i}=network{i,3};
db_new(i)= 1;

```

```

end

react_new=react_new';
prod_new=prod_new';
db_new=db_new';

nodes_new=union(react_new,prod_new);

dist =inf(length(nodes_new));
next=zeros(length(nodes_new));
B_new=nodes_new';

for i=1:length(nodes_new)
    dist(i,i)=0;
end

for j=1:length(react_new)

    matching1=strcmp(B_new,react_new{j});
    row_w=find(matching1==1);
    matching2=strcmp(B_new,prod_new{j});
    column_w=find(matching2==1);
    dist(row_w,column_w)=db_new(j);

end

for k=1:length(nodes_new)
    for i=1:length(nodes_new)
        for j=1:length(nodes_new)
            if dist(i,k)+dist(k,j)<dist(i,j)
                dist(i,j)=dist(i,k)+dist(k,j);
                next(i,j)=k;
            end
        end
    end
end

end

%%%%%%%%%%%%%%
% Path Reconstruction %
%%%%%%%%%%%%%%

path_w=cell(length(nodes_new));

for i=1:length(nodes_new)

    for j=1:length(nodes_new)
        path_w{i,j}=Path4(i,j,next);
    end
end

%Delete the unnecessary variables
clear ('col','counter','i','j','k','C','G','Tree','TF','S','ans','pred');
clear('produced','reactors','source','target');

% Warshall's verification
%%%%%%%%%%%%%%

source=dependencies(:,1);
target=dependencies(:,2);

for i=1:length(dependencies)

```

```

A=source(i);
B=target(i);
dokim=strcmp(nodes_new,A);
ps1=find(dokim==1);
dokim=strcmp(nodes_new,B);
ps2=find(dokim==1);
d=dist(ps1,ps2);
disp(d); % If dist between two nodes=1--> there is no in-between path.
        %There is only the reaction between them!!!
end

%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%          Checking the dependencies file          %
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%
% Here we check which connections nodes (dependencies file) exist in the
% network data.
%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

fprintf('-----\n');
fprintf('\n');
fprintf('          Checking the dependencies file          \n');
fprintf('\n');
fprintf('-----\n');

objective={};
ct=0;
for i=1:length(dependencies)

    A=source(i);
    B=target(i);

    matching=strcmp(A,nodes_new);
    pos1=find(matching==1);

    matching=strcmp(B,nodes_new);
    pos2=find(matching==1);

    if isempty(pos1)==1
        fprintf('-----\n');
        fprintf('\n');
        disp(['The source node' cellstr(A)]);
        disp('does NOT exist in the network data');
        fprintf('\n');
        fprintf('-----\n');
        ct=ct+1;
        objective{i}={};
    elseif isempty(pos2)==1
        fprintf('-----\n');
        fprintf('\n');
        disp(['The target node' cellstr(B)]);
        disp('does NOT exist in the network data');
        fprintf('\n');
        fprintf('-----\n');
        ct=ct+1;
        objective{i}={};
    elseif isempty(pos1)==1 && isempty(pos2)==1
        fprintf('-----\n');
        fprintf('\n');
        disp([' The source node ' cellstr(A)]);
        disp([' and the target node ' cellstr(B)]);

```

```

disp(' do NOT exist in the network data');
fprintf('\n');
fprintf('-----\n');
ct=ct+1;
objective{i}={};
else
objective{i}=path_w{pos1,pos2};
end

end

objective=objective';

if ct~=0
disp([num2str(ct) ' dependencies do NOT exist in the network data ']);
elseif ct==length(dependencies)
disp( ' None dependency relation exists in the network data ');
else
disp( ' All the dependencies exist in the network data ');
end

%%
%%%%%%%%%%
%                               %
%               Objective Pathways               %
%                               %
%%%%%%%%%%
%
% Above we have already found the Objective Pathways. As Objective Path
% we call the path that results from Warshall's Algorithm. After this, we
% check if this path is acceptable by seeing if each path-node is a signal
% node. We find the non-empty paths positions in nepp matrix and we check
% if it is acceptable path.
%
%%%%%%%%%%

fprintf('-----\n');
fprintf('|                               |\n');
fprintf('|               Objective Pathways               |\n');
fprintf('|                               |\n');
fprintf('-----\n');

nepp=[]; % Array with non-empty-path-positions (direct path positions)
rpp=[]; % Array with rest-path-positions
k=1;
m=1;

for i=1:length(objective)

if isempty(objective{i})==0

nepp(k)=i;
k=k+1;

else

rpp(m)=i;
m=m+1;

end

end

nepp=nepp';
rpp=rpp';

```



```

                %%%%%%%%%%
                % Signal Nodes %
%%%%%%%%%%%%%
%
% Here we find the number of signal nodes
%

signod=zeros(length(signals),1);
for i=1:length(signals)

