



ΓΕΩΠΟΝΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ
ΤΜΗΜΑ ΒΙΟΤΕΧΝΟΛΟΓΙΑΣ

**ΜΕΤΑΠΤΥΧΙΑΚΟ ΠΡΟΓΡΑΜΜΑ ΣΠΟΥΔΩΝ
«ΒΙΟΛΟΓΙΑ ΣΥΣΤΗΜΑΤΩΝ»**

ΜΕΤΑΠΤΥΧΙΑΚΗ ΔΙΑΤΡΙΒΗ

Εφαρμογές της Θεωρίας της Πληροφορίας
στα Βιολογικά Συστήματα

Μαρία Ι. Κωστάκου

Επιβλέπων Καθηγητής: Ηλίας Ηλιόπουλος

Αθήνα, 2019

**ΜΕΤΑΠΤΥΧΙΑΚΟ ΠΡΟΓΡΑΜΜΑ ΣΠΟΥΔΩΝ
«ΒΙΟΛΟΓΙΑ ΣΥΣΤΗΜΑΤΩΝ»**

ΜΕΤΑΠΤΥΧΙΑΚΗ ΔΙΑΤΡΙΒΗ

**«Εφαρμογές της Θεωρίας της Πληροφορίας στα Βιολογικά
Συστήματα»**

«Applications of Information Theory in Biological Systems»

Μαρία Ι. Κωστάκου

Τριμελής Εξεταστική Επιτροπή

Επιβλέπων Καθηγητής

Ηλίας Ηλιόπουλος, Καθηγητής ΓΠΑ

Μέλη Επιτροπής

Εμμανουήλ Φλεμετάκης, Αν. Καθηγητής ΓΠΑ

Δημήτριος Βλαχάκης, Επίκ. Καθηγητής ΓΠΑ

Περίληψη

Η Βιολογία Συστημάτων είναι πλέον ο νέος τρόπος προσέγγισης των βιολογικών φαινομένων που κατανοεί και αποδέχεται την εγγενή πολυπλοκότητά τους χωρίς να την περιορίζει. Η κατανόηση σε επίπεδο συστήματος είναι μια ισχυρή εναλλακτική στην παραδοσιακή αναγωγική προσέγγιση που χρησιμοποιούν σήμερα πολλοί κλάδοι της Βιολογίας. Η συστημική θεώρηση απαιτεί στενή συνεργασία με άλλους επιστημονικούς κλάδους όπως τα Μαθηματικά, η Φυσική και η Επιστήμη των Υπολογιστών. Στο πλαίσιο αυτό μελετάται η Θεωρία της Πληροφορίας του Claude Shannon η οποία τις τελευταίες δεκαετίες έχει σημειώσει σημαντική πρόοδο στην παροχή εννοιολογικών πλαισίων για τον προσδιορισμό της δυναμικής των Βιολογικών Συστημάτων.

Στην παρούσα διπλωματική εργασία γίνεται μια συστηματική ανασκόπηση που αφορά στη συνεισφορά της Θεωρίας της Πληροφορίας στην κατανόηση και μελέτη διάφορων βιολογικών προβλημάτων. Τα Βιολογικά Συστήματα που εξετάζονται κατατάσσονται με βάση τον τύπο των στοιχείων του συστήματος και τις αλληλεπιδράσεις μεταξύ τους, σε Γενετικά Ρυθμιστικά Δίκτυα, Μεταβολικά, Πρωτεϊνικά και Σηματοδοτικά Δίκτυα. Οι μελέτες που παρουσιάζονται σε αυτή την ανασκόπηση δείχνουν ότι οι προσεγγίσεις της Θεωρίας της Πληροφορίας είναι σε θέση να παράγουν νέα γνώση συμβάλλοντας σε μια πιο συνολική κατανόηση των Βιολογικών Συστημάτων.

Λέξεις Κλειδιά: Θεωρία Πληροφορίας Shannon, Εντροπία, Αμοιβαία Πληροφορία, Γενετικά Ρυθμιστικά Δίκτυα, Μεταβολικά Δίκτυα, Πρωτεϊνικά Δίκτυα, Σηματοδοτικά Δίκτυα

Abstract

Systems Biology is the new way of approaching biological phenomena that takes into account the inherent complexity. Understanding at the system level is a powerful alternative to the traditional reductionist approach, used in many branches of Biology. System thinking requires interdisciplinary approaches and brings together principles and concepts from areas of study as Mathematics, Physics and Computer Science. In this context, Claude Shannon's Information Theory has made significant progress in providing conceptual frameworks for identifying the dynamics of biological systems.

This thesis provides a systematic review of most of the existing studies on the contribution of Information Theory in inferring and understanding of various biological problems. Biological systems are classified based on the type of system components and the interactions between them in Gene Regulatory, Metabolic, Protein-Protein Interaction and Signaling Networks. The related studies show that information-theoretic methods can generate new knowledge, contributing to a better understanding of biological systems.

Keywords: Shannon's Information Theory, Entropy, Mutual information, Gene Regulatory Networks, Protein-Protein interaction Networks, Signaling Networks

Στη μνήμη της γιαγιάς μου Κωνσταντίνας,

Ευχαριστίες

Με αφορμή τη διεκπεραίωση και ολοκλήρωση της παρούσας διατριβής, θα ήθελα να εκφράσω την ευγνωμοσύνη μου σε όλους τους καθηγητές του μεταπτυχιακού προγράμματος σπουδών "Βιολογία Συστημάτων" για τις γνώσεις που μου παρείχαν. Ειδικότερα, θα ήθελα να ευχαριστήσω τον καθηγητή μου κ. Κωνσταντίνο Γιαλούρη για την αμέριστη συμπαράσταση, υποστήριξη και καθοδήγησή του καθ' όλη τη διάρκεια των σπουδών μου και σε όλα τα στάδια εκπόνησης της διπλωματικής μου διατριβής. Επιπλέον, ευχαριστώ θερμά τον καθηγητή κ. Ηλία Ηλιόπουλο και τα μέλη της τριμελούς επιτροπής μου κ. Δημήτρη Βλαχάκη και κ. Εμμανουήλ Φλεμετάκη για τις πολύτιμες συμβουλές και τα σχόλιά τους στην παρούσα διατριβή.

Τέλος, ευχαριστώ θερμά τους φίλους και την οικογένειά μου για την υπομονή και την συμπαράστασή τους στην προσπάθειά μου.

Περιεχόμενα

1	Κεφάλαιο 1. Εισαγωγή	9
1.1	Η Γενική Θεωρία Συστημάτων.....	10
1.2	Αναλογίες και Αυτό-ομοιότητα	11
1.3	Θερμοδυναμική και Πολύπλοκα Συστήματα	12
1.4	Πληροφορία και Εντροπία στα Βιολογικά Συστήματα	14
2	Κεφάλαιο 2. Βασικά στοιχεία της Θεωρίας της Πληροφορίας	16
2.1	Η Θεωρία της Πληροφορίας του Shannon.....	16
2.2	Εντροπία Shannon	18
2.3	Σύνθετη Εντροπία.....	19
2.4	Υπό Συνθήκη Εντροπία	19
2.5	Σχετική Εντροπία	20
2.6	Αμοιβαία Πληροφορία	20
2.7	Διακριτό Κανάλι και Χωρητικότητα Καναλιού	22
2.8	Θεωρία Ρυθμού-Παραμόρφωσης.....	23
2.9	Εντροπία Kolmogorov-Sinai και Εντροπία Δικτύου.....	23
2.10	Αρχή Μέγιστης Εντροπίας	24
3	Κεφάλαιο 3. Εφαρμογές της Θεωρίας της Πληροφορίας	25
3.1	Γενετικά Ρυθμιστικά Δίκτυα	27
3.1.1	Εισαγωγή.....	27
3.1.2	Πίνακας Αμοιβαίας Πληροφορίας.....	28
3.1.3	Ο αλγόριθμος REVEAL.....	29
3.1.4	Δίκτυα Συνάφειας (Relevance Networks-RELNET)	29
3.1.5	Ο αλγόριθμος C3NET	34
3.2	Μεταβολικά Δίκτυα	34
3.2.1	Εισαγωγή.....	34
3.2.2	Ανάλυση Μεταβολικών Ροών	36
3.2.3	Αντίστροφη Μηχανική Μεταβολικών Δικτύων	39
3.2.4	Μελέτη Ευρωστίας Δικτύου	42
3.3	Πρωτεϊνικά Δίκτυα	43
3.3.1	Εισαγωγή.....	43
3.3.2	Αναγνώριση Πρωτεϊνικών Συμπλόκων	46
3.3.3	Αναγνώριση Βιοδεικτών	48
3.3.4	Μελέτη Ευρωστίας Δικτύου	51

3.4	Σηματοδοτικά Δίκτυα	52
3.4.1	Εισαγωγή.....	52
3.4.2	Μέτρηση της ποσότητας της μεταφερόμενης πληροφορίας	55
3.4.3	Εκτίμηση της Χωρητικότητας Καναλιού στη μετάδοση σήματος	56
3.4.4	Μελέτη της κατευθυντικής απόκρισης χημειοτακτικών κυττάρων	58
4	Κεφάλαιο 4. Επίλογος	60
5	Βιβλιογραφία.....	64

Κεφάλαιο 1. Εισαγωγή

Η παρούσα διπλωματική εργασία αποτελεί μια συστηματική ανασκόπηση της συνεισφοράς της Θεωρίας της Πληροφορίας στη μελέτη των βιολογικών συστημάτων. Η Θεωρία της Πληροφορίας ενώ αναπτύχθηκε για πρώτη φορά στον κλάδο των επικοινωνιών έχει ευρεία εφαρμοσιμότητα σε διάφορους τομείς της επιστήμης και μπορεί να αξιοποιηθεί αποτελεσματικά σε άλλους τομείς όπως τα Μαθηματικά, η Επιστήμη Υπολογιστών, τα Οικονομικά, η Φυσική, Βιοπληροφορική.

Στο πρώτο κεφάλαιο τίθεται το θεωρητικό υπόβαθρο για τη μελέτη των πολύπλοκων συστημάτων. Περιγράφονται τα γενικά χαρακτηριστικά πολύπλοκων συστημάτων με έμφαση στα Βιολογικά Συστήματα και γίνεται μια πρώτη σύνδεση μεταξύ θερμοδυναμικών εννοιών και Πληροφορίας.

Στο δεύτερο κεφάλαιο περιγράφονται βασικές έννοιες και ορισμοί της Θεωρίας της Πληροφορίας. Η Θεωρία της Πληροφορίας σχετίζεται με την ικανότητα ενός συστήματος να μεταδίδει πληροφορίες. Βασικά στοιχεία που μελετώνται είναι η ροή, η αξιοποίηση και η αποθήκευση της πληροφορίας τα οποία είναι απαραίτητα για την κατανόηση των αρχών που διέπουν τις δυναμικές διεργασίες των βιολογικών συστημάτων. Οι έννοιες αυτές πρόκειται να χρησιμοποιηθούν στη συνέχεια σε εφαρμογές που αφορούν προβλήματα της Βιολογίας Συστημάτων.

Στο τρίτο κεφάλαιο γίνεται μια συστηματική ανασκόπηση της υπάρχουσας βιβλιογραφίας. Περιγράφονται ορισμένες επιτυχημένες εφαρμογές της Θεωρίας της Πληροφορίας σε Βιολογικά Συστήματα όπως τα Γενετικά Ρυθμιστικά, τα Μεταβολικά, τα Πρωτεϊνικά και τα Σηματοδοτικά Δίκτυα. Στα Γενετικά Ρυθμιστικά Δίκτυα, οι μέθοδοι της Θεωρίας της Πληροφορίας προσεγγίζουν κατά κύριο λόγο το πρόβλημα της ανακατασκευής δικτύων. Αντιστοίχως, στα Μεταβολικά Δίκτυα οι μέθοδοι που εφαρμόζονται αφορούν τη δυναμική των Μεταβολικών Δικτύων και την ανθεκτικότητά τους στις διαταραχές. Στα Πρωτεϊνικά Δίκτυα προσεγγίζονται προβλήματα όπως η αναγνώριση πρωτεϊνικών συμπλόκων, η αναγνώριση βιοδεικτών και η μελέτη ευρωστίας δικτύου. Επιπλέον, μέθοδοι με βάση τη Θεωρία της Πληροφορίας εφαρμόζονται στα Σηματοδοτικά Δίκτυα για στη μελέτη του ρυθμού μετάδοσης σημάτων, την εκτίμηση της Χωρητικότητας Καναλιού των δικτύων αυτών και τη μελέτη χημειοτακτικών μηχανισμών.

Η ικανότητα ανάλυσης και κατανόησης της δυναμικής των βιολογικών συστημάτων με ελλιπή και θορυβώδη δεδομένα αποτελεί μεγάλη πρόκληση για τις μεθοδολογίες που χρησιμοποιούνται για να περιγράψουν κάθε ένα από τα παραπάνω προβλήματα. Η παρούσα ανασκόπηση έχει σαν στόχο να προσεγγίσει τη στατιστική-στοχαστική φύση βιολογικών προβλημάτων που έχουν μελετηθεί θεωρητικά, αναδεικνύοντας την υπολογιστική αξία της Θεωρίας της Πληροφορίας στη μοντελοποίησή τους.

2.1 Η Γενική Θεωρία Συστημάτων

Η Γενική Θεωρία Συστημάτων (ΓΘΣ) αναπτύχθηκε από τον Αυστριακό βιολόγο Ludwig von Bertalanffy (1901-1972) τη δεκαετία του 1940. Ο Bertalanffy εργάστηκε με σκοπό την ενοποίηση της επιστήμης, αναζητώντας κοινές αρχές και νόμους που έχουν ισχύ σε διαφορετικά φαινόμενα, φυσικά, βιολογικά και κοινωνικά. Στο βιβλίο του με τίτλο *"General System theory: Foundations, Development, Applications"* αναφέρει: *"...υπάρχουν μοντέλα, αρχές και νόμοι που εφαρμόζονται σε γενικευμένα συστήματα ή στις υποκατηγορίες τους, ανεξαρτήτως του είδους τους, τη φύση των συστατικών στοιχείων τους, και τις σχέσεις των "δυνάμεων" μεταξύ τους. Φαίνεται εύλογη η αναζήτηση μιας θεωρίας όπου καθολικές αρχές εφαρμόζονται γενικά σε κάθε σύστημα."* Υποστηρίζει ότι είναι απαραίτητο στη σύγχρονη Βιολογία να μελετώνται όχι μόνο τα μέρη και οι διεργασίες ξεχωριστά, αλλά να μελετώνται μαζί ως σύνολο λαμβάνοντας υπόψη τη δομή και την οργάνωση του εκάστοτε βιολογικού συστήματος. Αυτό προτείνεται διότι τα διάφορα μέρη και διεργασίες έχουν διαφορετική συμπεριφορά αν ληφθούν ως απομονωμένες οντότητες και διαφορετική ως σύνολο. Η Γενική Θεωρία Συστημάτων ασχολείται με τα προβλήματα ενός συστήματος τα οποία βασίζονται στην πληρότητά του, δηλαδή στην λειτουργία του ως σύνολο. Ειδικότερα επικεντρώνεται στη διάταξη και οργάνωση των επιμέρους στοιχείων και στις μεταξύ τους σχέσεις οι οποίες τα συνδέουν και δημιουργούν ένα σύστημα. Η χαρακτηριστική αυτή οργάνωση των επιμέρους στοιχείων καθορίζει το σύστημα και είναι ανεξάρτητη της φύσης των στοιχείων της.

Όπως υποστηρίζει ο Bertalanffy, μία ενοποιητική σύλληψη και κατανόηση του κόσμου δεν θα μπορούσε να βασιστεί στην μάταιη ιδέα της αναγωγής όλων των επιπέδων της πραγματικότητας στο επίπεδο της Φυσικής. Εντούτοις, η ισομορφία των νόμων στα διαφορετικά επιστημονικά πεδία είναι κάτι που πρέπει να λαμβάνεται υπόψη από τους μελετητές καθώς παρεμφερείς έννοιες, μοντέλα και νόμοι εμφανίζονται συχνά σε τελείως διαφορετικούς επιστημονικούς κλάδους. Η Γενική Θεωρία Συστημάτων δεν αποτελεί απλώς μια εφαρμογή γνωστών μαθηματικών νόμων σε διαφορετικών ειδών προβλήματα, αλλά στοχεύει στη λύση προβλημάτων που δεν έχουν απαραιτήτως προφανή λύση. Η κλασική επιστήμη εστίαζε κυρίως σε φαινόμενα είτε αιτιοκρατικά είτε στοχαστικά χωρίς να εξετάζει τη σύνδεση μεταξύ των στοιχείων και διεργασιών τους. Στο σημείο αυτό η Γενική Θεωρία Συστημάτων καλείται να προσδώσει, έναν νέο τρόπο μαθηματικής σκέψης που να καθορίζει τις έννοιες που σχετίζονται με την οργάνωση των οντοτήτων που συνιστούν τα συστήματα και ταυτόχρονα να τις ποσοτικοποιεί (Von Bertalanffy, 1968).