    A=signals(i);
    matching=strcmp(nodes_new,A);
    pos=find(matching==1);

    if isempty(pos)==1
        signod(i)=0;
    else
        signod(i)=pos;
    end

end

%%%%%%%%%%%%%

                %%%%%%%%%%
                % Stimuli Nodes %
                %%%%%%%%%%
%%%%%%%%%%%%%
%
% Here we find the number of stimuli nodes
%

stimnod=zeros(length(stimuli),1);
for i=1:length(stimuli)

    A=stimuli(i);
    matching=strcmp(nodes_new,A);
    pos=find(matching==1);

    if isempty(pos)==1
        stimnod(i)=0;
    else
        stimnod(i)=pos;
    end

end

%%%%%%%%%%%%%

%%%%%%%%%%%%%
% Presenting the direct paths %
%%%%%%%%%%%%%
%
% Here we present the direct Warshall's path
%

finalgr={};
fgct=1;
ctr=0;

for i=1:length(nepp)

    flag=0;

```

```

A=source(nepp(i));
B=target(nepp(i));

helpy=nodes_new(objective{nepp(i)});

for j=1:length(helpy)

    compr2=strcmp(helpy{j,1},signals);
    pst2=find(compr2);
    compr1=strcmp(A,stimuli);
    pst1=find(compr1);

    if sum(compr2)==1 && dependmtr(pst1,pst2)==0
        flag=1;
    end

end

if flag==0

    fprintf('-----\n');
    disp([' From source ' cellstr(A)]);
    disp([' we re going to target ' cellstr(B) ]);
    disp('through the path: ');
    fprintf('\n');
    disp (nodes_new(objective{nepp(i)}));
    fprintf('-----\n');
    fprintf('\n');

    helpa={};
    bl=1;
    helpa(bl)=A;

    for k=1:length(helpy)
        bl=bl+1;
        helpa{bl}=helpy{k};
    end

    helpa(bl+1)=B;
    helpa=helpa';

    for k=1:length(helpa)-1

        finalgr{fgct,1}=helpa{k};
        finalgr{fgct,2}=helpa{k+1};
        fgct=fgct+1;

    end

else

    ctr=ctr+1;
    rpp(length(rpp)+1)=nepp(i);

end

end

end

%Delete the unnecessary variables
clear('A','B','matching','pos1','pos2','i','ps1','ps2','dokim','d','match');
clear('data','db_new','nodes','path','prod','react','textdata','tran_clos');
clear('ans','k','j');

pause;

%%

```

```

%%%%%%%%%%%%%%
%
%           No possible Pathways           %
%
%%%%%%%%%%%%%%
%
% Here we find the NO possible pathways for the dependencies data we have.
% As no possible pathway we reference to the path through which we will
% never reach the target node because it aint exist. And if there is
% NO Warshall's path, we seek other possible paths through the source
% nodes. If there is NO other possible path, then there is NO path at all.
%
%%%%%%%%%%%%%%

```

```

fprintf('-----\n');
fprintf('|                               |\n');
fprintf('|           No possible Pathways |\n');
fprintf('|                               |\n');
fprintf('-----\n');

```

```

rlbmtr=zeros(length(rpp),1); % Reliability Matrix shows if there is a possible path --> 1
% or NOT --> 0 !!!

```

```

for i=1:length(rpp);

s=source(rpp(i));
t=target(rpp(i));

mtcs=strcmp(s,nodes_new);
nds=find(mtcs==1);

mtct=strcmp(t,nodes_new);
ndt=find(mtct==1);

d=dist(nds,:);

if length(unique(d))>2

    rlbmtr(i,1)=1;

end

end

```

```

%%%%%%%%%%%%%%
% Presenting the NO possible paths %
%%%%%%%%%%%%%%
%
% Here we present the NO paths because there is no Warshall's path and not
% any other possible path through other nodes
%

```

```

for i=1:length(rlbmtr)

if rlbmtr(i)==0

fprintf('-----\n');
disp([' From source ' cellstr(source(rpp(i)))]);
disp('there is NO possible path: ');
fprintf('\n');
disp(['to target ' cellstr(target(rpp(i))) ]);
disp(' in this network !!! ');
fprintf('\n');
fprintf('-----\n');


```

```

fprintf('\n');

end

end

pause;

%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%                               %
%               Alternative Paths               %
%                               %
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%
% So far, we have already found the direct Warshall paths NOT passing
% through a signal node, and we have also found the NO possible paths from
% one source node to its target node (based on the dependencies data).
% So, here we examine the rest dependencies:
%
% a) if there is a direct path (NOT the Warshall path)through some other
% nodes then we have a desirable path. If this path goes through a
% signal node then is not permittable and we face what we call
% "CONFLICT", because we can ONLY go to the target node through
% another signal node.
% b) if there is no other direct pathm then we face the similar with above
% case of NO possible pathways.
%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

```

```

fprintf('-----\n');
fprintf(|                               |\n);
fprintf(|               Alternative Paths               |\n);
fprintf(|                               |\n);
fprintf('-----\n');

```

```

% Finding the array with the possible alternative pathways positions for
% each resting source node
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

```

```

m=1;
pos=[];

for i=1:length(rpp);

if rlbmtr(i)==1

s=source(rpp(i));
t=target(rpp(i));

mtcs=strcmp(s,nodes_new);
nds=find(mtcs==1);

mtct=strcmp(t,nodes_new);
ndt=find(mtct==1);

d=dist(nds,:);
uni=unique(d);

l=1;
for j=1:length(uni)

if uni(j)~=0 && uni(j)~=inf

```

```

for k=1:length(d)
    if dist(nds,k)==uni(j)
        pos(m,l)=k; % Rows are the rest nodes - columns are the alternative dist positions
        l=l+1; % Next column
    end
end
end
end
end

m=m+1; % Next row
end
end

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Finding the dependencies nodes %
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

sourcenod=zeros(length(source),1);
targetnod=zeros(length(target),1);

for i=1:length(dependencies)

    A=source(i);
    match1=strcmp(A,nodes_new);
    sourcenod(i)=find(match1==1);

    B=target(i);
    match2=strcmp(B,nodes_new);
    targetnod(i)=find(match2==1);

end
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

% Finding the rest array with the source nodes with existing alter path
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

rest=[];
j=1;
for i=1:length(rlbmtr)

    if rlbmtr(i)==1

        rest(j,1)=rpp(i);
        j=j+1;

    end
end

% Presenting the alternative paths for the rest source nodes
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

rest1=rest;
for i=1:size(pos,1)

    for j=1:nnz(pos(i,:))

```

```

pth=path_w{pos(i,j),targetnod(rest(i))};

ct=0;
if isempty(pth)==0

for k=1:size(pth,2)

    if sum(pth(1,k)==signod)==0

        ct=ct+1;

    else

        posit2=find(pth(1,k)==signod);
        a=(sourcenod(rest(i))==stimnod);
        posit1=find(a);

        if dependmtr(posit1,posit2)==1 || dependmtr(posit1,posit2)==-1

            ct=ct+1;

        end

    end

end

end
end

pth1=path_w{sourcenod(rest(i)),pos(i,j)};

ct1=0;
if isempty(pth1)==0

for k=1:size(pth1,2)

    if sum(pth1(1,k)==signod)==0

        ct1=ct1+1;

    else

        posit2=find(pth1(1,k)==signod);
        a=(sourcenod(rest(i))==stimnod);
        posit1=find(a);

        if dependmtr(posit1,posit2)==1

            ct1=ct1+1;

        end

    end

end

end

end

if isempty(pth1)==0 && isempty(pth)==0
    rest1(i)=inf;
if ct==size(pth,2) && ct1==size(pth1,2)

    fprintf('-----\n');
    disp([' From source: ' cellstr(source(rest(i)))]);
    disp([' we go to target: ' cellstr(target(rest(i)))]);
    disp(' through the path: ');
    disp(nodes_new(pth1));
    disp(nodes_new(pos(i,j)));

```

```

disp(nodes_new(pth));
fprintf('-----\n');
rest1(i)=0;

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

helpa={};
bl=1;
helpa(bl)=cellstr(source(rest(i)));

for k=1:length(nodes_new(pth1))
    bl=bl+1;
    helpa(bl)=nodes_new(pth1(k));
end

bl=bl+1;
helpa(bl)=nodes_new(pos(i,j));

for k=1:length(nodes_new(pth))
    bl=bl+1;
    helpa(bl)=nodes_new(pth(k));
end

helpa(bl+1)=cellstr(target(rest(i)));
helpa=helpa';

for k=1:length(helpa)-1

    finalgr{fgct,1}=helpa{k};
    finalgr{fgct,2}=helpa{k+1};
    fgct=fgct+1;

end

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

end

end

end

end

pause;

%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%                               %
%           Finding Conflicts           %
%                               %
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%
% Here we find the conflicts ! If there is a path with a signal node
% through it, then we have a conflict. If there is NO possible path neither
% an alternative path, then we present the NO existing paths. We check
% if the Path_w of rest1 is empty. If it is true, then there is NO possible
% pathway.
%
% Also, we permit to the pathways with only one conflict into the final network
%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

fprintf('-----\n');
fprintf('|                               |\n');
fprintf('|           Finding Conflicts           |\n');

```