Θεωρείται ότι ο Bertalanffy είναι ο πρόδρομος της Βιολογίας Συστημάτων, καθώς εισήγαγε στη Βιολογία τη συστημική θεώρηση από το 1928. Η Βιολογία Συστημάτων είναι ένα διεπιστημονικό πεδίο της επιστήμης της Βιολογίας που επικεντρώνεται στις αλληλεπιδράσεις των βιολογικών συστημάτων με βάση μια ολιστική (αντί για αναγωγική) προσέγγιση. Δηλαδή, η Βιολογία Συστημάτων έχει μετατοπίσει το ενδιαφέρον της στην

ολοκληρωμένη μελέτη των βιολογικών δράσεων (γενετικών, πρωτεϊνικών, μεταβολικών κ.α) που αλληλεπιδρούν συνεργιστικά για την επίτευξη των βιολογικών λειτουργιών. Ακόμη, μεταξύ των "συνόρων" των βιολογικών συστημάτων, αναδύονται χρήσιμες πληροφορίες που δεν ήταν εμφανείς με την κλασική προσέγγιση της Βιολογίας. Κατά τον Bertalanffy είναι σχεδόν αδύνατο να υποβιβάσουμε απλουστευτικά τα βιολογικά επίπεδα στο χαμηλότερο επίπεδο των κατασκευασμάτων των νόμων της Φυσικής. Ωστόσο, είναι πολύ πιθανό να μπορούμε να βρούμε μοντέλα και νόμους μέσα σε καθένα από τα διάφορα επίπεδα. Η ενοποιητική αρχή είναι ότι μπορούν να βρεθούν μορφές οργάνωσης σε όλα τα επίπεδα, γι' αυτό, μια σημαντική επιδίωξη της Βιολογίας Συστημάτων είναι η μοντελοποίηση και η διερεύνηση των *αναδυόμενων ιδιοτήτων* που προκύπτουν μεταξύ κλιμάκων (Von Bertalanffy, 1962).

2.2 Αναλογίες και Αυτό-ομοιότητα

Κάνοντας μια σύντομη αναδρομή στο παρελθόν, αρκετοί φιλόσοφοι πίστευαν ότι είναι καλό στη φιλοσοφία να αναζητούμε τις ομοιότητες, ακόμα και σε πράγματα που απέχουν πολύ μεταξύ τους. Ο Αριστοτέλης υποστήριζε ότι οι ομοιότητες αποτελούν τα "όργανα" των συλλογισμών στα οποία στηρίζονται οι υποθέσεις και είναι χρήσιμες στους ρητορικούς συλλογισμούς. Ωστόσο, αποκαλύπτει την αδυναμία της αναλογικής επιχειρηματολογίας από την πλευρά της απόδειξης και τονίζει ότι χρειάζεται ιδιαίτερη προσοχή σε πιθανές "επιφανειακές" ομοιότητες των αναλογιών καθώς τα διαφορετικά αντικείμενα δεν ταυτίζονται απλά παραλληλίζονται μέσω της σύγκρισης παρόμοιων σχέσεων που έχουν (Lloyd, 1966).

Η μελέτη των πολύπλοκων συστημάτων συχνά βασίζεται στη χρήση αναλογιών. Η αναλογία είναι ένα είδος ομοιότητας το οποίο μας βοηθά στη κατανόηση και την ανάλυση άγνωστων συστημάτων με βάση την υπάρχουσα γνώση από άλλα συστήματα καλά μελετημένα. Θα μπορούσαμε να πούμε ότι είναι η ομοιότητα σε ένα πιο σαφές και πιο εννοιολογικό πλαίσιο ενώ η βασική διαφορά αναλογίας και ομοιότητας έγκειται κατά κύριο λόγο στις προθέσεις του στοχαστή. Για παράδειγμα, όμοια αντικείμενα έχουν κάποια κοινά γνωρίσματα και αν επρόκειτο τα γνωρίσματα αυτά να απομονωθούν, τότε θεωρούμε τα γνωρίσματα αυτά ως ανάλογα. Έτσι, δύο συστήματα είναι ανάλογα εάν συμφωνούν σε σαφώς καθορισμένες σχέσεις μεταξύ των αντίστοιχων μερών τους (Ρόγια, 1990). Εντούτοις, η χρήση των αναλογιών χρειάζεται ιδιαίτερη προσοχή έτσι ώστε να εφαρμόζεται σωστά η μεθοδολογία προσέγγισης πολύπλοκων συστημάτων, λαμβάνοντας υπόψη τις ιδιαιτερότητες του προς μελέτη συστήματος (Nicolis & Nicolis, 2009).

Αξιοσημείωτο χαρακτηριστικό των πολύπλοκων συστημάτων είναι ότι τα μέρη από τα οποία αποτελείται ένα αντικείμενο μπορεί να μοιάζουν με το σύνολο. Αντικείμενα των

οποίων τα μέρη του συνόλου έχουν παρόμοια χαρακτηριστικά με το σύνολο λέγονται αυτό-όμοια και έχουν ισχύ σε διαφορετικά επίπεδα και κλίμακες (Mittleton-Kelly, 2004). Κατά τον Lin (Lin, 1998) στην περίπτωση μελέτης ενός πιθανοτικού μοντέλου μπορεί να δοθεί ένας πληροφοριακός-θεωρητικός ορισμός για την ομοιότητα. Σύμφωνα με το *Θεώρημα Ομοιότητας* ισχύει ότι η ομοιότητα μεταξύ ενός στοιχείου A και ενός στοιχείου B, $sim(A,B)$, με πιθανότητα P μετράται από τον λόγο μεταξύ της ποσότητας πληροφοριών που απαιτείται για να δηλωθεί η συνάφεια των A και B και των πληροφοριών που απαιτούνται για να περιγραφούν πλήρως τα A και B ως:

$$sim(A, B) = \frac{\log P(\text{common}(A, B))}{\log P(\text{description}(A, B))}$$

2.3 Θερμοδυναμική και Πολύπλοκα Συστήματα

Ως *Σύστημα* ορίζεται ένα σύνολο στοιχείων που αλληλεπιδρούν ή σχετίζονται μεταξύ τους. Το σύστημα έχει σαφώς καθορισμένα όρια, τα οποία καθορίζουν ποια στοιχεία αποτελούν μέρος του ή όχι, τα δε στοιχεία που βρίσκονται εκτός των ορίων αποτελούν το περιβάλλον του συστήματος. Η κύρια κατηγοριοποίηση των συστημάτων περιλαμβάνει τα *ανοιχτά συστήματα* τα όποια ανταλλάσσουν ύλη, ενέργεια ή πληροφορία με το περιβάλλον τους και τα *κλειστά συστήματα* τα οποία εν αντιθέσει δεν αλληλεπιδρούν με το περιβάλλον τους (Nicolis & Nicolis, 2007).

Πολύπλοκο Σύστημα είναι το σύστημα που απαρτίζεται από πολλά υποσυστήματα τα οποία μπορεί να αλληλεπιδρούν μεταξύ τους. Τα πολύπλοκα συστήματα παρουσιάζουν μια σχετική δυσκολία στη μοντελοποίηση τους εξαιτίας των ιδιαίτερων σχέσεων μεταξύ των στοιχείων τους και των αλληλεπιδράσεων με το περιβάλλον τους. Οι Nicolis και Prigogine υποστηρίζουν ότι η πολυπλοκότητα είναι ένα φαινόμενο που έχει βαθιές ρίζες στους νόμους της φύσης, όπου συστήματα με μεγάλο αριθμό αλληλεπιδρώντων υπομονάδων είναι πανταχού παρόντα (Nicolis & Nicolis, 2007; Prigogine, 1987). Για να κατανοηθεί ποσοτικά η έννοια της πολυπλοκότητας χρησιμοποιούνται εργαλεία της Στατιστικής Φυσικής και της Επιστήμης των Υπολογιστών. Στο πλαίσιο αυτό μπορούμε να ορίσουμε την *πολυπλοκότητα* ως την ποσότητα της πληροφορίας που είναι απαραίτητη για την περιγραφή ενός συστήματος, σε δεδομένο επίπεδο κλίμακας (Bar-Yam, 1999).

Οι Θερμοδυναμικοί Νόμοι στην κλασσική Φυσική ισχύουν για τα κλειστά συστήματα. Συγκεκριμένα, κατά τον Πρώτο Θερμοδυναμικό Νόμο, η συνολική ενέργεια ενός απομονωμένου συστήματος (δεν ανταλλάσει ενέργεια και ύλη με το περιβάλλον) παραμένει σταθερή για όλες τις αλλαγές που πραγματοποιούνται σε αυτό. Αυτό αποτελεί μια διατύπωση της *Αρχής Διατήρησης της Ενέργειας* και δηλώνει ότι η ενέργεια δεν μπορεί

να δημιουργηθεί ή να καταστραφεί, μόνο να μετατραπεί από τη μία μορφή στην άλλη (Dolev & Elitzur, 2000; Silow & Jørgensen, 2011).

Ο Δεύτερος Θερμοδυναμικός Νόμος ορίζει ότι κατά τη διάρκεια οποιασδήποτε μεταβολής ενός απομονωμένου συστήματος η Εντροπία αυξάνεται (ή διατηρείται κατά τις αντιστρεπτές μεταβολές). Η Εντροπία του συστήματος αποκτά τη μέγιστη τιμή της όταν επέλθει *θερμοδυναμική ισορροπία*. Η δεύτερη Αρχή της Θερμοδυναμικής μπορεί να διατυπωθεί με διαφορετικούς τρόπους. Αφενός η Εντροπία όπως θα δούμε και στη συνέχεια είναι μέτρο της πιθανότητας, και σε δεδομένο σύστημα τείνει στη κατάσταση της πιο πιθανής κατανομής που είναι η τάση προς τη *Μέγιστη Εντροπία* ή τη *μέγιστη πιθανή κατανομή*, δηλαδή τη *μέγιστη διαταραχή* του συστήματος. Ο Eddington επεσήμανε τη σημασία του 2ου Θερμοδυναμικού Νόμου κατατάσσοντάς τον στην ανώτατη θέση ανάμεσα στους νόμους της φύσης, αναγνωρίζοντας τις απεριόριστες δυνατότητες που προσφέρει στην επιστημονική έρευνα ένας "τόσος απλός" νόμος. Υποστηρίζει ότι "*αν βρεθεί κάποια θεωρία που αντικρούεται με τον νόμο αυτό τότε είναι καταδικασμένη να αποτύχει παταγωδώς*" (Eddington, 1948).

Κατά τον Τρίτο Θερμοδυναμικό Νόμο, η Εντροπία ενός συστήματος στην απόλυτη θερμοκρασία του μηδενός $T=0$ είναι μηδέν ($S=0$). Η φυσική ερμηνεία ορίζει ότι είναι αδύνατο να φτάσει η θερμοκρασία ενός συστήματος στο απόλυτο μηδέν της κλίμακας Kelvin σε πεπερασμένο αριθμό βημάτων (Dolev & Elitzur, 2000; Silow & Jørgensen, 2011).

Παραδείγματα που σχετίζονται άμεσα με τον Δεύτερο Θερμοδυναμικό Νόμο μπορεί να είναι ένα θερμό αντικείμενο που ψύχεται, ένα αέριο που διαχέεται μέσω του αέρα, ένα υγρό που αναμειγνύεται με ένα άλλο. Όλες αυτές είναι αυθόρμητες διαδικασίες οι οποίες έχουν σαν κοινό χαρακτηριστικό ότι η ενέργεια τείνει να διασπείρεται με το πέρασ του χρόνου και σαφώς πρόκειται για μη αντιστρεπτές διαδικασίες. Η Εντροπία είναι ένα μέτρο της τάσης αυτής που μετρά το πόσο διασκορπισμένη είναι η ενέργεια ανάμεσα στα σωματίδια του συστήματος και πώς αυτά διασπείρονται. Καθώς τα σωματίδια στο σύστημα κινούνται και αλληλεπιδρούν μεταξύ τους είναι πολύ πιο πιθανό να διασπείρονται στο χώρο από το να μείνουν συγκεντρωμένα, μέχρι τελικά το σύστημα θα έρθει σε κατάσταση *Μέγιστης Εντροπίας* ή *αλλιώς θερμοδυναμικής ισορροπίας*, όπου η ενέργεια είναι κατανεμημένη ομοιόμορφα.

Παρόλο που η Εντροπία αυξάνεται με το χρόνο σε ένα κλειστό σύστημα, ένα ανοιχτό σύστημα μπορεί να την κρατήσει στο εσωτερικό του χαμηλά, αυξάνοντας ταυτόχρονα την Εντροπία της γύρω περιοχής. Μέχρι πρόσφατα οι θερμοδυναμικοί νόμοι ήταν δυνατό να εφαρμοστούν μόνο σε κλειστά συστήματα. Στη μελέτη των ανοιχτών συστημάτων έκανε σημαντική πρόοδο ο Ilya Prigogine ο οποίος τιμήθηκε με Νόμπελ Χημείας το 1977 για τον προσδιορισμό των *δομών έκλυσης* (dissipative structures) και του ρόλου τους σε συστήματα μακράν της θερμοδυναμικής ισορροπίας. Τις τελευταίες δεκαετίες, οι θερμοδυναμικές

έννοιες έχουν επεκταθεί έτσι ώστε να μπορούν να λύσουν προβλήματα που αφορούν ανοιχτά συστήματα, πέρα από τη θερμοδυναμική ισορροπία (Weber et al., 1989).

Τα Πολύπλοκα Συστήματα παρουσιάζουν τη δική τους ιδιαίτερη φαινομενολογία, με βασικό στοιχείο τις *αναδυόμενες ιδιότητες* οι οποίες ορίζονται ως πρόσθετα χαρακτηριστικά που διαθέτει ένα σύστημα εκτός του συνόλου των ιδιοτήτων των στοιχείων που περιέχει το σύστημα δηλαδή, είναι κάτι παραπάνω από το άθροισμα των στοιχείων τους (Nicolis & Nicolis, 2009). Πιο συγκεκριμένα, διακρίνουμε την *αναδυόμενη πολυπλοκότητα* σαν αποτέλεσμα της συμπεριφοράς πολλών και απλών στοιχείων που αλληλεπιδρούν με τρόπο που το σύνολο που συνιστούν είναι πολύπλοκο. Διακρίνουμε ακόμη την *αναδυόμενη απλότητα* ως αποτέλεσμα πολύπλοκων στοιχείων τα οποία σε μικρή κλίμακα μπορεί να συμπεριφέρονται με τρόπο πολύπλοκο, αλλά σε μεγαλύτερη κλίμακα η συμπεριφορά του συνόλου να είναι απλή (Bar-Yam, 1999).

Βασικά χαρακτηριστικά της πολυπλοκότητας είναι η μη αντιστρεψιμότητα και η στοχαστικότητα (Prigogine, 1987). Η συνύπαρξη στο ίδιο σύστημα γνωστών και άγνωστων στοιχείων ή "τάξης" και "αταξίας" δυσκολεύουν την πρόβλεψη της δυναμικής του συστήματος και καθιστούν τη μη-γραμμικότητα (δηλαδή η εξίσωση που περιγράφει τη σχέση των μεταβλητών εμφανίζει όρους με δύναμη μεγαλύτερη του ενός) ως αναγκαία συνθήκη για τη μελέτη της πολυπλοκότητας. Μία ακόμη γενική αρχή των πολύπλοκων συστημάτων είναι ότι βρίσκονται στα όρια μεταξύ των ντετερμινιστικών και των πιθανοκρατικών προσεγγίσεων και ακολουθούν τους δικούς τους κανόνες όσον αφορά τη δομή και την οργάνωσή τους (Nicolis & Nicolis, 2009).

2.4 Πληροφορία και Εντροπία στα Βιολογικά Συστήματα

Βασικό χαρακτηριστικό των έμβιων συστημάτων είναι η ικανότητά τους να επεξεργάζονται πληροφορίες. Ο στατιστικός-πληροφορικός χαρακτήρας των θερμοδυναμικών εννοιών έχει βοηθήσει στην ερμηνεία και την ενοποίηση των βιολογικών συστημάτων σε όλες τις εκφάνσεις τους και σε κάθε επίπεδο της βιολογικής κλίμακας, από τις κυτταρικές οδούς μέχρι και τις οικολογικές διεργασίες. Έτσι, μπορούν να μελετηθούν οι μεταβολές της ενέργειας οι οποίες πραγματοποιούνται συνεχώς σε όλα τα έμβια συστήματα, οι έννοιες της τάξης και της αταξίας, όπως χαρακτηριστικά είναι η ζωή και ο θάνατος των οργανισμών, η προσαρμογή πληθυσμών και η εξαφάνιση. Στο πλαίσιο αυτό έχουν υιοθετηθεί τρεις βασικές αρχές για τα Βιολογικά Συστήματα:

1. Τα Βιολογικά Συστήματα σταθεροποιούνται μακριά από τη θερμοδυναμική ισορροπία μέσω της αυτοοργάνωσης¹ και αποτελούν δομές έκλυσης².
2. Τα Βιολογικά Συστήματα που εκμεταλλεύονται την υπάρχουσα "ελεύθερη" ενέργεια από το περιβάλλον τους πιο αποτελεσματικά συγκριτικά με τους ανταγωνιστές τους, δύνανται να αφήσουν περισσότερους απογόνους. Αυτό αφορά όχι μόνο στην πιο αποδοτική χρήση των πόρων αλλά και στην ενίσχυση της αυτονομίας και της ακεραιότητάς τους.
3. Η πολυπλοκότητα της πληροφορίας με τη μορφή γενετικής ποικιλότητας υπόκειται στον Δεύτερο Θερμοδυναμικό Νόμο, κατά τον οποίο αποτρέπεται η αλάνθαστη αντιγραφή σε ανοιχτά συστήματα με πεπερασμένες πηγές ενέργειας υπαγορεύοντας την επέκταση του χώρου της γενετικής φάσης στον χρόνο (Weber et al., 1989).