```

fprintf(' \n');
fprintf('-----\n');

rest2=rest1;
for i=1:length(rest1)

    ctcf=0;
    ctcf1=0;
    ctcf2=0;

    if rest1(i)~=0

        helpyp=path_w{sourcenod(rest(i)),targetnod(rest(i))};
        helpy=nodes_new(helpyp);

        if isempty(helpyp)==1

            fprintf('-----\n');
            disp([' From source ' cellstr(source(rpp(i)))]);
            disp('there is NO possible path: ');
            fprintf('\n');
            disp(['to target ' cellstr(target(rpp(i))) ]);
            disp(' in this network !!! ');
            fprintf('\n');
            fprintf('-----\n');
            fprintf('\n');
            rest2(i)=0;

        else

            for k=1:size(helpyp,2)

                posit2=find(helpyp(k)==signod);
                a=(sourcenod(rest(i))==stimnod);
                posit1=find(a);

                if sum(helpyp(k)==signod)==1 && dependmtr(posit1,posit2)==0

                    fprintf('-----\n');
                    disp([' From source: ' cellstr(source(rest(i)))]);
                    disp([' we go to target: ' cellstr(target(rest(i)))]);
                    disp(' through the path: ');
                    disp(nodes_new(helpyp));
                    fprintf('\n');
                    disp('but the signal conflict is ');
                    disp(nodes_new(helpyp(k)))
                    fprintf('-----\n');
                    rest2(i)=1;
                    ctcf=ctcf+1;

                end

            end

        end

    elseif rest1(i)==inf

        pth1=path_w{sourcenod(rest(i)),pos(i,j)};

        posit2=find(pth1(k)==signod);
        a=(sourcenod(rest(i))==stimnod);

```



```

posit1=find(a);

for k=1:size(pth1,2)

    if sum(pth1(k)==signod)==1 && dependmtr(posit1,posit2)==0

        fprintf('-----\n');
        disp([' From source: ' cellstr(source(rest(i)))]);
        disp([' we go to intermediate node: ' cellstr(pos(i,j))]);
        disp(' through the path: ');
        disp(nodes_new(pth1));
        disp(nodes_new(pth));
        fprintf('\n');
        disp('but the signal conflict is ');
        disp(nodes_new(pth1(k)))
        fprintf('-----\n');
        rest2(i)=1;
        ctcf1=ctcf1+1;

    end

end

pth=path_w{pos(i,j),targetnod(rest(i))};

for k=1:size(pth,2)

    posit2=find(pth(k)==signod);
    a=(sourcenod(rest(i))==stimmnod);
    posit1=find(a);

    if sum(pth(k)==signod)==1 && dependmtr(posit1,posit2)==0

        fprintf('-----\n');
        disp([' From intermediate node: ' cellstr(pos(i,j))]);
        disp([' we go to target: ' cellstr(target(rest(i)))]);
        disp(' through the path: ');
        disp(nodes_new(pth1));
        disp(nodes_new(pth));
        fprintf('\n');
        disp('but the signal conflict is ');
        disp(nodes_new(pth(k)));
        fprintf('-----\n');
        rest2(i)=1;
        ctcf2=ctcf2+1;

    end

end

end

if ctcf==1

    helpa={};
    bl=1;
    helpa(bl)=source(rest(i));

    for k=1:length(helpyp)
        bl=bl+1;
        helpa{bl}=helpy{k};
    end

    helpa(bl+1)=target(rest(i));

```

```

helpa=helpa';

for k=1:length(helpa)-1

    finalgr{fgct,1}=helpa{k};
    finalgr{fgct,2}=helpa{k+1};
    fgct=fgct+1;

end

end

if ctcf1+ctcf2==1

    helpa={};
    bl=1;
    helpa(bl)=cellstr(source(rest(i)));

    for k=1:length(nodes_new(pth1))
        bl=bl+1;
        helpa(bl)=nodes_new(pth1(k));
    end

    for k=1:length(nodes_new(pth))
        bl=bl+1;
        helpa(bl)=nodes_new(pth(k));
    end

    helpa(bl+1)=cellstr(target(rest(i)));
    helpa=helpa';

    for k=1:length(helpa)-1

        finalgr{fgct,1}=helpa{k};
        finalgr{fgct,2}=helpa{k+1};
        fgct=fgct+1;

    end

end

end

%Delete the unnecessary variables
clear('A','B','B_new','ans','column_w','ct','ct1','ctr','d','flag','helpy');
clear('helpyp','i','ind','j','k','l','m','match1','match2','row_w','rest');
clear('s','t','uni','pth','pth1','pth2','matching1','matching2','mtcs');
clear('mtct','nds','ndt')

pause;

%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%                               %
%           Presenting the Conflicts           %
%                               %
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%
% Here, we present finally the conflict nodes of dependencies.
%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

ct=0;

```

```

for i=1:length(rest2)

    if rest2(i)==1

        ct=ct+1;

    end

end

fprintf('\n');
disp([' ' We have ' num2str(ct) ' Conflicts in the network ' ]);
fprintf('\n');

%Delete the unnecessary variables
clear('ct','i');

pause;

%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%                               %
%           Exporting the Results           %
%                               %
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

for i=1:length(finalgr)

    if isempty(finalgr{i,1})==0 && isempty(finalgr{i,2})==0

        A=finalgr{i,1};
        B=finalgr{i,2};

        for j=1:length(finalgr)
            if j~=i && isempty(finalgr{j,1})==0 && isempty(finalgr{j,2})==0

                if (strcmp(A,finalgr{j,1})==1 && strcmp(B,finalgr{j,2})==1)

                    finalgr{j,1}={};
                    finalgr{j,2}={};

                end

            end

        end

        end

    end

    p=fopen('f.dot','w');

    if p== 0
        fprintf('Error in opening a file\n');
    else
        fprintf('Successfully opened the file\n');
    end

    fprintf(p,'digraph G{');
    fprintf(p,'\n');

    for i=1:length(finalgr)

        if isempty(finalgr{i,1})==0
            fprintf(p,'%s -> %s; \n',finalgr{i,1},finalgr{i,2});
        end

    end

```

```

end

for i=1:length(dependencies)
for j=1:length(finalgr)

A=dependencies(i,1);

if isempty(finalgr{j,1})==0

if (strcmp(A,finalgr{j,1})==1 || strcmp(A,finalgr{j,2})==1)
fprintf(p,"%s" [shape=rect, style="filled,rounded", fillcolor=cornflowerblue] \n', dependencies{i,1});
end

end

end

end

for i=1:length(dependencies)
for j=1:length(finalgr)

B=dependencies(i,2);

if isempty(finalgr{j,2})==0

if (strcmp(B,finalgr{j,1})==1 || strcmp(B,finalgr{j,2})==1)
fprintf(p,"%s" [shape=ellipse, style="filled,rounded", fillcolor=firebrick1] \n', dependencies{i,2});
end

end

end

end

fprintf(p,');
fclose all;

%%%
%%%%%%%%%%
%
%           Final Network Reactions Counter           %
%
%%%%%%%%%%

ct=0;

for i=1:length(finalgr)

if isempty(finalgr{i,1})==0

ct=ct+1;

end

end

fprintf('\n');
fprintf('\n');
fprintf('-----\n');
fprintf('\n');
fprintf('\n');
disp([' Reactions in the final network : ', num2str(ct) ]);
fprintf('\n');
fprintf('\n');
fprintf('-----\n');

```

```

fprintf('\n');
fprintf('\n');

%%
%%%%%%%%%%
%
%           Graphs Comparison           %
%
%
%%%%%%%%%%

% Merging the finalgr array
%%%%%%%%%%

fnet={};

k=1;

for i=1:length(finalgr)

    if isempty(finalgr{i,1})==0 && isempty(finalgr{i,2})==0

        fnet{k,1}=finalgr{i,1};
        fnet{k,2}=finalgr{i,2};
        k=k+1;

    end

end

% Reading Teo's network
%%%%%%%%%%

[num,txtdata2]=xlsread('teonet.xls');
clear num;
teonet=txtdata2;
clear txtdata2;

% First Estimation
%%%%%%%%%%
%
% As a first comparison - estimation method of both graphs we can use
% reactions number's division in each final graph
%

if length(fnet) > length(teonet)

    estim = (length(teonet)/length(fnet))*100 ;

else

    estim = (length(fnet)/length(teonet))*100 ;

end

fprintf('\n');
fprintf('\n');

disp([' A first similarity estimation between the two graphs is: ' num2str(estim) '%']);

fprintf('\n');
fprintf('\n');

% Jaccard Index
%%%%%%%%%%

```

```

%
% The Jaccard index, also known as the Jaccard similarity coefficient is a
% statistic used for comparing the similarity and diversity of sample sets.
% The Jaccard coefficient measures similarity between sample sets, and is
% defined as the size of the intersection divided by the size of the union
% of the sample sets.
%

```

```

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%                               %
%   Bill Solutions Matrix       %
%                               %
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

```

```

bill=[ ] ;
k=1;

```

```

for i=1:length(fnet)

```

```

    A=fnet{i,1};
    B=fnet{i,2};

```

```

    for j=1:length(network)

```

```

        if strcmp(A,network{j,1})==1 && strcmp(B,network{j,3})==1

```

```

            bill(k)=j ;
            k=k+1 ;

```

```

        end

```

```

    end

```

```

end

```

```

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%                               %
%   Teo Solutions Matrix       %
%                               %
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

```