Τα Βιολογικά Συστήματα κατά τον Bertalanffy είναι σύνθετα ως προς την επεξεργασία της πληροφορίας. Η δομή τους υποστηρίζει την εισαγωγή σε αυτά ύλης με υψηλής περιεκτικότητας πληροφορίας και ενέργειας με χαμηλή Εντροπία, ενώ εξέρχονται από το σύστημα ως ύλη και ενέργεια με υψηλή Εντροπία. Συνοπτικά, η συνολική πληροφορία που εισέρχεται στο βιολογικό σύστημα διασπάται μέσα σε αυτό έτσι ώστε να διατηρηθεί η κατάσταση υψηλής οργάνωσής του. Οι οργανισμοί διαθέτουν τα ακόλουθα τρία βασικά στοιχεία: την οργάνωση, τη δυναμική ροή των διεργασιών τους και τη δυνατότητα διατήρησης της πληροφορίας. Η πιο σημαντική διαφορά μεταξύ βιολογικών και φυσικών συστημάτων είναι ότι στα πρώτα ορισμένες πληροφορίες διατηρούνται στο χρόνο, αντίθετα με τα πιο πολλά φυσικά συστήματα. Κάθε οργανισμός δηλαδή, προέρχεται από άλλους του ίδιου είδους και φέρει χαρακτηριστικά του παρελθόντος πολλά εκ των οποίων είναι κοινά σε γενεές που προηγήθηκαν (Von Bertalanffy, 1952).

Ένα από τα πιο θεμελιώδη κριτήρια για τους έμβιους οργανισμούς είναι ότι αποτελούν ανοιχτά συστήματα (Prigogine, Nicolis, & Babloyantz, 1972) καθώς, με τον όρο αυτό υποδηλώνεται η συνεχής ανταλλαγή ύλης και ενέργειας με το περιβάλλον τους. Οι έμβιοι οργανισμοί δεν αποτελούν στατικές δομές αλλά δομές στις οποίες λαμβάνουν χώρα συνεχείς διεργασίες όπου δομικά συστατικά και ουσίες που αποδίδουν ενέργεια διασπώνται και αναδομούνται. Οι διεργασίες αυτές είναι απαραίτητες για τον έμβιο οργανισμό έτσι ώστε να διατηρείται στη λεγόμενη σταθερή κατάσταση (steady state), αποφεύγοντας έτσι την αύξηση της Εντροπίας και αυξάνοντας πιθανότατα την τάξη και την οργάνωσή του. Τα άλλα χαρακτηριστικά των έμβιων οργανισμών όπως ο μεταβολισμός, η ανάπτυξη, η απόκριση σε ερεθίσματα, η αναπαραγωγή κ.α. προκύπτουν ουσιαστικά από τη

¹ Η αυθόρμητη ανάδυση της συνολικής συνοχής από τις τοπικές αλληλεπιδράσεις ανάμεσα σε αρχικά ανεξάρτητες συνιστώσες (Heylighen, 2002).

² Μοτίβα όπως π.χ. τα Bénard cells, τα οποία παρουσιάζουν δυναμική αυτοοργάνωση. Τέτοιες δομές είναι αναγκαστικά ανοιχτά συστήματα (Heylighen, 2002).

βασική ανάγκη των οργανισμών να διατηρούνται στη σταθερή κατάσταση (Von Bertalanffy, 1968).

Η αμοιβαία σχέση Εντροπίας και Πληροφορίας, καλείται να προσδώσει μια νέα οπτική στα αινίγματα της Βιολογίας. Η ροή της πληροφορίας στον χώρο, στον χρόνο και μεταξύ συστημάτων και κλιμάκων παρέχει μια πολύ σημαντική επεξηγηματική δυνατότητα σε φαινόμενα που διαφορετικά θα φαίνονταν μεταξύ τους ασυσχέτιστα (Dittrich, 2015).

Κεφάλαιο 2. Βασικά στοιχεία της Θεωρίας της Πληροφορίας

3.1 Η Θεωρία της Πληροφορίας του Shannon

Η Θεωρία της Πληροφορίας η οποία αναπτύχθηκε για πρώτη φορά από τον Claude Shannon το 1948 στη δημοσίευσή του με τίτλο “*A Mathematical Theory of Communication*” έχει βρει τις τελευταίες δεκαετίες πολλαπλές εφαρμογές σε διαφορετικούς τομείς της επιστήμης όπως η Στατιστική, η Πληροφορική, τα Οικονομικά, η Φυσική, η Τεχνητή Νοημοσύνη και προσφάτως η Βιοπληροφορική και η Βιολογία Συστημάτων. Η Θεωρία της Πληροφορίας έχει τις βάσεις της στη θεωρία πιθανοτήτων και τη Στατιστική, ενώ πολλά από τα μαθηματικά που εφαρμόζονται αναπτύχθηκαν πρώτα στο πεδίο της Θερμοδυναμικής από τους Boltzmann, Gibbs και Landauer.

Ο Shannon ανέπτυξε το μέτρο για την Πληροφορία με βάση την παρατήρηση ενός γεγονότος με πιθανότητα p , έτσι η Πληροφορία εκφράζεται με όρους πιθανότητας (p). Το μέτρο της Πληροφορίας $I(p)$ ορίζεται σύμφωνα με τα παρακάτω αξιώματα:

α. Η Πληροφορία είναι ένας μη αρνητικός πραγματικός αριθμός:

$$I(p) \geq 0$$

β. Η εμφάνιση ενός γεγονότος με πιθανότητα 1, δεν παράγει καμιά πληροφορία:

$$I(1) = 0$$

γ. Η Πληροφορία που αποκτήθηκε από δύο ανεξάρτητα γεγονότα είναι το σύνολο της πληροφορίας που αποκτήθηκε για καθένα από τα γεγονότα ξεχωριστά:

$$I(p_1 \cdot p_2) = I(p_1) + I(p_2)$$

δ. Μικρές αλλαγές στην πιθανότητα, δεν πρέπει να έχουν ως αποτέλεσμα μεγάλες αλλαγές στην $I(p)$. Αυτό σημαίνει ότι η $I(p)$ πρέπει να είναι συνεχής και μονότονη συνάρτηση της p .

Από τα παραπάνω αξιώματα προκύπτουν οι εξής κανόνες:

$$\text{i. } I(p^n) = n \cdot I(p)$$

$$\text{ii. } I(p) = I\left(\left(p^{\frac{1}{n}}\right)^n\right) = n \cdot I\left(p^{\frac{1}{n}}\right)$$

$$\text{iii. } I\left(p^{\frac{n}{m}}\right) = \frac{n}{m} \cdot I(p)$$

Λόγω της ιδιότητας της συνέχειας της συνάρτησης, και σύμφωνα με τον παραπάνω κανόνα ισχύει ότι $0 < p \leq 1$ και για έναν πραγματικό αριθμό $r > 0$ έχουμε:

$$I(p^r) = r \cdot I(p)$$

Από τα παραπάνω αξιώματα και τον κανόνα iii μπορούμε να συμπεράνουμε ότι:

$$I(p) = -\log_b(p) = \log_b\left(\frac{1}{p}\right), \quad \text{για } b \text{ θετικό.}$$

Οι μονάδες που χρησιμοποιούνται για τη μέτρηση της Πληροφορίας καθορίζονται από τη βάση του λογαρίθμου b . Έτσι όταν χρησιμοποιείται ο φυσικός λογάριθμος e η μονάδα είναι το nat (από το natural unit) και χρησιμοποιείται συνήθως για συνεχείς τυχαίες μεταβλητές. Όταν χρησιμοποιείται ο δεκαδικός λογάριθμος, τότε η μονάδα είναι το Hartley. Η επικρατέστερη μονάδα, με βάση λογαρίθμου το 2 είναι το bit (από το binary digit), λόγω της εκτεταμένης χρήσης του δυαδικού συστήματος στους υπολογιστές. Ο πίνακας που ακολουθεί δείχνει τις διάφορες μονάδες μέτρησης της Πληροφορίας.

b	Όνομα μονάδας
2	Bits
3	trits
e	Nats
10	Hartleys

Πίνακας 1: Μονάδες μέτρησης της Πληροφορίας

Αξίζει να σημειωθεί ότι η Πληροφορία και η Εντροπία είναι συναρτήσεις της κατανομής μιας τυχαίας μεταβλητής X οι οποίες εξαρτώνται μόνο από τις πιθανότητες που χαρακτηρίζουν την τυχαία μεταβλητή και όχι από τις πιθανές τιμές της τυχαίας μεταβλητής. Είναι ακόμη εύκολο να δειχθεί ότι $H(X) \geq 0$ για κάθε πιθανή κατανομή της πιθανότητας του X .

Η Πληροφορία εναλλακτικά μπορεί να μετρηθεί και ως μείωση της αβεβαιότητας. Η Πληροφορία, δηλαδή, μπορεί να θεωρηθεί ως η ικανότητα μείωσης της αβεβαιότητας. Πρακτικά, η ποσότητα της πληροφορίας προκύπτει από τη διαφορά της *εκ των προτέρων* (a priori) αβεβαιότητας και της *εκ των υστέρων* (posteriori) αβεβαιότητας. Η εκ των προτέρων αβεβαιότητα μειώνεται μέσω της απόκτησης Πληροφορίας από την εκτέλεση πειραμάτων και αναζήτησης δεδομένων και στοιχείων, ενώ η εκ των υστέρων αβεβαιότητα υπολογίζεται μετά την παρατήρηση των δεδομένων (Klir, 2005).

3.2 Εντροπία Shannon

Η Εντροπία είναι θεμελιώδης έννοια για τη Θεωρία της Πληροφορίας. Η έννοια της Εντροπίας μπορεί να ερμηνευθεί με τους εξής δύο τρόπους:

- α. Η Εντροπία για μια τυχαία μεταβλητή αναφέρεται στην ποσότητα της αβεβαιότητας αυτής.
- β. Δείχνει το μέσο περιεχόμενο Πληροφορίας ή την αναμενόμενη τιμή της Πληροφορίας που απαιτείται για να χαρακτηριστεί μια τυχαία μεταβλητή (μήνυμα).

Η Εντροπία μπορεί να μετρηθεί μηχανικά ή στατιστικά. Μηχανικά μετράται με όρους θερμότητας ή ενεργειακών ροών, ενώ στατιστικά μετράται με εκτιμήσεις πιθανοτήτων που περιλαμβάνουν τον αριθμό των μικροκαταστάσεων στις οποίες μπορεί να βρεθεί το σύστημα. Τα Βιολογικά Συστήματα εκδηλώνουν τις λειτουργίες τους και μεταβάλλονται με πολλούς και ποικίλους τρόπους και ως εκ τούτου η Εντροπία μηχανικά είναι πολύ δύσκολο να μετρηθεί, ωστόσο είναι εφικτό να μετρηθεί στατιστικά. Η Θεωρία της Πληροφορίας επιτρέπει τη στατιστική μέτρηση της Εντροπίας. Στατιστικά μπορούμε να πούμε ότι η Εντροπία παράγεται όσο νέες μικροκαταστάσεις γίνονται διαθέσιμες σε ένα σύστημα (Brooks, Leblond, & Cumming, 1984).

Η Εντροπία μπορεί να οριστεί μαθηματικά ως εξής: Αν X μια τυχαία μεταβλητή που παίρνει τις τιμές της από το \mathbf{X} με συνάρτηση μάζας πιθανότητας $p(x) = Pr\{X=x\}$, $x \in \mathbf{X}$, η Εντροπία του X ορίζεται ως:

$$H(X) = -\sum_{x \in X} p(x) \log p(x) \quad (1)$$

Η Εντροπία του Shannon μπορεί να χρησιμοποιηθεί και για τον ορισμό άλλων μετρήσεων της Πληροφορίας, οι οποίες αναδεικνύουν τις σχέσεις μεταξύ δύο τυχαίων μεταβλητών X και Y . Οι πιο βασικές σχετικές μετρήσεις που θα αναλυθούν στη συνέχεια είναι οι ακόλουθες:

- i. Η Σύνθετη Εντροπία, η οποία μετράει τη συνολική Πληροφορία των X και Y .
- ii. Η Υπό Συνθήκη Εντροπία, η οποία μετράει την Πληροφορία της X , όταν η Y είναι γνωστή και αντιστρόφως.
- iii. Η Σχετική Εντροπία, η οποία μετράει την ομοιότητα των X και Y .
- iv. Η Αμοιβαία Εντροπία, η οποία μετράει τη σχέση των X και Y , υπό την έννοια ότι δείχνει πόσο μειώνεται η Πληροφορία της X όταν μαθαίνουμε την Y και αντιστρόφως.

3.3 Σύνθετη Εντροπία

Ο ορισμός της Εντροπίας μπορεί να επεκταθεί για ένα ζεύγος τυχαίων μεταβλητών X και Y αντίστοιχα αν αντικατασταθούν στην παραπάνω σχέση. Η Σύνθετη ή Από-Κοινού Εντροπία (Joint Entropy) μετράει την συνολική πληροφορία των X και Y . Η Σύνθετη Εντροπία διακριτών τυχαίων μεταβλητών με κοινή πιθανότητα κατανομής $p(x,y)$ μπορεί να οριστεί ως:

$$H(X, Y) = -\sum_{x \in X} \sum_{y \in Y} p(x, y) \log p(x, y) \quad (2)$$

3.4 Υπό Συνθήκη Εντροπία

Η Υπό Συνθήκη ή Δεσμευμένη Εντροπία (Conditional Entropy) μετράει την Πληροφορία της X όταν η Y είναι γνωστή και αντίστροφα. Η Υπό Συνθήκη Εντροπία μιας τυχαίας μεταβλητής Y δεδομένης της τυχαίας μεταβλητής X ορίζεται ως εξής:

$$H(Y|X) = \sum_{x \in X} p(x) H(Y|X = x)$$

$$\begin{aligned}
&= - \sum_{x \in X} p(x) \sum_{y \in Y} p(y|x) \log p(y|x) \\
&= - \sum_{x \in X} \sum_{y \in Y} p(x, y) \log p(y|x) \tag{3}
\end{aligned}$$

Μπορεί να αποδειχθεί ότι $H(Y|X) \leq H(Y)$ το οποίο σημαίνει ότι γνωρίζοντας την τιμή της X δεν θα πρέπει να αυξάνεται η αβεβαιότητα για την Y .

Ακόμη εφαρμόζοντας τον Κανόνα της Αλυσίδας η Σύνθετη Εντροπία δύο τυχαίων μεταβλητών X, Y υπολογίζεται προσθέτοντας την Εντροπία της μίας στην Υπό Συνθήκη Εντροπία της άλλης ως εξής:

$$\begin{aligned}
H(X, Y) &= H(X) + H(Y|X) \\
&= H(Y) + H(X|Y)
\end{aligned}$$

3.5 Σχετική Εντροπία

Η Σχετική Εντροπία (Relative Entropy), γνωστή και ως *Kullback-Leibler distance*, μετράει την απόσταση μεταξύ δυο κατανομών πιθανότητας. Ορίζεται ως εξής:

$$D(p||q) = \sum_{x \in X} p(x) \log \frac{p(x)}{q(x)} \tag{4}$$

όπου p και q είναι κατανομές πιθανότητας στον χώρο του X . Η Σχετική Εντροπία είναι πάντα μη αρνητικός αριθμός και είναι μηδέν αν $p=q$.

3.6 Αμοιβαία Πληροφορία

Η Αμοιβαία Πληροφορία (Mutual Information) μετράει τη σχέση των τυχαίων μεταβλητών X και Y , όταν λάβουμε κάποια πληροφορία για την πρώτη ή για τη δεύτερη αντίστοιχα. Δηλαδή, η απόκτηση πληροφορίας για την X μπορεί να συνεπάγεται μείωση της ποσότητας της αβεβαιότητας για την Y και αντιστρόφως.

Έχουμε ότι $p(x,y)$ η κοινή πιθανότητα κατανομής των τυχαίων μεταβλητών X, Y και $p(x), p(y)$ είναι περιθώριες κατανομές πιθανότητας (marginal probability distributions) αντίστοιχα. Τότε η Αμοιβαία Πληροφορία της X γνωρίζοντας την Y ή $I(X;Y)$ είναι η Σχετική Εντροπία ανάμεσα στη κοινή πιθανότητα κατανομής και την πιθανότητα κατανομής του X και Y :

$$\begin{aligned} I(X;Y) &= \sum_{x \in X} \sum_{y \in Y} p(x,y) \log \frac{p(x,y)}{p(x)p(y)} \\ &= D(p(x,y) \| p(x)p(y)) \end{aligned} \quad (5)$$

Η Αμοιβαία Πληροφορία μπορεί να εκφραστεί με όρους Εντροπίας ως εξής:

$$\begin{aligned} I(X;Y) &= H(X) - H(X|Y) \\ &= H(Y) - H(Y|X), \text{ (σύμφωνα με τον Κανόνα της Αλυσίδας)} \\ &= H(X) + H(Y) - H(X,Y) \\ &= I(Y;X) \end{aligned} \quad (5.1)$$

Η Αμοιβαία Πληροφορία μπορεί να θεωρηθεί ως μια γενικευμένη συσχέτιση ανάλογη του συντελεστή γραμμικής συσχέτισης r , που μπορεί όμως να εφαρμοστεί και σε μη γραμμικά συστήματα (Roufston, 1997).

Μία γενικευμένη εκδοχή της Αμοιβαίας Πληροφορία για τυχαίες μεταβλητές με αιτιώδεις σχέσεις είναι η Κατευθυνόμενη Πληροφορία (directed information) (Mousavian et al., 2016). Η μετρική αυτή περιέχει επιπλέον το στοιχείο της κατεύθυνσης της ροής της πληροφορίας σε κανάλια με αιτιατή ανάδραση και ορίζεται μέσω των αιτιατών υπό συνθήκη πιθανοτικών κατανομών. Η Κατευθυνόμενη Εντροπία για δύο τυχαίες ακολουθίες $X^N = (X_1, \dots, X_N)$ και $Y^N = (Y_1, \dots, Y_N)$ ορίζεται ως:

$$I(X^N \rightarrow Y^N) = \sum_{n=1}^N I(X^n; Y^n | Y^{n-1}) \quad (6)$$

όπου $I(X^N; Y^N)$ είναι η Αμοιβαία Πληροφορία

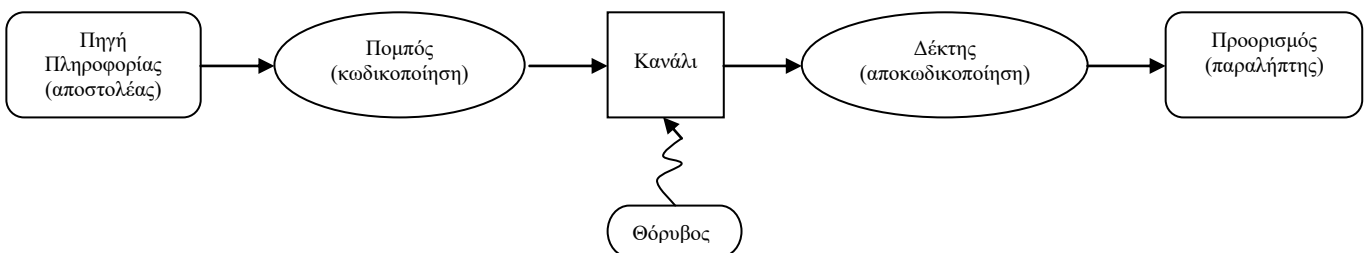
3.7 Διακριτό Κανάλι και Χωρητικότητα Καναλιού

Η διαδικασία της μετάδοσης της πληροφορίας απαιτεί ένα φυσικό σύστημα-φορέα που να συνδέει χωρικά και χρονικά την πηγή της πληροφορίας με τον αποδέκτη. Στην πηγή, το μήνυμα κωδικοποιείται και αποστέλλεται μέσω ενός ενθόρυβου καναλιού στον προορισμό όπου αποκωδικοποιείται (Σχήμα 1). Ένα κανάλι επικοινωνίας καλείται κανάλι "χωρίς μνήμη" (memoryless), εάν η κατανομή πιθανότητας της εξόδου δεν εξαρτάται εξαρχής από τις προηγούμενες εισόδους ή εξόδους του καναλιού και εξαρτάται μόνο από την είσοδο του καναλιού στην τρέχουσα χρονική στιγμή. Η έννοια του θορύβου εμπεριέχει όλους τους ανεπιθύμητους παράγοντες που μπορεί να παρεμβάλλονται στη διαδικασία της επικοινωνίας, με συνέπεια απώλειες στην ακεραιότητα του μηνύματος. Με βάση αυτό το μοντέλο επικοινωνίας προκύπτουν οι αναλογίες στα Βιολογικά Συστήματα.

Χωρητικότητα Καναλιού (channel capacity) ονομάζεται η μέγιστη τιμή του ρυθμού μετάδοσης της πληροφορίας (μονάδες πληροφορίας /μονάδες χρόνου) μέσα από το κανάλι επικοινωνίας. Η κατάλληλη μέτρηση στην περίπτωση αυτή είναι η Αμοιβαία Πληροφορία και συγκεκριμένα η μέγιστη Αμοιβαία Πληροφορία (Mousavian et al., 2016). Τότε, η Χωρητικότητα Καναλιού C σε ένα διακριτό, χωρίς μνήμη κανάλι επικοινωνίας ορίζεται ως εξής:

$$C = \max_{p(x)}(I(X; Y)) \quad (7)$$

όπου $p(x)$ η πιθανότητα κατανομής της τυχαίας μεταβλητής εισόδου x και η μέγιστη τιμή για το I λαμβάνεται από όλες τις πιθανές κατανομές $p(x)$



Σχήμα 1: Σχηματική απεικόνιση του γενικού μοντέλου επικοινωνίας (Shannon 1948).

3.8 Θεωρία Ρυθμού-Παραμόρφωσης

Η Θεωρία Ρυθμού-Παραμόρφωσης παρέχει τις θεωρητικές βάσεις για τη συμπίεση δεδομένων με απώλειες. Η Θεωρία Ρυθμού-Παραμόρφωσης περιγράφει το πρόβλημα της εύρεσης του ελάχιστου αριθμού bits/έξοδο που θα πρέπει να μεταδίδονται, ρυθμός R , μέσω του καναλιού έτσι ώστε να ανακατασκευαστεί η πηγή (σήμα εισόδου) στον δέκτη, χωρίς να υπερβαίνει έναν προκαθορισμένο ρυθμό σφαλμάτων (παραμόρφωση). Η απόσταση μεταξύ της τυχαίας μεταβλητής X και της εικόνας της \hat{X} ονομάζεται παραμόρφωση (distortion) $d(X, \hat{X})$. Η στατιστική μέση τιμή της παραμόρφωσης είναι:

$$D = E[d(X, \hat{X})]$$

Η σχέση μεταξύ R και D είναι η συνάρτηση του Ρυθμού-Παραμόρφωσης $R(D)$ και μπορεί να οριστεί ως ένα πρόβλημα ελαχιστοποίησης ως εξής:

$$R(D) = \min_{E[d(X, \hat{X})] \leq D} I(X; \hat{X}) \quad (8)$$

3.9 Εντροπία Kolmogorov-Sinai και Εντροπία Δικτύου

Σε δίκτυα δυναμικών συστημάτων, τα πιο σημαντικά μέτρα που ποσοτικοποιούν την περιεχόμενο της πληροφορίας είναι η Εντροπία Δικτύου και η Εντροπία Kolmogorov-Sinai (KS) (Mesbahi et al., 2015; Kazienko et al., 2017). Η Εντροπία ως μέτρο της τυχαιότητας για τα δυναμικά συστήματα εισήχθη από τους Kolmogorov και Sinai. Η Εντροπία Kolmogorov-Sinai αποτελεί μια γενίκευση της Εντροπίας του Shannon η οποία χρησιμοποιείται για την ποσοτικοποίηση της πολυπλοκότητας των δυναμικών συστημάτων. Όσο πιο πολύ προβλέψιμη είναι η αβεβαιότητα τόσο υψηλότερη είναι η Εντροπία του. Ενωιολογικά σχετίζεται με τη Εντροπία του Shannon, όπου η εκ των προτέρων αβεβαιότητα μειώνεται μέσω της απόκτησης πληροφορίας (Mesbahi et al., 2015). Ακόμη, σχετίζεται με την έννοια της Εντροπίας στη θερμοδυναμική όπου η αταξία συνεπάγεται αύξηση της Εντροπίας και το γεγονός ότι η αταξία και η αβεβαιότητα (ή μη-προβλεψιμότητα) είναι στενά συνδεδεμένες. Πολύ συνοπτικά, η Εντροπία Kolmogorov-Sinai H μπορεί να οριστεί επιλέγοντας μια διαμέριση $\{W_i\}$ η οποία να μεγιστοποιεί το ρυθμό "παραγωγής πληροφορίας" $H\{W_i\}$ (Sinai, 2009; Stumpf, Balding, & Girolami, 2011). Η Εντροπία Kolmogorov-Sinai για ένα δυναμικό σύστημα περιγράφεται από τη σχέση:

$$H = \sup_{\{W_i\}} H(\{W_i\}) \quad (9)$$

όπου το άνω φράγμα (sup) βρίσκεται για όλες τις πεπερασμένες διαμερίσεις $\{W_i\}$

Η έννοια της Εντροπίας Δικτύου σχετίζεται με την ικανότητα του δικτύου να αντεπεξέρχεται σε τυχαίες αλλαγές στη δομή του (Demetrius & Manke, 2005). Η Εντροπία Δικτύου είναι η Εντροπία ενός στοχαστικού πίνακα Markov $P = (p_{ij})$ όπου $p_{ij} \geq 0$, $\sum_j p_{ij} = 1$ και π η στάσιμη κατανομή, που σχετίζεται με έναν πίνακα γειτνίασης (adjacency matrix) $A = (a_{ij})$. Τότε ο συγκεκριμένος πίνακας που καθορίζεται στα πλαίσια μιας μεταβολικής διατύπωσης, όπου λ η κυρίαρχη ιδιοτιμή του πίνακα A με (v_i) το κυρίαρχο ιδιοδιάνυσμα και M_A το σύνολο των στοχαστικών πινάκων που ικανοποιούν την ιδιότητα $a_{ij} = 0 \leftrightarrow p_{ij} = 0$. Η παράμετρος λ θα πρέπει να ικανοποιεί τη μεταβολική διατύπωση ως εξής:

$$\log \lambda = \sup_p \left[- \sum_{i,j} \pi_i p_{ij} \log p_{ij} + \sum_{i,j} \pi_i p_{ij} \log a_{ij} \right] \quad (10)$$

όπου το άνω φράγμα για όλους τους πιθανούς στοχαστικούς πίνακες λαμβάνεται για τον μοναδικό στοχαστικό πίνακα $P=(p_{ij})$ που ορίζεται ως:

$$p_{ij} = \frac{a_{ij} v_j}{\lambda v_i}$$

Τότε, η Εντροπία Δικτύου ορίζεται ως:

$$H = - \sum_{ij} \pi_i p_{ij} \log p_{ij} \quad (11)$$

3.10 Αρχή Μέγιστης Εντροπίας

Η Αρχή Μέγιστης Εντροπίας διατυπώθηκε αρχικά από τον E. T. Jaynes το 1957 ο οποίος παρατήρησε μια φυσική αντιστοιχία μεταξύ της Στατιστικής Μηχανικής και της Θεωρίας της Πληροφορίας. Ισχυρίστηκε ότι η Εντροπία της Στατιστικής Μηχανικής και η Εντροπία του Shannon είναι ουσιαστικά το ίδιο. Συνεπώς, η Στατιστική Μηχανική θα πρέπει να θεωρηθεί ως μια συγκεκριμένη εφαρμογή ενός γενικού εργαλείου λογικής συμπερασματολογίας και Θεωρίας της Πληροφορίας. Συνοπτικά η Αρχή της Μέγιστης Εντροπίας είναι μία μέθοδος για την εξαγωγή συμπερασμάτων υπό συνθήκες αβεβαιότητας ή ελλιπούς γνώσης και στοχεύει στην εύρεση της κατάλληλης κατανομής πιθανοτήτων σύμφωνα με την διαθέσιμη πληροφορία.

Σύμφωνα με την αρχή αυτή, η καλύτερη κατανομή είναι αυτή που ικανοποιεί όλους τους γνωστούς περιορισμούς και ταυτόχρονα ακολουθεί την πιθανότητα κατανομής της

Μέγιστης Εντροπίας, δηλαδή την πιθανότητα κατανομής που μεγιστοποιεί την Εντροπία του Shannon (εξ. 1). Σύμφωνα με τον Jaynes, (όσον αφορά στο πρόβλημα της πρόβλεψης) "η μεγιστοποίηση της Εντροπίας δεν είναι μια εφαρμογή ενός νόμου της φυσικής, αλλά απλώς μια μέθοδος συλλογισμού που εξασφαλίζει ότι δεν έχουν εισαχθεί αυθαίρετες υποθέσεις". Είναι δηλαδή η λιγότερο μεροληπτική εκτίμηση για τις δεδομένες πληροφορίες και είναι απόλυτα μη δεσμευτική όσον αφορά τις πληροφορίες που λείπουν. Ακόμη, η πιθανότητα που μεγιστοποιεί την Εντροπία είναι αυτή με τη μικρότερη Πληροφορία. Για περισσότερες πληροφορίες σχετικά με τις μεθόδους μεγιστοποίησης της Εντροπίας για διακριτές και συνεχείς κατανομές πιθανότητας, μπορεί κανείς να δει την αντίστοιχη βιβλιογραφία (Jaynes, 1957).

Κεφάλαιο 3. Εφαρμογές της Θεωρίας της Πληροφορίας

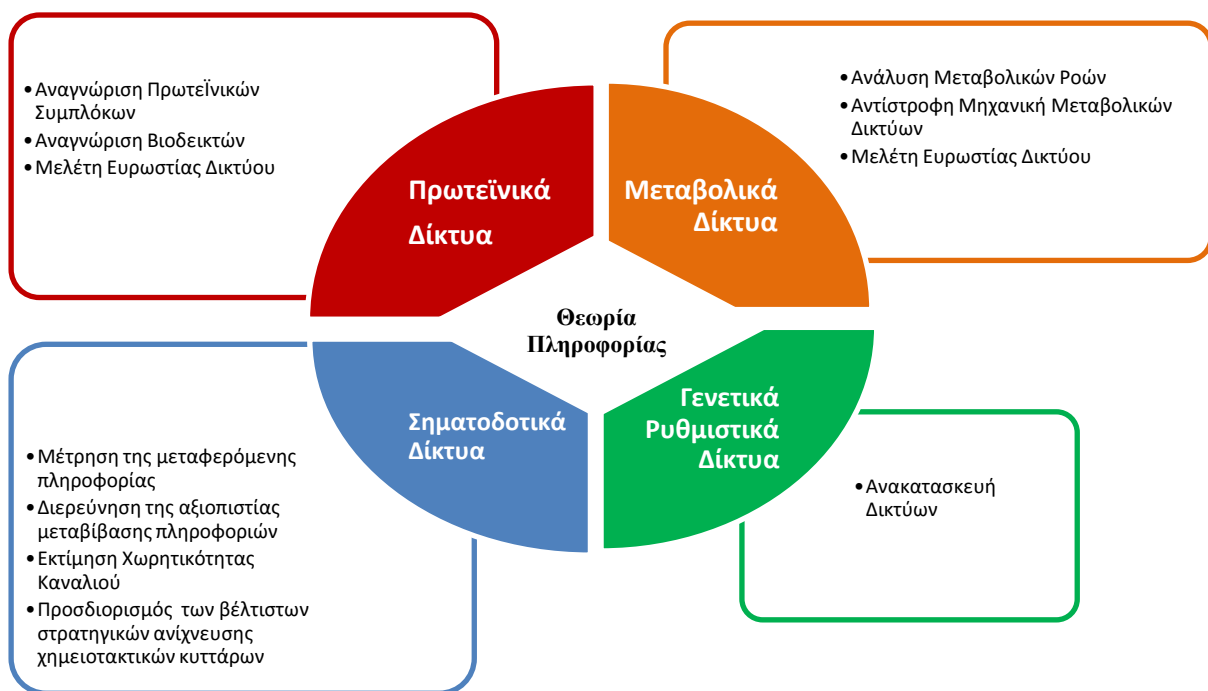
Η Βιολογία Συστημάτων εστιάζει στη μελέτη της δυναμικής των συσχετιζόμενων στοιχείων που απαρτίζουν ένα βιολογικό σύστημα κατά τη διάρκεια της λειτουργίας του. Η χρήση της έννοιας του δικτύου εξυπηρετεί πολύ καλά τη βιολογική πραγματικότητα. Κάθε κόμβος αντιπροσωπεύει μια βιολογική οντότητα (γονίδιο, πρωτεΐνη, κλπ.) και κάθε ακμή τις αλληλεπιδράσεις μεταξύ τους. Στο κεφάλαιο αυτό μέθοδοι της Θεωρίας της Πληροφορίας χρησιμοποιούνται στην διερεύνηση της δυναμικής και της δομής των βιολογικών δικτύων, από την εύρεση του βαθμού συσχέτισης δυο βιολογικών οντοτήτων μέχρι την ανακατασκευή ενός βιολογικού δικτύου.

Η βιολογική πληροφορία που αποθηκεύεται στα γονίδια, μπορεί να μελετηθεί με ποικίλους τρόπους. Σύμφωνα με τη Θεωρία της Πληροφορίας, οι βασικές πτυχές της επικοινωνίας είναι το μήνυμα που κωδικοποιείται από ένα σύνολο συμβόλων, ο πομπός, το μέσο στο οποίο μεταδίδονται οι πληροφορίες, ο δέκτης και ο θόρυβος³ (Σχήμα 1). Αντιστοίχως, στη βιολογική έκφραση της γενετικής πληροφορίας ο πομπός είναι το γονιδίωμα, το μήνυμα είναι το RNA messenger και ο δέκτης είναι το κυτταρόπλασμα του κυττάρου με τα συστατικά που περιέχει. Με ανάλογο τρόπο η θεωρία μπορεί να χρησιμοποιηθεί σε μεγαλύτερες κλίμακες όπως για παράδειγμα στην περιγραφή σημάτων στο νευρικό σύστημα των ζώων. Στην περίπτωση αυτή ένα όργανο αίσθησης μπορεί να συλλέξει πληροφορίες από τον οργανισμό και το περιβάλλον του μεταφέροντας τις

³ Ως θόρυβος στο σημείο αυτό, εννοείται ο μοριακός θόρυβος ο οποίος αναφέρεται στις εγγενείς στοχαστικές διακυμάνσεις των βιοχημικών διεργασιών του κυττάρου λόγω της διακριτικότητας των μορίων και της τυχαίας σύγκρουσής τους (Leng & Sharrock, 2012). Ο θόρυβος επηρεάζει όλες τις βιολογικές διεργασίες και υφίσταται σε όλες τις βιολογικές κλίμακες από το μοριακό και κυτταρικό επίπεδο έως και τη δυναμική των οργανισμών και των πληθυσμών. Οι λειτουργικοί ρόλοι του θορύβου στα διάφορα επίπεδα οργάνωσης μπορεί να διαφέρουν σημαντικά (Tsimring, 2015).