```

teodor=[ ] ;
k=1;

```

```

for i=1:length(teonet)

```

```

    A=teonet{i,1};
    B=teonet{i,2};

```

```

    for j=1:length(network)

```

```

        if strcmp(A,network{j,1})==1 && strcmp(B,network{j,3})==1

```

```

            teodor(k)=j ;
            k=k+1 ;

```

```

        end

```

```

    end

```

```

end

```

```

Jaccard = [length(intersect(bill,teodor)) / length(union(bill,teodor))]*100;

```

```

fprintf('-----\n');
fprintf(|                               |\n);
fprintf(|           Jaccard Index         |\n);
fprintf(|                               |\n);
fprintf('-----\n');

fprintf('\n');
fprintf('\n');
disp([' The Jaccard similarity index between the two graphs is: ' num2str(Jaccard) '%']);
fprintf('\n');
fprintf('\n');

```

```

%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

```

```

fprintf('-----\n');
fprintf(|                               |\n);
fprintf(|           THE END                 |\n);
fprintf(|                               |\n);
fprintf('-----\n');

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%                               %
%           END                 %
%                               %
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

```

### **Subroutine Path\_Construction**

```

function x=Path4(i,j,next)

    if dist(i,j)==inf
        x=NaN;

    end

    intermediate = next(i,j);
    if intermediate ==0
        x=[]; % the direct edge from i to j gives the shortest path
    else
        x=[Path4(i,intermediate,next),intermediate,Path4(intermediate,j,next)];
    end

```

## Βιβλιογραφία

### **Βιβλία – Παραδόσεις**

- [ 1 ] *Kitano H, “Foundations of Systems Biology”* (MIT - 2001).
- [ 2 ] *Robert Sedgewick, “Algorithms in C”*, Princeton University, December 1990.
- [ 3 ] *Desmond S. T. Nicholl, “An Introduction to Genetic Engineering”*, University of the West of Scotland, Paisley, UK, Third Edition, Cambridge 2008.
- [ 4 ] *Myer Kutz, “Biomedical Engineering and Design Handbook”*, Second Edition, McGraw-Hill Companies, 2009.
- [ 5 ] *Κ.Χ. Γιαννάκογλου, “Μέθοδοι Βελτιστοποίησης στην Αεροδυναμική”*, Ε.Μ.Π., Αθήνα 2006, 4η Έκδοση.
- [ 6 ] *Κωνσταντίνα Σπ. Νικήτα, “Προσομοίωση Φυσιολογικών Συστημάτων”*, Εκδόσεις Τζιόλα, Αθήνα 2011.
- [ 7 ] *Λ. Αλεξόπουλος, “Σημειώσεις μαθήματος «Εμβιομηχανική και Βιοϊατρική Τεχνολογία», Κεφάλαιο 1: Επιστήμη της Βιοϊατρικής τεχνολογίας σήμερα”*, Ε.Μ.Π., Αθήνα 2011.
- [ 8 ] *Λ. Αλεξόπουλος, “Σημειώσεις μαθήματος «Εμβιομηχανική και Βιοϊατρική Τεχνολογία», Κεφάλαιο 2: Εισαγωγή στη Βιολογία”*, Ε.Μ.Π., Αθήνα 2011.
- [ 9 ] *Λ. Αλεξόπουλος, “Σημειώσεις μαθήματος «Εμβιομηχανική και Βιοϊατρική Τεχνολογία», Κεφάλαιο 5: Βιολογικές Μετρήσεις (Bioinstrumentation)”*, Ε.Μ.Π., Αθήνα 2011.
- [ 10 ] *Λ. Αλεξόπουλος, “Σημειώσεις μαθήματος «Εμβιομηχανική και Βιοϊατρική Τεχνολογία», Κεφάλαιο 6: Συστημική Βιολογία (Systems Biology)”*, Ε.Μ.Π., Αθήνα 2011.
- [ 11 ] *Jeremy M Berg, John L Tymoczko, and Lubert Stryer, “Biochemistry”*, 5th edition, New York 2002.



## Δημοσιεύσεις – Επιστημονικές Διατριβές

- [ 12 ] Michael J. Berridge, **“Cell Signalling Pathways”**, Cell Signalling Biology, Module 2, 2012
- [ 13 ] Shao-shan Carol Huang and Ernest Fraenkel, **“Integrating Proteomic, Transcriptional, and Interactome Data Reveals Hidden Components of Signaling and Regulatory Networks”**, Science Signaling 2 (81), ra40. [DOI: 10.1126/scisignal.2000350] , 28 July 2009.
- [ 14 ] Paul J. Kersey, Jorge Duarte, Allyson Williams, Youla Karavidopoulou, Ewan Birney and Rolf Apweiler, **“The International Protein Index: An integrated database for proteomics experiments”**, 2004 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim, Proteomics 2004, 4, 1985–1988, DOI 10.1002/pmic.200300721.
- [ 15 ] Vassily Hatzimanikatis, Leila H. Choe and Kelvin H. Lee, **“Proteomics: Theoretical and Experimental Considerations”**, Biotechnol. Prog. 1999, 15, 312-318.
- [ 16 ] Jiri Petrak, Robert Ivanek, Ondrej Toman, Radek Cmejla, Jana Cmejlova, Daniel Vyoral, Jan Zivny and Christopher D. Vulpe, **“Deja vu in proteomics. A hit parade of repeatedly identified differentially expressed proteins”**, Proteomics 2008, 8, 1744–1749, DOI 10.1002/pmic.200700919.
- [ 17 ] Alexander Schmidt, Cornelia Ciosto, Josef Kellermann and Friedrich Lottspeich, **“A new strategy for quantitative proteomics using isotope-coded protein labels”**, Max-Planck-Institute of Biochemistry, Martinsried, Germany.
- [ 18 ] Lennart Martens, Henning Hermjakob, Philip Jones, Marcin Adamski, Chris Taylor, David States, Kris Gevaert, Joel Vandekerckhove and Rolf Apweiler, **“PRIDE: The proteomics identifications database”**, Proteomics 2005, 5, 3537–3545, DOI 10.1002/pmic.200401303.
- [ 19 ] Mark P. Molloy, Erin E. Brzezinski, Junqi Hang, Michael T. McDowell, Ruth A. VanBogelen, **“Overcoming technical variation and biological variation in quantitative proteomics”**, Proteomics 2003, 3, 1912–1919, DOI 10.1002/pmic.200300534.
- [ 20 ] Tomas Helikar<sup>1</sup>, Naomi Kochi, John Konvalina and Jim A. Rogers, **“Boolean Modeling of Biochemical Networks”**, The Open Bioinformatics Journal, 2011, 5, 16-25.
- [ 21 ] **“Graph Algorithms Using Depth First Search, Analysis of Algorithms”**, John Reif, Ph.D. Distinguished Professor of Computer Science Duke University.
- [ 22 ] **“Theory of Algorithms / Algorithms in Bioinformatics, Dynamic Programming Part II”**, 2007 Pearson Addison-Wesley.

- [ 23 ] *Federica Eduati, Alberto Corradin, Barbara Di Camillo, Gianna Toffolo, “A Boolean Approach to Linear Prediction for Signaling Network Modeling”, PLoS ONE, September 2010 | Volume 5 | Issue 9 | e12789.*
- [ 24 ] *Albert-László Barabási & Zoltán N. Oltvai, “Network Biology: Understanding the Cell’s Functional Organization”, NATURE REVIEWS | GENETICS, VOLUME 5 | FEBRUARY 2004, 101 – 114, doi:10.1038/nrg1272.*
- [ 25 ] *Nurcan Tuncbag, Scott McCallum, Shao-shan Carol Huang and Ernest Fraenkel , “SteinerNet: a web server for integrating ‘omic’ data to discover hidden components of response pathways”, 25 May 2012, Nucleic Acids Research, 2012, Vol. 40, Web Server issue W505–W509, doi:10.1093/nar/gks445.*
- [ 26 ] *Reka Albert, “Boolean Modeling of Genetic Regulatory Networks”, Lect. Notes Phys. 650, 459–481 (2004).*
- [ 27 ] *Rui-Sheng Wang, Assieh Saadatpour and Reka Albert, “Boolean modeling in systems biology: an overview of methodology and applications”, Phys. Biol. 9 (2012) 055001 (14pp), doi:10.1088/1478-3975/9/5/055001.*
- [ 28 ] *Pedro Mendes and Douglas B. Kell, “Non-linear optimization of biochemical pathways: applications to metabolic engineering and parameter estimation”, Bioinformatics, Vol. 14 no. 10 1998, 869 - 883.*
- [ 29 ] *Richard H. Byrd, Peihuang Lu, Jorge Nocedal and Ciyou Zhu, “A Limited Memory Algorithm for bound Constrained Optimization”, Northwestern University, May 1994.*