πληροφορίες αυτές στον εγκέφαλο όπου αποθηκεύονται και επεξεργάζονται (Haynie, 2008).

Στο κεφάλαιο αυτό θα παρουσιαστούν οι πιο σημαντικές τεχνικές εφαρμογές της Θεωρίας της Πληροφορίας σε βιολογικά συστήματα: Συγκεκριμένα, θα μελετηθούν αποκλειστικά δίκτυα συστημάτων της Μοριακής Βιολογίας, των οποίων τα στοιχεία προς μελέτη είναι μόρια (πρωτεΐνες, γονίδια και μεταβολίτες). Με βάση τον τύπο των στοιχείων του συστήματος και τις αλληλεπιδράσεις μεταξύ τους, τα Βιολογικά Συστήματα μπορούν να κατηγοριοποιηθούν σε τέσσερις μεγάλες κατηγορίες Γενετικά Ρυθμιστικά Δίκτυα, Μεταβολικά Δίκτυα, Πρωτεϊνικά και Σηματοδοτικά Δίκτυα. Το παρακάτω σχήμα (Σχήμα 2) δείχνει συνοπτικά τις εφαρμογές της Θεωρίας της Πληροφορίας που θα μελετηθούν στη συνέχεια του κεφαλαίου.



Σχήμα 2: Εφαρμογές της Θεωρίας της Πληροφορίας σε Βιολογικά Συστήματα

4.1 Γενετικά Ρυθμιστικά Δίκτυα

4.1.1 Εισαγωγή

Τα Γενετικά Ρυθμιστικά Δίκτυα αναπαριστούν σύνολα γονιδίων και τις μεταξύ τους αλληλεπιδράσεις που αφορούν στις ρυθμιστικές σχέσεις μεταξύ των μεταγραφικών παραγόντων και των γονιδίων που στοχεύουν. Μοντελοποιούνται συνήθως ως δίκτυα τα οποία αποτελούνται από κόμβους και ακμές. Οι κόμβοι αντιπροσωπεύουν γονίδια ή τις κωδικοποιούμενες από αυτά πρωτεΐνες ή μεταβολίτες, ενώ οι ακμές αντιπροσωπεύουν τις μοριακές αλληλεπιδράσεις μεταξύ πρωτεϊνών ή μεταξύ DNA και πρωτεϊνών ή άλλες έμμεσες σχέσεις μεταξύ γονιδίων (Karlach & Shamir, 2008).

Τα Γενετικά Ρυθμιστικά Δίκτυα είναι ίσως το πιο σημαντικό είδος βιολογικού δικτύου καθώς διαδραματίζουν εξέχοντα ρόλο σε κάθε έκφανση της λειτουργίας του κυττάρου. Αποτελούν "κέντρα ελέγχου" της κυτταρικής διαφοροποίησης, του μεταβολισμού, του κυτταρικού κύκλου και της μεταγωγής σήματος. Η μοντελοποίηση των δικτύων αυτών αποτελεί μεγάλη πρόκληση για τη Βιολογία Συστημάτων καθώς μπορεί να συντελέσει στη καλύτερη κατανόηση των πτυχών της φυσιολογίας, της ανάπτυξης και της παθογένειας των κυττάρων και της σχέσης φαινότυπου-γονότυπου. Έχει ακόμη εξέχουσα σημασία για την πρόοδο των θεραπευτικών μεθόδων και την επιτάχυνση και βελτιστοποίηση των βιοτεχνολογικών επιτευγμάτων (Costanzo et al., 2019; Karlach & Shamir, 2008; Mousavian et al., 2016).

Βασικός στόχος της μελέτης των Γενετικών Ρυθμιστικών Δικτύων είναι η *ανακατασκευή* τους, δηλαδή η εκ νέου σύνθεση ή αναδόμησή τους από πειραματικά δεδομένα. Απαραίτητη προϋπόθεση για την ανακατασκευή των δικτύων αυτών είναι η πλήρης κατανόηση της ροής της αιτιότητας. Μια τέτοια κατανόηση χρειάζεται έναν διεπιστημονικό συνδυασμό που περιλαμβάνει θεωρία, πειράματα, τεχνολογίες υψηλής απόδοσης (high-throughput) και μοντελοποίηση. Με την πρόοδο των τεχνολογιών υψηλής απόδοσης μπορούμε να λάβουμε μετρήσεις για την συνολική απόκριση ενός βιολογικού συστήματος από δεδομένα έκφρασης σε πλήθος περιπτώσεων. Η ανακατασκευή των Γενετικών Ρυθμιστικών Δικτύων γίνεται με μεθόδους Αντίστροφης Μηχανικής (Reverse Engineering) και αποτελεί πολύτιμο εργαλείο για τη διερεύνηση των σχέσεων αιτίου-αποτελέσματος μεταξύ γονιδίων, βάσει των παρατηρούμενων επιπέδων έκφρασης (Davidson & Levin, 2005).

Τα δεδομένα για την ανάλυση της γονιδιακής έκφρασης που χρησιμοποιούν οι αλγόριθμοι που θα εξεταστούν στη συνέχεια, προέρχονται από την τεχνολογία των μικροσυστοιχιών. Συνοπτικά, οι μικροσυστοιχίες γονιδίων αποτελούν δισδιάστατες διατάξεις σημείων στα οποία είναι ακινητοποιημένες αλληλουχίες χιλιάδων διαφορετικών

γονιδίων. Η διαδικασία βασίζεται στην ανίχνευση μέσω υβριδισμού, μια τεχνική που χρησιμοποιεί μόρια νουκλεϊκού οξέος σεσημασμένα με φθορίζουσα χρωστική προκειμένου να ταυτοποιήσει συμπληρωματικά μόρια. Οι μικροσυστοιχίες έχουν το πλεονέκτημα ότι περιέχουν έναν πολύ μεγάλο αριθμό γονιδίων σε πολύ μικρό μέγεθος. Μπορούν να χρησιμοποιηθούν προκειμένου να αναλύσουν τη γονιδιακή έκφραση από εκατοντάδες ή και χιλιάδες γονίδια με ένα μόνο δείγμα. Οι μικροσυστοιχίες ανιχνεύουν ποιά γονίδια ενεργοποιούνται ή καταστέλλονται σε διάφορα στάδια ανάπτυξης ενός οργανισμού, ή σε απόκριση σε στα διάφορα ερεθίσματα του περιβάλλοντος (Davis, et al., 1995).

Στην ενότητα αυτή θα εξεταστεί το πώς οι μετρικές της Θεωρίας της Πληροφορίας συνεισφέρουν στην ανίχνευση πολύπλοκων ρυθμιστικών σχέσεων από δεδομένα γονιδιακής έκφρασης. Τα μέτρα της Εντροπίας και της Αμοιβαίας πληροφορίας ελέγχουν τα επίπεδα έκφρασης τα οποία καθιστούν δύο γονίδια αλληλοεξαρτώμενα. Οι μετρικές αυτές έχουν εφαρμοστεί κυρίως στα Γενετικά Ρυθμιστικά Δίκτυα για τον εντοπισμό μεγάλων Γενετικών Ρυθμιστικών Δικτύων από δεδομένα χαμηλής έκφρασης (Mousavian, Kanousi, et al., 2016). Στη συνέχεια της ενότητας θα αναλυθούν οι αλγόριθμοι ανακατασκευής Γενετικών Ρυθμιστικών Δικτύων που εμπεριέχουν τις εν λόγω μετρικές. Μια μεγάλη υποκατηγορία των αλγόριθμων ανακατασκευής για τα Γενετικά Ρυθμιστικά Δίκτυα είναι τα *Δίκτυα Αμοιβαίας Πληροφορίας*, δίκτυα δηλαδή που βασίζονται στο μέτρο της Αμοιβαίας Πληροφορίας.

4.1.2 Πίνακας Αμοιβαίας Πληροφορίας

Το μέτρο της Αμοιβαία Πληροφορίας, όπως είδαμε και στο προηγούμενο κεφάλαιο ποσοτικοποιεί την στοχαστική σχέση ανάμεσα σε δύο τυχαίες διακριτές μεταβλητές. Η Αμοιβαία Πληροφορία μπορεί να συλλάβει τη μη-γραμμική εξάρτηση μεταξύ τυχαίων μεταβλητών, που σημαίνει ότι μπορεί να εφαρμοστεί στα Γενετικά Ρυθμιστικά Δίκτυα για τον εντοπισμό έμμεσων ρυθμιστικών σχέσεων μεταξύ δύο γονιδίων. Οι μέθοδοι που θα ακολουθήσουν απαιτούν αρχικά τον υπολογισμό του Πίνακα της Αμοιβαίας Πληροφορίας (MIM-Mutual Information Matrix). Πρόκειται για έναν τετραγωνικό πίνακα, που περιέχει δηλαδή τον ίδιο αριθμό σειρών και στηλών, όπου το στοιχείο του $MIM_{i,j}$ δίνεται από την Αμοιβαία Πληροφορία μεταξύ των X_i και X_j ως εξής:

$$\begin{aligned} MIM_{ij} &= H(X_i) + H(X_j) - H(X_i, X_j) \\ &= I(X_i, X_j) \end{aligned}$$

$$= \sum_{x_i \in X} \sum_{x_j \in X} p(x_i, x_j) \log\left(\frac{p(x_i, x_j)}{p(x_i)p(x_j)}\right), \quad (3.1)$$

όπου X_i και X_j με $X_i \in X$, $i = 1, \dots, n$, είναι τυχαίες διακριτές μεταβλητές που δηλώνουν τα επίπεδα έκφρασης των γονιδίων i και j αντίστοιχα (Olsen, Meyer, & Bontempi, 2009).

4.1.3 Ο αλγόριθμος REVEAL

Ο αλγόριθμος REVEAL (REVerse Engineering Algorithm) είναι ο πρώτος γενικός αλγόριθμος Αντίστροφης Μηχανικής βασισμένος στη Θεωρία της Πληροφορίας που δημιουργήθηκε για τη μελέτη των Γενετικών Ρυθμιστικών Δικτύων. Η απεικόνισή τους βασίζεται στο δυαδικό δίκτυο⁴ (Boolean network). Οι κόμβοι αντιστοιχούν σε ένα γονίδιο το οποίο μπορεί να βρίσκεται σε δύο διαφορετικές (ιδεατές) καταστάσεις, την ενεργοποιημένη και την απενεργοποιημένη οι οποίες παίρνουν αντίστοιχα τις τιμές 1 και 0. Ο αλγόριθμος λαμβάνει διακριτά δεδομένα, ωστόσο για να αναπαρασταθεί η συνεχής συμπεριφορά των έμβιων συστημάτων γίνεται χρήση πολλαπλών διακριτών δεδομένων (που προσεγγίζουν, δηλαδή, τα συνεχή) (Liang, Fuhrman, & Somogyi, 1998).

Ο αλγόριθμος χρησιμοποιεί το μέτρο της Αμοιβαίας Πληροφορίας για τον εντοπισμό λειτουργικών σχέσεων μεταξύ δύο γονιδίων. Η Αμοιβαία Πληροφορία υπολογίζει πρακτικά το βαθμό συσχέτισης δύο γονιδίων. Ο βαθμός αυτός, όπως μπορούμε να δούμε από την (3.1) όταν λάβει την τιμή 0 δηλώνει ότι οι τυχαίες μεταβλητές είναι ανεξάρτητες ενώ όταν δύο μεταβλητές συνδέονται επιστρέφει την τιμή της Εντροπίας της περιθώριας κατανομής. Θα πρέπει να σημειωθεί, ωστόσο, ότι ο αλγόριθμος λειτουργεί καλά όταν το δείγμα για κάθε γονίδιο είναι σχετικά μικρό (Sanguinetti & Huynh-Thu, 2019).

4.1.4 Δίκτυα Συνάφειας (Relevance Networks-RELNET)

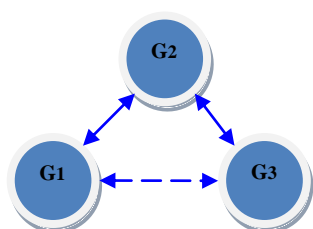
Τα Δίκτυα Συνάφειας αναπτύχθηκαν για πρώτη φορά στον τομέα της Λειτουργικής Γονιδιωματικής (Functional Genomics) για την ομαδοποίηση γονιδίων από δεδομένα έκφρασης RNA. Η τεχνική που χρησιμοποιείται υπολογίζει την Αμοιβαία Πληροφορία για

⁴ Τα Boolean networks (BNs) αποτελούν συστήματα διακριτού χρόνου, των οποίων οι μεταβλητές μπορούν να πάρουν μόνο δύο πιθανές τιμές (0,1). Εισήχθησαν για πρώτη φορά στο πεδίο της Βιολογίας για να περιγράψουν ποιοτικά τις γονιδιακές αλληλεπιδράσεις (Kauffman, 1969).

κάθε ζεύγος γονιδίων ενός συνόλου δεδομένων και ορίζει ένα όριο (κατώφλι). Η Αμοιβαία Πληροφορία είναι το κύριο εργαλείο προκειμένου να αποφανθούμε αν και πώς ένα γονίδιο επηρεάζει ένα άλλο. Αν βρεθεί υψηλή τιμή της Αμοιβαίας Πληροφορίας για ένα ζεύγος γονιδίων, τότε αυτό ερμηνεύεται ως πιθανή βιολογική συσχέτιση ανάμεσα στα δύο και το αντίθετο. Αυξάνοντας το όριο της Αμοιβαίας Πληροφορίας, το δίκτυο περιλαμβάνει όλο και πιο ισχυρές πιθανές συσχετίσεις, ενώ χαμηλώνοντάς το εισέρχονται ολοένα και πιο ασθενείς πιθανές συσχετίσεις (Butte & Kohane, 1999). Τα Δίκτυα Συνάφειας μας επιτρέπουν να βρούμε τα γονίδια που σχετίζονται βιολογικά, χωρίς να παρέχουν επιπλέον πληροφορία για το είδος των ρυθμιστικών σχέσεων. Στη συνέχεια του κεφαλαίου θα αναλυθούν οι πιο γνωστές μέθοδοι που έχουν προταθεί για το "φιλτράρισμα" των έμμεσων ή ψευδών συνδέσεων οι οποίες βασίζονται σε Δίκτυα Συνάφειας.

4.1.4.1 Ο αλγόριθμος ARACNE

Το ARACNE (Algorithm for the Reconstruction of Accurate Cellular Networks) είναι ένας αλγόριθμος ανακατασκευής γενετικών ρυθμιστικών δικτύων από δεδομένα γονιδιακής έκφρασης μικροσυστοιχιών (steady-state microarray data). Ο αλγόριθμος βασίζεται στην υπόθεση ότι το επίπεδο έκφρασης ενός γονιδίου μπορεί να θεωρηθεί ως μια τυχαία μεταβλητή και ότι οι αμοιβαίες σχέσεις μεταξύ γονιδίων μπορούν να ανιχνευτούν από τις στατιστικές εξαρτήσεις.



Σχήμα 3: Οι κόμβοι του δικτύου απεικονίζουν γονίδια και οι ακμές τις ρυθμιστικές σχέσεις μεταξύ τους

Αρχικά, ο αλγόριθμος υπολογίζει την Αμοιβαία Πληροφορία για κάθε ζεύγος γονιδίων, $I(G_1, G_2) \equiv I_{G_1, G_2}$. Στη συνέχεια γίνεται εκτίμηση των παραμέτρων της Αμοιβαίας Πληροφορίας με χρήση μεθόδων τυχαιοποίησης (π.χ ο εκτιμητής Gaussian Kernel) και τίθεται το όριο (κατώφλι) σημαντικότητας αυτής (I_0) ώστε να υπάρχει στατιστική ανεξαρτησία. Ο αλγόριθμος στο σημείο αυτό χρησιμοποιεί σαν φίλτρο μια πιο ειδική

μετρική, τη λεγόμενη Ανισότητα Επεξεργασίας Δεδομένων (Data Processing Inequality-DPI), μια ιδιότητα που προκύπτει από τον Κανόνα της Αλυσίδας για την Αμοιβαία πληροφορία (εξ. 5.1.). Για παράδειγμα, έχουμε το γονίδιο G_2 το οποίο ρυθμίζει τα G_1 και G_3 (Εικόνα 3.1). Το G_2 πρόκειται να έχει υψηλή τιμή Αμοιβαίας πληροφορίας με το G_1 και G_3 και αντίστοιχα υψηλές τιμές θα έχουν τα G_1 και G_3 . Αυτό σημαίνει ότι δεν είναι εφικτό να εντοπιστεί κάποια διαφοροποίηση ως προς τις ρυθμιστικές σχέσεις μεταξύ των τριών γονιδίων μόνο με τη χρήση της Αμοιβαίας Πληροφορίας. Ωστόσο, με την εφαρμογή της Ανισότητας Επεξεργασίας Δεδομένων αφαιρείται από το δίκτυο η σχέση που θα λάβει τη μικρότερη τιμή Αμοιβαίας Πληροφορίας, η οποία ισοδυναμεί με τη μικρότερη συσχέτιση. Σύμφωνα με τη Ανισότητα Επεξεργασίας Δεδομένων ισχύει η σχέση:

$$I(G_1, G_3) \leq \min [I(G_1, G_2), I(G_2, G_3)]$$

Η μέθοδος αυτή ελαχιστοποιεί τις έμμεσες ή πιο ασθενείς υποψήφιες συσχετίσεις και επιλέγει να κρατήσει μόνο αυτές που αφορούν τα άμεσα συσχετιζόμενα ζεύγη γονιδίων. Τα δεδομένα εξόδου της ανάλυσης του ARACNE είναι ένα δίκτυο που προβλέπει τις ρυθμιστικές σχέσεις μεταξύ γονιδίων (A.A. Margolin et al., 2006; Basso et al., 2005; Mousavian, Kavousi, et al., 2016)

Το πλεονέκτημα του ARACNE σε σχέση με άλλους αλγορίθμους είναι ότι έχει χαμηλή υπολογιστική πολυπλοκότητα και χρησιμοποιεί το πλήρες εύρος δυναμικού των δεδομένων χωρίς να κάνει αυθαίρετες υποθέσεις σχετικά με τον τοπολογία του δικτύου. Σημαντικό επίσης χαρακτηριστικό για την αξιοπιστία του αλγορίθμου είναι ότι οι προβλέψεις του είναι βιοχημικά επικυρωμένες (Margolin et al., 2006).