- [ 30 ] *Carmen G. Moles, Pedro Mendes and Julio R. Banga, “Parameter Estimation in Biochemical Pathways: A Comparison of Global Optimization Methods”, Genome Res. 2003 13: 2467-2474, 10.1101/gr.1262503.*
- [ 31 ] *Alexander Mitsos, Ioannis N. Melas, Paraskeuas Siminelakis, Aikaterini D. Chairakaki, Julio Saez-Rodriguez, Leonidas G. Alexopoulos, “Identifying Drug Effects via Pathway Alterations using an Integer Linear Programming Optimization Formulation on Phosphoproteomic Data”, 2009, PLoS Comput Biol 5(12): e1000591. doi:10.1371/journal.pcbi.1000591.*
- [ 32 ] *Alexander Mitsos, Ioannis N. Melas, Paraskeuas Siminelakis, Aikaterini D. Chairakaki, Julio Saez-Rodriguez, Leonidas G. Alexopoulos, “Non Linear Programming (NLP) Formulation for Quantitative Modeling of Protein Signal Transduction Pathways”, 2012, PLoS ONE 7(11): e50085. doi:10.1371/journal.pone.0050085.*
- [ 33 ] *Boris Kholodenko, Michael B. Yaffe and Walter Kolch, “Computational Approaches for Analyzing Information Flow in Biological Networks”, 17 April 2012, Science Signaling 5 (220), re1. [DOI: 10.1126/scisignal.2002961].*
- [ 34 ] *Dana Pe’er, Aviv Regev, Gal Elidan and Nir Friedman, “Inferring subnetworks from perturbed expression profiles”, Israel. April 3, 2001.*

- [ 35 ] *Deukwoo Kwon, Mahlet G. Tadesse, Najun Sha, Ruth M. Pfeiffer and Marina Vannucci, “Identifying Biomarkers from Mass Spectrometry Data with Ordinal Outcome”, Cancer Informatics 2007: 3 19–28.*
- [ 36 ] *Mukesh Bansal, Vincenzo Belcastro, Alberto Ambesi-Impiombato and Diego di Bernardo, “How to infer gene networks from expression profiles”, Molecular Systems Biology 3; Article number 78; doi:10.1038/msb4100120.*
- [ 37 ] *Bree B. Aldridge, John M. Burke, Douglas A. Lauffenburger and Peter K. Sorger, “Physicochemical modelling of cell signalling pathways”, Nature Cell Biology, Volume 8, Munber 11, November 2006.*
- [ 38 ] *Melas Ioannis, Kretsos Kosmas, Alexopoulos Leonidas, “Leveraging systems biology approaches in clinical pharmacology”, Biopharmaceutics & Drug Disposition.*
- [ 39 ] *Hiroaki Kitano, “Systems Biology: A Brief Overview”, Science 295, 1662 (2002), DOI: 10.1126/science.1069492.*
- [ 40 ] *Σακελλαρόπουλος Θεόδωρος, “Μοντελοποίηση και Βελτιστοποίηση Θορύβου σε Πολυπλεκτικά Πειράματα Elisa”, Διπλωματική Εργασία, Εργαστήριο Εμβιομηχανικής και Συστημικής Βιολογίας, Ε.Μ.Π., Αθήνα, Οκτώβριος 2012.*
- [ 41 ] *Μεσσήνης Δημήτρης, “Ανάπτυξη Τεχνικής Μέτρησης Βιολογικών Σημάτων Υψηλής Ευαισθησίας”, Διπλωματική Εργασία, Εργαστήριο Εμβιομηχανικής και Συστημικής Βιολογίας, Ε.Μ.Π., Αθήνα, Οκτώβριος 2011.*

### **Ιστότοποι**

- [ 42 ] <http://en.wikipedia.org/wiki/Bioinformatics>
- [ 43 ] [http://en.wikipedia.org/wiki/Cell\\_signaling](http://en.wikipedia.org/wiki/Cell_signaling)
- [ 44 ] [http://en.wikipedia.org/wiki/Signal\\_transduction](http://en.wikipedia.org/wiki/Signal_transduction)
- [ 45 ] <http://www.ndsu.edu/pubweb/~mcclean/plsc431/geneexpress/eukaryex9.htm>
- [ 46 ] <http://www.ncbi.nlm.nih.gov/books/NBK21205/>
- [ 47 ] <http://www.cytoscape.org>
- [ 48 ] <http://biodata.mshri.on.ca/osprey>
- [ 49 ] <http://www.biomedcentral.com/1752-0509/1/2>
- [ 50 ] [http://en.wikipedia.org/wiki/Floyd%E2%80%93Warshall\\_algorithm](http://en.wikipedia.org/wiki/Floyd%E2%80%93Warshall_algorithm)

[ 51 ] <http://datastructures.itgo.com/graphs/transclosure.htm>

[ 52 ] [www.luminexcorp.com/](http://www.luminexcorp.com/)

[ 53 ] <http://bioinformatics.biol.uoa.gr/msc/gr/general.html>