Μία προέκταση του αλγορίθμου για συνεχή χρονικά δεδομένα (time-course data ή time series) είναι το TimeDelay-ARACNE. Η βασική ιδέα στην οποία στηρίζεται είναι ότι η έκφραση ενός γονιδίου μια συγκεκριμένη χρονική στιγμή μπορεί να εξαρτάται από τα επίπεδα έκφρασης ενός άλλου γονιδίου σε προηγούμενη χρονική στιγμή (ή χρονικές στιγμές). Ο αλγόριθμος TimeDelay-ARACNE ακολουθεί τα εξής τρία βήματα:

1. Εντοπίζει για κάθε γονίδιο την αρχική χρονική στιγμή των αλλαγών στην έκφρασή του.
2. Κατασκευάζει το δίκτυο.
3. Αφαιρεί τις λιγότερο σχετικές συνδέσεις από το δίκτυο.

Η χρονική προσθήκη στον αλγόριθμο μπορεί να δώσει μια πολύ σημαντική πληροφορία για το δίκτυο η οποία αφορά τον εντοπισμό της "κατεύθυνσης" μεταξύ των συνδέσεων. Αυτό πρακτικά μας επιτρέπει να διακρίνουμε τα ρυθμιστικά γονίδια από τα ρυθμιζόμενα. Ο αλγόριθμος χρησιμοποιείται με μεγάλη ακρίβεια στην ανακατασκευή

μικρών κατευθυνόμενων βιολογικών δικτύων, ανιχνεύοντας τις αλληλεπιδράσεις των γονιδίων ακόμη και όταν είναι διαθέσιμος μικρός αριθμός μετρήσεων (Zorropoli, Morganello, & Ceccarelli, 2010).

4.1.4.2 Ο αλγόριθμος CLR

Ο αλγόριθμος CLR (Context Likelihood of Relatedness) χρησιμοποιεί μεταγραφικά προφίλ μικροσυστοιχιών ενός οργανισμού για τον συστηματικό προσδιορισμό των μεταγραφικών ρυθμιστικών αλληλεπιδράσεων υπό πλήθος διαφορετικών συνθηκών (πχ. απόκριση σε αλλαγές pH, αντιβιοτικά, συγκέντρωση οξυγόνου κ.α). Ο αλγόριθμος είναι κατά βάση μια προέκταση των Δικτύων Συνάφειας. Υπολογίζει αρχικά την Αμοιβαία Πληροφορία για κάθε ζεύγος γονιδίων και έπειτα λαμβάνοντας υπόψη τις εμπειρικές κατανομές των τιμών της Αμοιβαίας Πληροφορίας υπολογίζει την τυπική τιμή (Z-score). Συγκεκριμένα, αντί να εξεταστεί η πληροφορία $I(X_i, X_j)$ μεταξύ των μεταβλητών X_i , και X_j ο αλγόριθμος εξετάζει το Z-score που υπολογίζεται ως εξής:

$$z_{ij} = \sqrt{z_i^2 + z_j^2}$$

όπου $z_i = \max(0, \frac{I(X_i, X_j) - \hat{\mu}_i}{\hat{\sigma}_i})$ και $\hat{\mu}_i$, $\hat{\sigma}_i$ η μέση και η τυπική απόκλιση από την εμπειρική κατανομή των τιμών της Αμοιβαίας Πληροφορίας $I(X_i, X_k)$ με $k=1, \dots, n$. Δηλαδή, το Z-score της κάθε ρυθμιστικής σχέσης εξαρτάται από την κατανομή των τιμών της Αμοιβαίας Πληροφορίας για όλους τους πιθανούς ρυθμιστές (z_i) και την κατανομή των τιμών της Αμοιβαίας Πληροφορίας για τα υποθετικά γονίδια-στόχους (z_j) (Faith et al., 2007).

4.1.4.3 Ο αλγόριθμος MRNET

Ο αλγόριθμος MRNET χρησιμοποιείται για την ανάλυση μεγάλων Γενετικών Ρυθμιστικών Δικτύων από δεδομένα μικροσυστοιχιών. Ο αλγόριθμος μοντελοποιεί το δίκτυο επιλέγοντας ένα υποσύνολο χαρακτηριστικών που έχουν τη μεγαλύτερη συσχέτιση ως προς μια κλάση (συνάφεια-relevance) και τη μικρότερη συσχέτιση ανά ζεύγη μεταξύ τους (πλεονασμός-redundancy). Στον αλγόριθμο αυτό τα χαρακτηριστικά που εξετάζονται, δηλαδή οι αλληλεπιδράσεις μεταξύ γονιδίων ρυθμιστών και γονιδίων στόχων ταξινομούνται με την τεχνική της Μέγιστης Συνάφειας-Ελάχιστου Πλεονασμού (maximum relevance/minimum redundancy-MRMR) (Meyer et al, 2007).

Η τεχνική αυτή βασίζεται στη λεγόμενη *προς τα εμπρός επιλογή* (forward selection). Πρακτικά κάθε γονίδιο θεωρείται γονίδιο στόχος και σε κάθε βήμα ο αλγόριθμος επιλέγει

το γονίδιο που έχει τη μεγαλύτερη τιμή Αμοιβαίας Πληροφορίας με το γονίδιο στόχο, δηλαδή παρουσιάζει τη μέγιστη συνάφεια (η οποία αντιστοιχεί σε ισχυρές αλληλεπιδράσεις). Ταυτόχρονα, ο αλγόριθμος επιλέγει το γονίδιο με τη χαμηλότερη τιμή Αμοιβαίας Πληροφορίας σε σχέση με τα προηγούμενα επιλεγμένα γονίδια (ελάχιστος πλεονασμός) η οποία αντιστοιχεί σε ασθενείς αλληλεπιδράσεις. Με τον τρόπο αυτό το γονίδιο που μεγιστοποιεί την ακόλουθη σχέση συμπεριλαμβάνεται στο σύνολο των "επιλεγμένων" γονιδίων (X_S):

$$S_j = I(X_j, X_T) - \frac{1}{|X_S|} \sum_{X_k \in X_S} I(X_j, X_k)$$

Στην παραπάνω σχέση η μεταβλητή X_T αναφέρεται στο γονίδιο στόχο και η X_S δηλώνει το σύνολο των επιλεγμένων γονιδίων που φαίνεται να σχετίζονται με το γονίδιο στόχο.

Κύριος περιορισμός του αλγορίθμου, κοινός και σε μεθόδους όπως το ARACNE και το CLR, είναι ότι δεν καθορίζεται η κατεύθυνση των αλληλεπιδράσεων, καθώς η Αμοιβαία Πληροφορία είναι συμμετρικό μέτρο (Meyer et al, 2007). Ένας ακόμη περιορισμός είναι το γεγονός ότι η ποιότητα του επιλεγμένου σετ γονιδίων εξαρτάται αποκλειστικά από την σωστή επιλογή της πρώτης μεταβλητής και δεν αποδίδει καλά εάν αυτή είναι λανθασμένη (Meyer et al., 2010).

Ο αλγόριθμος MRNET Backward (MRNETB), αποτελεί μια βελτιωμένη εκδοχή του MRNET και έχει σαν στόχο να αντιμετωπίσει τον περιορισμό της επιλογής της "σωστής" πρώτης μεταβλητής. Το MRNETB ξεπερνά αυτόν τον περιορισμό με την εφαρμογή της μεθόδου του *προς τα πίσω περιορισμού* (backward elimination) (Meyer et al., 2010). Πρακτικά ο αλγόριθμος εισάγει όλες τις υποψήφιες μεταβλητές στο μοντέλο και σταδιακά απορρίπτει σε κάθε βήμα εκείνη που είναι λιγότερο σημαντική, η οποία είναι αυτή με τη μικρότερη τιμή P-value. Η διαδικασία συνεχίζεται έως ότου οι μεταβλητές που έχουν απομείνει να έχουν μεμονωμένα μικρότερες τιμές P-value από ορισμένες τιμές (οι οποίες είναι συνήθως 0.05 ή 0.10) (Agresti, 2007). Η μέθοδος του προς τα πίσω περιορισμού έχει φανεί ότι υπερτερεί σε σχέση με τη μέθοδο της προς τα εμπρός επιλογής ως προς την επιλογή των μεταβλητών, επομένως και στην απόδοση του προβλεπόμενου δικτύου (Meyer et al., 2010).

Οι πιο πολλές προσεγγίσεις επιλογής χαρακτηριστικών από δεδομένα γονιδιακής έκφρασης μικροσυστοιχιών δεν μπορούν να διαχειριστούν χρονικά πολυμεταβλητά δεδομένα χωρίς να εφαρμόσουν κάποια μέθοδο ταξινόμησης δεδομένων (data flattening), με αποτέλεσμα να χάνεται η χρονική πληροφορία. Μία παραλλαγή της μεθόδου MMRMR που έχει προταθεί για χρονικά πολυμεταβλητά δεδομένα γονιδιακής έκφρασης είναι το TMRMR. Συνοπτικά, η μέθοδος εκτιμά τη συνάφεια ενός γονιδίου υπολογίζοντας τον μέσο όρο των στατιστικών τιμών F για κάθε χρονική στιγμή και τον πλεονασμό μεταξύ γονιδίων

με τη χρήση της χρονικής δυναμικής στρέβλωσης (dynamic time warping), μια προσέγγιση που μετρά την ομοιότητα μεταξύ δυο χρονικών σειρών. Με το TMRMR επιτυγχάνεται μεγαλύτερη ακρίβεια υπολογισμών και επιλογή χαρακτηριστικών μεγαλύτερης διακριτικής ικανότητας (Radovic, Ghalwash, et al. 2017).

4.1.5 Ο αλγόριθμος C3NET

Το C3NET είναι ακόμη ένας αλγόριθμος επαγωγής Γενετικών Ρυθμιστικών Δικτύων ο οποίος αποτελείται από δύο βήματα. Αρχικά, όπως και στους αλγορίθμους Relevance, ARACNE και CLR όλες οι μη σημαντικές τιμές Αμοιβαίας Πληροφορίας εξαλείφονται από τον πίνακα, με βάση ένα καθορισμένο κατώφλι. Στο δεύτερο βήμα ο αλγόριθμος επιλέγει τη σχέση (ακμή) με την υψηλότερη τιμή Αμοιβαίας Πληροφορίας για κάθε γονίδιο. Η σχέση αυτή συνδέεται ταυτόχρονα με τις γειτονικές σχέσεις με υψηλή ΑΠ για κάθε γονίδιο. Αυτό υποδηλώνει ότι ο μέγιστος αριθμός των επιλεγμένων ακμών θα να είναι ίσος με τον αριθμό των γονιδίων που εξετάζονται. Ο τεχνική που χρησιμοποιεί ο αλγόριθμος φαίνεται να δίνει καλύτερα αποτελέσματα συγκριτικά με άλλες παρόμοιες μεθόδους που έχουν αναφερθεί. Σημαντικό στοιχείο του C3NET για την καλή του απόδοση είναι η χαμηλή υπολογιστική πολυπλοκότητα (Altay & Emmert-Streib, 2011).

4.2 Μεταβολικά Δίκτυα

4.2.1 Εισαγωγή

Ο μεταβολισμός είναι το άθροισμα των βιοχημικών αντιδράσεων που καταλύονται από ένζυμα σε ένα κύτταρο. Όλες αυτές οι αντιδράσεις σχηματίζουν "μεταβολικές οδούς" που συνδέονται μεταξύ τους κι έτσι διαμορφώνουν ένα δυναμικό κύκλωμα, το οποίο ονομάζεται Μεταβολικό Δίκτυο. Ο μεταβολισμός αποτελεί, υπό μία έννοια, τον διαμεσολαβητή μεταξύ των οργανισμών και του περιβάλλοντός κι αυτό διότι εξωτερικοί παράγοντες (όπως θερμοκρασία, συγκέντρωση ουσιών κ.α.) επηρεάζουν τον οργανισμό μέσω του μεταβολισμού του. Η πρόσφατη πρόοδος στην ανάλυση του γονιδιώματος και των ομικών τεχνολογιών έχουν καταστήσει δυνατή την εξαγωγή πληροφοριών που σχετίζονται με το μεταβολισμό στη γενωμική κλίμακα. Ως εκ τούτου, έχει σημειωθεί μεγάλη

πρόοδος στην ανακατασκευή δικτύων και την κατασκευή μοντέλων πρόβλεψης (Hütt et al., 2007).

Τα Μεταβολικά Δίκτυα μπορούν να περιγραφούν από το σύνολο των βιοχημικών σχέσεων μεταξύ μεταβολιτών και ενζύμων (κυρίως πρωτεΐνες) εντός του κυττάρου. Η αναπαράστασή τους γίνεται με τη βοήθεια γράφων όπου συνήθως οι ακμές αντιπροσωπεύουν τα ένζυμα και οι κόμβοι τους μεταβολίτες. Τα ένζυμα αλληλεπιδρούν με τους μεταβολίτες έτσι ώστε να καταλύουν μεταβολικές αντιδράσεις ή να μεταφέρουν μεταβολίτες (Walhout, Vidal, & Dekker, 2012). Υπάρχουν πολλοί τρόποι αναπαράστασης των Μεταβολικών Δικτύων με τη μορφή γράφων, η επιλογή εκ των οποίων εξαρτάται από τις ανάγκες και τον σκοπό της μελέτης. Οι πιο κοινοί είναι οι εξής:

- α. Το "δίκτυο αντίδρασης", στο οποίο κάθε κόμβος αναπαριστά μια αντίδραση και κάθε ακμή ενώνει τα προϊόντα μιας αντίδρασης με τα αντιδρώντα της άλλης.
- β. Το "δίκτυο υποστρώματος", όπου οι κόμβοι αντιπροσωπεύουν τους μεταβολίτες και οι ακμές τις μεταβολικές αντιδράσεις (Chalancon, Kruse, & Babu, 2013).
- γ. Το "διμερές δίκτυο"⁵ (bipartite), το οποίο περιλαμβάνει δύο κατηγορίες κόμβων, μία για τους μεταβολίτες και μία για τις αντιδράσεις. Οι ακμές ενώνουν μεταβολίτες με αντιδράσεις (για τα υποστρώματα) ή αντιδράσεις με μεταβολίτες (για τα προϊόντα) (Van Helden, et al., 2002).

Η Μεταβολομική⁶ (Metabolomics) αποτελεί μια πειραματική προσέγγιση που αναλύει τις διαφορές στο επίπεδο του μεταβολίτη σε πειραματικά δείγματα. Έχει αναδειχθεί σε πολύτιμη τεχνική για την ανάλυση μικρών οργανικών μορίων σε βιολογικά συστήματα (πχ. οργανικά οξέα, αμινοξέα, σάκχαρα, στεροειδή κ.α). Έχει εφαρμογές στην ανίχνευση βιοδεικτών που σχετίζονται με ασθένειες, ή μεταβολές της δραστηριότητας των μονοπατιών που οφείλονται σε φαρμακευτικές ουσίες σε πλήθος περιπτώσεων (πχ καρκίνος, διαβήτης, κυστική ίνωση, τοξικολογία κ.α). Η βελτίωση των τεχνολογιών που χρησιμοποιούνται περισσότερο στη Μεταβολομική, όπως η φασματομετρία μάζας και η τεχνική αναλυτικής χημείας NMR, έχει συντελέσει στην δημιουργία μεγάλων βάσεων δεδομένων, που αποτελούνται από διάφορες κατηγορίες μεταβολιτών (Cicek, et al., 2013).

Οι τεχνικές της Μεταβολομικής που είναι βασισμένες στη φασματομετρία μάζας, είναι ευαίσθητες ως προς την ταυτόχρονη ανάλυση μεγάλου αριθμού ενώσεων. Αν και υφίστανται περιορισμούς όσον αφορά τις δυνατότητες ποσοτικοποίησης, η τεράστια ποσότητα διαθέσιμων πληροφοριών σε ένα μόνο πείραμα (LC-MS ή GC-MS) μπορεί να

⁵ Διμερές (bipartite) ονομάζεται ένας γράφος G εάν οι κορυφές του μπορούν να χωριστούν σε δύο υποσύνολα X και Y, έτσι ώστε κάθε ακμή του G να συνδέει ένα στοιχείο του X με ένα στοιχείο του Y.

⁶ Ένας πιο πλήρης ορισμός είναι ο εξής: Μεταβολομική είναι το πεδίο μελέτης που ασχολείται με την ποσοτική μέτρηση της δυναμικής πολυπαραμετρικής μεταβολικής αντίδρασης των έμβιων συστημάτων σε παθοφυσιολογικά ερεθίσματα ή γενετική τροποποίηση (Zhao, Vaziri, & Lin, 2015)

παρέχει λεπτομερείς πληροφορίες σχετικά με τα πρότυπα αλλαγής του μεταβολίτη σε ολόκληρο το μεταβολικό δίκτυο. Η τεχνική NMR χρησιμοποιείται συμπληρωματικά με τη φασματομετρία μάζας η οποία αν και είναι περιορισμένη όσον αφορά την ευαισθησία, είναι ικανή να διασαφηνίσει τη μοριακή δομή. Σημαντικό πρόσθετο χαρακτηριστικό της τεχνικής NMR είναι ότι παρέχει ποσοτικές πληροφορίες, ικανές να εντοπίσουν τα επίπεδα των ανιχνευόμενων ενώσεων.

Η ερμηνεία και η επαγωγή συμπερασμάτων από σύνθετα σύνολα μεταβολομικών δεδομένων αποτελεί τα τελευταία χρόνια εξαιρετική πρόκληση, ενώ ταυτόχρονα γίνονται μεγάλες προσπάθειες για την ενσωμάτωση των μεταβολομικών δεδομένων στην Γονιδιωματική και Πρωτεωμική. Βασικός στόχος της Μεταβολομικής είναι να μετρηθεί και να διερευνηθεί η βιοχημική ανάλυση ολόκληρης της διαδρομής από τον γονότυπο στον φαινότυπο προσφέροντας έτσι νέες γνώσεις στη Βιολογία και την Ιατρική (Padmanabhan, et. al., 2014).

4.2.2 Ανάλυση Μεταβολικών Ροών

Τα κύτταρα έχουν τη δυνατότητα να προσαρμόζουν τις μεταβολικές ροές τους, σαν απόκριση στις αλλαγές του περιβάλλοντος. Η ακριβής μέτρηση των ροών του μεταβολικού μονοπατιού, δηλαδή του ρυθμού με τον οποίο ένα υπόστρωμα μετατρέπεται σε προϊόν, αποτελεί θεμελιώδη στόχο για την κυτταρική Φυσιολογία και τη Μεταβολική Μηχανική (Lee, Park, & Kim, 2011). Για την ανάλυση του κυτταρικού μεταβολισμού έχουν αναπτυχθεί σημαντικά υπολογιστικά εργαλεία που περιλαμβάνουν τον προσδιορισμό των μεταβολικών ροών που καθορίζονται από τα ποσοστά των καταλυόμενων αντιδράσεων που συμμετέχουν σε ένα Μεταβολικό Δίκτυο. Αναλόγως τις ανάγκες της μελέτης, τα υπολογιστικά αυτά εργαλεία μπορούν να κατηγοριοποιηθούν σε τρεις κατηγορίες (Trinh, Wlaschin, & Srien, 2009): α. *Ανάλυση Μεταβολικών Ροών*, β. *Ανάλυση Ισορροπίας Ροών*, γ. *Ανάλυση Μεταβολικού Μονοπατιού*.

Η αναλυτική τεχνική που χρησιμοποιείται για τον ποσοτικό προσδιορισμό των ενδοκυτταρικών μεταβολικών ροών, ως αποτέλεσμα των καταλυτικών και μεταγραφικών αλληλεπιδράσεων, είναι γνωστή ως Ανάλυση των Μεταβολικών Ροών (Metabolic Flux Analysis -MFA). Η τεχνική βασίζεται στην στοιχειομετρία των μεταβολικών αντιδράσεων καθώς και στην εφαρμογή *ισοζυγίων μάζας* στους ενδοκυττάρους μεταβολίτες, σε κατάσταση ψευδο-ισορροπίας (Lee et al., 2011).

Η Ανάλυση Ισορροπίας Ροών (Flux Balance Analysis-FBA) είναι μια ευρέως διαδεδομένη υπολογιστική μέθοδος ανάλυσης ροών η οποία υπολογίζει την ενδοκυτταρική κατανομή ροών σε ένα μεταβολικό δίκτυο μεγιστοποιώντας ή ελαχιστοποιώντας ένα σύνολο γραμμικών αλγεβρικών εξισώσεων. Η προσέγγιση βασίζεται στην υπόθεση ότι οι

μεταβολικές αντιδράσεις είναι σε ισορροπία, δηλαδή οι συγκεντρώσεις των αντιδρώντων βρίσκονται σε σταθερή κατάσταση (Lee et al., 2011; Walhout et al., 2012).

Η Ανάλυση του Μεταβολικού Μονοπατιού (Metabolic Pathway Analysis) προσδιορίζει την τοπολογία του κυτταρικού μεταβολισμού με βάση μόνο τη στοιχειομετρική δομή και τους θερμοδυναμικούς περιορισμούς των αντιδράσεων όπου οι κινητικές παράμετροι δεν απαιτούνται για τους υπολογισμούς. Αυτός ο τύπος ανάλυσης έχει εφαρμοστεί με επιτυχία σε διάφορους οργανισμούς για να τη διερεύνηση της δομής του μεταβολικού δικτύου, της ευρωστίας, της ευθραυστότητα, της ρύθμισης κλπ. Ένα χρήσιμο εργαλείο της προσέγγισης είναι η Ανάλυση Στοιχειώδους Λειτουργίας (Elementary Mode Analysis - EMA), για τον προσδιορισμό της δομής ενός μεταβολικού δικτύου που συνδέει τον κυτταρικό φαινότυπο με τον αντίστοιχο γονότυπο. Η ανάλυση αυτή μπορεί να εντοπίσει όλες τις πιθανές μεταβολικές οδούς που είναι γνωστές ως "στοιχειώδεις λειτουργίες" (elementary modes) οι οποίες είναι εγγενείς σε ένα μεταβολικό δίκτυο. Κάθε στοιχειώδης λειτουργία περιλαμβάνει ένα ελάχιστο και ταυτόχρονα μοναδικό σύνολο ενζυμικών αντιδράσεων που μπορεί να υποστηρίξει τις κυτταρικές λειτουργίες (Trinh et al., 2009).

Ο μεταβολισμός μπορεί να εκφραστεί ως ένας σταθμισμένος, γραμμικός συνδυασμός στοιχειωδών λειτουργιών. Η κατανομή μιας μεταβολικής ροής ενός Μεταβολικού Δικτύου μπορεί να ανακατασκευαστεί με βάση τις στοιχειώδεις λειτουργίες και τις πιθανότητες χρήσης τους. Ο μεταβολισμός ενός κυττάρου αναμένεται να προσαρμοστεί σε μια δεδομένη κατάσταση μεταβάλλοντας τη χρήση των στοιχειωδών λειτουργιών, προκειμένου να πετύχει τη βέλτιστη μεταβολική κατάσταση (Srienc & Unrean, 2010).

Οι Unrean και Srienc απέδειξαν θεωρητικά ότι τα Μεταβολικά Δίκτυα εξελίσσονται προς καταστάσεις *μέγιστης παραγωγής Εντροπίας*. Δηλαδή, στην πιο πιθανή μεταβολική κατάσταση, η πιθανότητα χρήσης μεμονωμένων στοιχειωδών λειτουργιών έχει την κατανομή του νόμου του Boltzmann και ισχύει η Αρχή της Μέγιστης Εντροπίας (§2.10). Για να αποδειχτεί ότι ένα Μεταβολικό Δίκτυο εξελίσσεται προς μία κατάσταση Μέγιστης Εντροπίας πραγματοποίησαν πειράματα προσαρμοστικής εξέλιξης σε βακτήρια με μειωμένη μεταβολική λειτουργικότητα (περιορισμένο σύνολο στοιχειωδών λειτουργιών). Συνεπώς, η κατάληξη των πειραμάτων ήταν πως η κυτταρική λειτουργία έχει την τάση να λειτουργεί με τον μέγιστο ρυθμό παραγωγής Εντροπίας χρησιμοποιώντας κυρίως τα πιο αποδοτικά μονοπάτια τα οποία ήταν αυτά με τις μικρές Εντροπίες αντίδρασης. Συμπερασματικά, η εξέλιξη ενός πολύπλοκου βιολογικού δικτύου φαίνεται ότι μπορεί να προβλεφτεί ποσοτικά με τη βοήθεια βασικών αρχών της Στατιστικής Μηχανικής (Srienc & Unrean, 2010; Unrean & Srienc, 2011).

4.2.2.1 Ο αλγόριθμος ADEMA

Ο αλγόριθμος ADEMA (Algorithm for Determining Expected Metabolic Alterations) αποτελεί μια πολυμεταβλητή τεχνική ανάλυσης Μεταβολικών Δικτύων που βασίζεται στην Αμοιβαία Πληροφορία. Ο αλγόριθμος αποτελεί σημαντικό εργαλείο καθώς μπορεί να αναλύει τις μη γραμμικές σχέσεις πολλαπλών μεταβολιτών. Βασικός στόχος του είναι ο εντοπισμός των "αναμενόμενων" αλλαγών στο επίπεδο του μεταβολίτη σε σχέση με μια συγκεκριμένη κατάσταση. Η τεχνική έχει χρησιμοποιηθεί σε πλήθος μελετών α. για την ανίχνευση βιοδεικτών και β. για τον προσδιορισμό της δραστηριότητας βιοχημικών οδών που εμφανίζουν μεταβολές, οι οποίες οφείλονται σε ασθένειες ή φάρμακα.

Τα βήματα που ακολουθεί ο αλγόριθμος μπορούν να συνοψιστούν ως εξής:

1. Αρχικά, σχηματίζεται ένας πληθυσμός αποτελούμενος από άτομα μεταβλητών ομάδων και άτομα ομάδων ελέγχου και γίνεται μέτρηση των συγκεντρώσεων των μεταβολιτών ενδιαφέροντος.
2. Στο δεύτερο βήμα, κάθε μια από τις παρατηρήσεις ταξινομείται με βάση κάποια τιμή πιθανότητας
3. Στο τρίτο βήμα, σχηματίζεται το μεταβολικό (υπο)δίκτυο από τους μετρούμενους μεταβολίτες.
4. Στο τέταρτο βήμα εντοπίζονται τα υποσύνολα των μεταβολιτών που είναι σχετικά με το μεταβολικό δίκτυο.
5. Στο πέμπτο και τελευταίο βήμα υπολογίζονται οι αναμενόμενες τιμές για τα επίπεδα των μεταβολιτών (με βάση τις πιθανότητες στο βήμα 2) για τη μεταβλητή ομάδα και την ομάδα ελέγχου με τη χρήση της Αμοιβαίας Πληροφορίας. Στη συνέχεια γίνεται σύγκριση των δύο ομάδων προκειμένου να εντοπιστούν οι διαφορές στη μεταβλητή ομάδα με βάση την ομάδα ελέγχου.

Οι Cicek et al., έδειξαν ότι ο ADEMA προβλέπει καλύτερα τις μεταβολές στο επίπεδο του μεταβολίτη σε σχέση με άλλες τεχνικές, παρέχοντας αποτελέσματα με σημαντικά μεγαλύτερη ακρίβεια, προσδίδοντας τη δυνατότητα ενσωμάτωσης με άλλες τεχνικές υψηλής απόδοσης (Cicek et al., 2013).

4.2.3 Αντίστροφη Μηχανική Μεταβολικών Δικτύων

Καθώς τα Μεταβολικά Δίκτυα μοιράζονται κοινά χαρακτηριστικά με τα Γενετικά Ρυθμιστικά Δίκτυα, πλήθος μελετών έχουν εφαρμόσει επιτυχείς μεθόδους που έχουν αναπτυχθεί για τα Γενετικά Ρυθμιστικά στα Μεταβολικά Δίκτυα (Basso et al., 2005; Margolin et al., 2006; Nemenman et al., 2007). Αντίστοιχα με το επίπεδο γονιδιακής έκφρασης για τον εντοπισμό ρυθμιστικών σχέσεων ανάμεσα σε γονίδια, οι σχέσεις μεταξύ των μεταβολιτών μπορούν να μελετηθούν με τη χρήση των μέτρων της Θεωρίας της Πληροφορίας. Οι αλγόριθμοι ARACNE και CLR, που όπως είδαμε σε προηγούμενη ενότητα αναπτύχθηκαν για τη επαγωγή Γενετικών Ρυθμιστικών Δικτύων, έχουν χρησιμοποιηθεί επίσης για την ανακατασκευή Μεταβολικών Δικτύων (πχ. περίπτωση μελέτης του μεταβολισμού ερυθρών κυττάρων) (Rosato et al., 2018).

Οι μεταβολίτες μπορεί να εμφανίσουν συζευγμένες μεταβολές στη συγκέντρωση τους κάτω από διάφορες συνθήκες γονιδιακών διαταραχών και στρες, αλλά και λόγω της εγγενούς μεταβλητότητας ενός βιολογικού δικτύου. Η συγκέντρωση ενός μεταβολίτη μπορεί να χρησιμοποιηθεί ως δεδομένο εισόδου σε αλγόριθμους που χρησιμοποιούνται για την επαγωγή συμπερασμάτων στα Μεταβολικά Δίκτυα. Η Αμοιβαία Πληροφορία, εκτός των άλλων, έχει χρησιμοποιηθεί για τον προσδιορισμό της συσχέτισης μεταξύ ζευγών μεταβολιτών, αφού είναι εφικτό να εντοπίσει επιπλέον μη-γραμμικές συσχετίσεις μεταξύ μεταβολιτών, που δεν θα ήταν ανιχνεύσιμες με τη χρήση αποκλειστικά και μόνο του συντελεστή συσχέτισης Pearson (Numata, Ebenhoeh, & Knapp, 2009).

4.2.3.1 Ο αλγόριθμος MIRA

Ο MIRA (Mutual Information-based Reporter Algorithm) είναι ένας πολυπαραγοντικός και συνδυαστικός αλγόριθμος που βασίζεται στην Αμοιβαία Πληροφορία και χρησιμοποιείται για να υπολογίσει τη συνολική μεταγραφική απόκριση γύρω από έναν μεταβολίτη. Έχει εφαρμοστεί, κατά βάση, για την ανίχνευση μεταβολικών φαρμακευτικών στόχων και για την κατανόηση της σχέσης μεταξύ της γονιδιακής έκφρασης και της μεταβολικής δραστηριότητας. Ο αλγόριθμος αποτελείται από δύο βασικές μεταβλητές οι οποίες είναι α. ο γονότυπος (πχ. ομάδα μελέτης, ομάδα ελέγχου) και β. οι συνδυασμοί γονιδίων που σχετίζονται με μια αντίδραση και τα επίπεδα έκφρασής τους αντίστοιχα. Πιο αναλυτικά, τα βήματα που ακολουθούνται είναι τα εξής:

1. Κατασκευάζεται το μεταβολικό δίκτυο, το οποίο αναπαρίσταται με τη βοήθεια του γράφου $G(V,E)$, όπου V το σύνολο των κόμβων που αποτελούνται από μεταβολίτες, γονίδια και αντιδράσεις και E το σύνολο δύο τύπων ακμών που συνδέουν αντιδράσεις με γονίδια και μεταβολίτες με αντιδράσεις.

2. Δημιουργούνται σύνολα γονιδίων για κάθε μεταβολίτη με βάση τη σύνδεσή του με τις γειτονικές αντιδράσεις.
3. Τα επίπεδα έκφρασης των γονιδίων ταξινομούνται.
4. Υπολογίζεται η Αμοιβαία Πληροφορία μεταξύ των συνδυασμών των επιπέδων γονιδιακής έκφρασης και της μεταβλητής ομάδας, για κάθε ομάδα γονιδίων.
5. Υπολογίζεται η μέση Αμοιβαία Πληροφορία για να υπολογιστεί στη συνέχεια η μέση μεταγραφική απόκριση γύρω από έναν μεταβολίτη.
6. Οι μεταβολίτες κατατάσσονται με βάση τη μέση Αμοιβαία Πληροφορία και προσδιορίζονται οι *μεταβολίτες αναφοράς*⁷ (reporter metabolites).

Ο αλγόριθμος MIRA αποτελεί μια βελτίωση του αλγορίθμου RA (Reporter Algorithm) που έχει χρησιμοποιηθεί σε πολυάριθμες μελέτες για την ανίχνευση των θερμών σημείων (hotspots) στα Μεταβολικά Δίκτυα γύρω από τα οποία "επικεντρώνεται" η μεταγραφική ρύθμιση σε περιπτώσεις ασθένειας ή γενετικής διαταραχής (Cicek, Roeder, & Ozsoyoglu, 2015).

4.2.3.2 Ο αλγόριθμος MIDER

Ο αλγόριθμος MIDER (Mutual Information Distance and Entropy Reduction) είναι ένας αλγόριθμος με ευρύ φάσμα εφαρμογών σχετικών με την επαγωγή δομών δικτύου. Μπορεί να χρησιμοποιηθεί για την ανακατασκευή οποιουδήποτε είδους δικτύου συμπεριλαμβανομένων των μεταβολικών, γενετικών ρυθμιστικών και σηματοδοτικών (Villaverde *et al.*, 2014). Ο αλγόριθμος δέχεται ως δεδομένα εισόδου δεδομένα χρονοσειρών σχετικά με τα ποσοτικά χαρακτηριστικά των κόμβων του δικτύου και λαμβάνει υπόψη τις χρονικές υστερήσεις μεταξύ μεταβλητών.

Ο αλγόριθμος βασίζεται σε προσεγγίσεις σχετικές με τη Θεωρία της Πληροφορίας προκειμένου να προβλέψει τις σχέσεις μεταξύ μεταβλητών αλλά και να διακρίνει τις άμεσες από τις έμμεσες αλληλεπιδράσεις μεταξύ τους. Παρέχει αρχικά, μια αναπαράσταση του δικτύου στο οποίο η απόσταση μεταξύ των κόμβων υποδεικνύει τη στατιστική τους εγγύτητα. Έπειτα, γίνεται χρήση της τεχνικής της *μείωσης* της Εντροπίας, έτσι ώστε να διακρίνει τις άμεσες από τις έμμεσες αλληλεπιδράσεις και να καθοριστεί η κατεύθυνση αυτών. Πιο αναλυτικά, υλοποιούνται οι παρακάτω διαδικασίες:

⁷ Ως *μεταβολίτες αναφοράς* αναφέρονται οι μεταβολίτες γύρω από τους οποίους συμβαίνουν οι πιο σημαντικές μεταγραφικές μεταβολές (Patil & Nielsen, 2005).

1. Υπολογίζονται αρχικά οι πολυδιάστατες Εντροπίες με χρονική υστέρηση και η Αμοιβαία Πληροφορία από τα δεδομένα που εισάγονται, με βάση τις οποίες φτιάχνεται ένας πίνακας αποστάσεων ανάμεσα στις μεταβλητές. Στη συνέχεια γίνεται οπτικοποίηση του πίνακα των αποστάσεων σε χάρτη δύο διαστάσεων.
2. Η ισχύς μιας αλληλεπίδρασης ανάμεσα σε δύο μεταβλητές υπολογίζεται μέσω της τεχνικής της *μείωσης* της Εντροπίας της πρώτης μεταβλητής που οφείλεται στην προσθήκη μιας δεύτερης μεταβλητής. Η τεχνική βασίζεται στην Υπό Συνθήκη Εντροπία και εφαρμόζεται έτσι ώστε να βελτιώσει τον απεικονιζόμενο χάρτη, εντοπίζοντας τις άμεσες από τις έμμεσες αλληλεπιδράσεις. Συνοπτικά, αν ένα είδος μεταβλητής X^* συνδέεται με μια μεταβλητή Y (με X^* το υποσύνολο με το οποίο έχει προβλεφτεί ότι συνδέεται) τότε η μείωση της Εντροπίας της Y λόγω της X^* , $E_R(Y, X^*)$ είναι:

$$E_R(Y, X^*) = \frac{H(Y|X^*) - (Y|X^*, X^*)}{H(Y)} > K$$

όπου K είναι ένα κατώφλι το οποίο μπορεί να οριστεί από τον χρήστη ή να υπολογιστεί αυτόματα ως συνάρτηση των τιμών της Εντροπίας (Villaverde et al., 2014).

3. Αποδίδεται η κατευθυντικότητα για κάθε ζεύγος συνδέσεων με τη βοήθεια Εντροπίας Μεταφοράς (transfer entropy), ένα μη συμμετρικό μέτρο για την αιτιότητα. Για κάθε ζεύγος μεταβλητών (X , Y) για τις οποίες προβλέπεται μια σύνδεση, ο αλγόριθμος υπολογίζει τις δύο Εντροπίες Μεταφοράς και αποδίδει την αιτιότητα στην κατεύθυνση που αντιστοιχεί σε αυτή με τη μεγαλύτερη τιμή. Η Εντροπία Μεταφοράς βασίζεται στις Υπό Συνθήκη Εντροπίες με χρονική υστέρηση (t). Η Εντροπία Μεταφοράς $T_{X \rightarrow Y}$ από την X στην Y μπορεί να εκφραστεί ως:

$$T_{X \rightarrow Y} = H(Y^t | Y^{t-\tau}) - H(Y^t | Y^{t-\tau}, X^{t-\tau})$$

Τα αποτελέσματα από την εφαρμογή του MIDER σε επτά διαφορετικά προβλήματα αναφοράς υποδεικνύουν ότι ο αλγόριθμος είναι κατάλληλος για τη μοντελοποίηση όλων των ειδών των βιολογικών δικτύων. Σημαντικό στοιχείο του είναι ότι μπορεί να εντοπίζει πολλαπλές αλληλεπιδράσεις αποφεύγοντας τα false positives, ενώ ταυτόχρονα παρέχει μεγάλη ακρίβεια υπολογισμών. Η χρήση του δεν απαιτεί εκ των προτέρων γνώση από τον χρήστη ενώ ταυτόχρονα μπορεί να προσαρμοστεί εύκολα στις ανάγκες του, μέσω των επιλογών που παρέχει. Ως περιορισμός, μπορεί να θεωρηθεί το ότι απαιτείται μεγάλη ποσότητα δεδομένων για να είναι αξιόπιστες οι εκτιμήσεις πολλών μεταβλητών, ως εκ τούτου, αυτό μπορεί εύκολα να ξεπεραστεί (Villaverde et al., 2014).

4.2.4 Μελέτη Ευρωστίας Δικτύου

Η ευρωστία (robustness) είναι η ικανότητα ενός συστήματος να διατηρεί τις βασικές του λειτουργίες σε περιπτώσεις εσωτερικών και εξωτερικών διαταραχών και σφαλμάτων. Αποτελεί βασικό χαρακτηριστικό των πολύπλοκων συστημάτων συμπεριλαμβανομένων των πολύπλοκων δικτύων. Στα πολύπλοκα δίκτυα, η Ευρωστία αναφέρεται στην ικανότητα του δικτύου να εκτελεί τις βασικές του λειτουργίες ακόμη και αν κάποιος από τους κόμβους ή τους συνδέσμους λείπουν (Albert, Jeong, & Barabási, 2000).

Παρομοίως, η Ευρωστία στα μοριακά δίκτυα μπορεί να εφαρμοστεί σε διάφορα επίπεδα οργάνωσης, αναφορικά με τη βιωσιμότητα του οργανισμού και τις ενδοκυτταρικές του διεργασίες (Demetrius & Manke, 2005). Ακόμη, ο όρος έχει συνδεθεί και με την έννοια της *εξελισσιμότητας* (evolvability), ως μέτρο της ικανότητας ενός πληθυσμού να ανταποκρίνεται στη φυσική επιλογή, ή αλλιώς τη δυνατότητα προσαρμογής ενός πληθυσμού στο περιβάλλον. Για παράδειγμα, ένα σύστημα μπορεί να θεωρηθεί περισσότερο ή λιγότερο εύρωστο σε σχέση με έναν δεδομένο αριθμό μεταλλάξεων, όπου η καταλληλότητα μιας μετάλλαξης εξαρτάται από το επίπεδο οργάνωσης στο οποίο εστιάζει. Μπορεί να σχετίζεται με την αλλαγή αμινοξέων πρωτεϊνών, με ρυθμιστικές μεταλλάξεις σε γενετικά κυκλώματα, ή με αλλαγές στις ενζυμικές αντιδράσεις στα Μεταβολικά Δίκτυα (Wagner, 2008).

Οι Demetrius και Manke έδειξαν ότι η Εντροπία Δικτύου σχετίζεται με την ικανότητα ενός δικτύου να ανταπεξέρχεται σε τυχαίες αλλαγές στη δομή του. Δίνουν πρακτικά έναν δομικό χαρακτηρισμό για την έννοια της Ευρωστίας, προτίνοντας την αναπαράσταση της τοπολογίας του δικτύου με όρους Εντροπίας Δικτύου, ως μια δομική ιδιότητα δικτύων. Η Εντροπία Δικτύου μπορεί να οριστεί ως συνάρτηση του αριθμού των κόμβων και των συνδέσεων (κατευθυνόμενων) μεταξύ γειτονικών κόμβων (Demetrius & Manke, 2005).

Οι Grimbs *et al.* μελέτησαν ποσοτικά και συνέκριναν τις ιδιότητες *σταθερότητας* (stability) των μεταβολικών καταστάσεων χρησιμοποιώντας ως περίπτωση μελέτης το Μεταβολικό Δίκτυο των ανθρώπινων ερυθρών κυττάρων. Η σταθερότητα σε μια μεταβολική σταθερή κατάσταση (steady state) είναι μια συστημική ιδιότητα που προκαλείται από τις κινητικές παραμέτρους του συνόλου των ενζύμων. Ταυτόχρονα, μεταβολές στις κινητικές παραμέτρους των επιμέρους ενζύμων μπορεί να επιφέρουν αλλαγές στη σταθερότητα μιας δεδομένης σταθερής κατάστασης. Βασικός στόχος της μελέτης ήταν η ποσοτική εκτίμηση των ιδιοτήτων που σχετίζονται με τη σταθερότητα και την ευρωστία των μεταβολικών καταστάσεων. Τα μέτρα που προτάθηκαν είναι ο συντελεστής συσχέτισης Pearson, η Αμοιβαία Εντροπία και το τεστ Kolmogorov-Sinai. Το τεστ Kolmogorov-Sinai χρησιμοποιήθηκε για να εντοπίσει τον αντίκτυπο των παραμέτρων

που μελετήθηκαν για την εκτίμηση της σταθερότητας του μεταβολικού συστήματος (Grimbs et al., 2007).

Στα Βιοχημικά Συστήματα, η Θεωρία της Πληροφορίας χρησιμοποιείται επίσης για την εκτίμηση της ποσότητας της πολυπλοκότητας της συμπεριλαμβανόμενης πληροφορίας. Πιο συγκεκριμένα, η Εντροπία Kolmogorov-Sinai περιγράφει το ρυθμό με τον οποίο μια στοχαστική διαδικασία παράγει πληροφορία και θεωρείται ένα καλό μέτρο για την ποσοτικοποίηση της πληροφορίας στους ελκυστές (attractors) (de la Fuente, 2010). Στο πεδίο των δυναμικών συστημάτων, ένας ελκυστής είναι ένα σύνολο αριθμητικών τιμών ως προς τις οποίες ένα σύστημα τείνει να εξελιχθεί, για μια ευρεία ποικιλία αρχικών συνθηκών του συστήματος (Boeing, 2016). Η έννοια του ελκυστή ειδικότερα, έχει χρησιμοποιηθεί στο πεδίο της Βιολογίας για να εξηγήσει κυτταρικούς τύπους και κυτταρικές αλλοιώσεις σε Δυναμικά Βιοχημικά Δίκτυα (Samal & Weber, 2018).

4.3 Πρωτεϊνικά Δίκτυα

4.3.1 Εισαγωγή

Η ανάλυση των πρωτεϊνικών αλληλεπιδράσεων είναι θεμελιώδης για την κατανόηση της οργάνωσης, των διαδικασιών και των λειτουργιών του κυττάρου. Οι πρωτεΐνες πολύ συχνά δεν επιτελούν αυτόνομες λειτουργίες αλλά οργανώνονται σε πολύ-πρωτεϊνικά σύμπλοκα επιτελώντας πολύπλοκες διεργασίες. Ακόμη, σύνθετες διεργασίες όπως π.χ. η ανοσοαπόκριση, απαιτούν την ταυτόχρονη λειτουργία περισσότερων το ενός συμπλόκων πρωτεϊνών. Η οργάνωση των πρωτεϊνικών μορίων σε σύμπλοκα και η συνδυασμένη δράση αυτών αποτελεί ανοιχτό πεδίο έρευνας. Ο χαρακτηρισμός άγνωστων λειτουργιών πρωτεϊνών μπορεί να προβλεφτεί μέσω της σύγκρισης των αλληλεπιδράσεων με άλλες όμοιες πρωτεΐνες των οποίων οι λειτουργίες είναι γνωστές. Η χρήση τεχνικών όπως π.χ. η φασματομετρία μάζας και οι μικροσυστοιχίες πρωτεϊνών έχουν εμπλουτίσει τα διαθέσιμα δεδομένα αλληλεπίδρασης πρωτεϊνών. Αυτό, έχει οδηγήσει σταδιακά στην ταυτοποίηση όλο και μεγαλύτερου αριθμού αλληλεπιδράσεων, τα οποία διευκολύνουν την κατασκευή δικτύων αλληλεπιδράσεων πρωτεϊνών (Vazquez, et al. , 2003; Zhang, 2009).

Ένα δίκτυο πρωτεϊνικών αλληλεπιδράσεων μπορεί να περιγραφεί ως ένα πολύπλοκο σύστημα πρωτεϊνών που συνδέονται με αλληλεπιδράσεις. Η πιο απλή αναπαράσταση ενός δικτύου πρωτεϊνών παίρνει τη μορφή ενός Γράφου που αποτελείται από κόμβους και ακμές. Οι πρωτεΐνες αντιπροσωπεύονται ως κόμβοι σε ένα τέτοιο γράφημα και οι αλληλεπιδράσεις μεταξύ τους ως ακμές. Οι αλληλεπιδράσεις είναι συχνά

δύο τύπων: α. στη πρώτη και απλούστερη περίπτωση, υπάρχει μια μη κατευθυνόμενη σύνδεση μεταξύ δύο πρωτεϊνών εάν σχηματίζουν σύμπλοκο πρωτεΐνης, το οποίο σχετίζεται με ορισμένες κυτταρικές λειτουργίες β. στη δεύτερη περίπτωση, υπάρχει μια κατευθυνόμενη σύνδεση από μία πρωτεΐνη σε μία άλλη εάν η πρώτη ρυθμίζει (θετικά ή αρνητικά) την δεύτερη (Jordan, Nguyen, & Liu, 2012).

Με βάση την αναπαράσταση ως Γράφο, διάφορες υπολογιστικές προσεγγίσεις, όπως η εξόρυξη δεδομένων, η μηχανική μάθηση και στατιστικές προσεγγίσεις, μπορούν να εφαρμοστούν για να αποκαλύψουν την οργάνωση των Πρωτεϊνικών Δικτύων σε διαφορετικά επίπεδα. Μια πιο λεπτομερής εξέταση της μορφής του γράφου μπορεί να αποφέρει ποικίλες γνώσεις για το σύστημα. Για παράδειγμα, οι γειτονικές πρωτεΐνες στον Γράφο θεωρείται γενικά ότι μοιράζονται λειτουργίες. Έτσι, οι λειτουργίες μιας πρωτεΐνης μπορούν να προβλεφθούν εξετάζοντας τις πρωτεΐνες με τις οποίες αλληλεπιδρά και τα πρωτεϊνικά σύμπλοκα στα οποία ανήκει. Επιπλέον, πυκνές συνδέσεις στο δίκτυο είναι πιθανό να σχηματίσουν σύμπλοκα πρωτεϊνών που λειτουργούν ως μονάδα σε μια συγκεκριμένη βιολογική διαδικασία. Ακόμη, η μελέτη των τοπολογικών χαρακτηριστικών του δικτύου (π.χ. αν είναι ανεξάρτητο κλίμακας, μικρό δίκτυο ή διέπεται από τον νόμο δύναμης) μπορεί να ενισχύσει την κατανόηση του βιολογικού συστήματος (Zhang, 2009).

Η Πρωτεομική (Proteomics) είναι το πεδίο μελέτης που πραγματεύεται την ανάλυση του πρωτεώματος, δηλαδή του συνόλου των πρωτεϊνών και των πεπτιδίων που κωδικοποιούνται και παράγονται από το γονιδίωμα ενός οργανισμού. Περιλαμβάνει τις εφαρμογές τεχνολογιών για την αναγνώριση και την ποσοτικοποίηση του συνόλου των πρωτεϊνών που περιέχονται σε ένα κύτταρο, ιστό ή οργανισμό. Η Πρωτεομική στοχεύει στην ταυτοποίηση των πρωτεϊνών ενός οργανισμού καθώς και στην κατανόηση της δομής και των λειτουργιών μιας συγκεκριμένης πρωτεΐνης. Επιπλέον, μελετά την αλληλεπίδραση των πρωτεϊνών σε σύμπλοκα, την ποσοτικοποίησή τους και την συσχέτιση με την κυτταρική λειτουργία. Η Πρωτεομική προσφέρει σημαντικές πληροφορίες για τη γονιδιακή έκφραση και έχει βασικό ρόλο στην κατανόηση μηχανισμών παθογένειας, την έγκαιρη διάγνωση νόσων, την πρόγνωση καθώς επίσης και στη την παρακολούθηση της ανάπτυξης νόσων. Επιπλέον, έχει εφαρμογές στην ανάπτυξη φαρμάκων, αλλά και πλήθος άλλων βιολογικών και βιοτεχνολογικών εφαρμογών (πχ. στην επιστήμη φυτών και τροφίμων) (Aslam, et al., 2017).

Η εκτεταμένη εφαρμογή πειραματικών τεχνικών υψηλής απόδοσης για τη μελέτη των πρωτεϊνικών αλληλεπιδράσεων έχει οδηγήσει στη δημιουργία μια μεγάλης τράπεζας δεδομένων Πρωτεομικής. Οι βασικότερες κατηγορίες τεχνικών είναι οι εξής:

- α. Οι μικροσυστοιχίες πρωτεϊνών, οι οποίες μπορούν να συλλάβουν ταυτόχρονα την έκφραση χιλιάδων πρωτεϊνών.

- β. Το διϋβριδικό σύστημα ζυμομυκήτων (yeast two-hybrid system- Y2H), το οποίο χρησιμοποιείται για την ανίχνευση πιθανών δυαδικών αλληλεπιδράσεων για οποιοδήποτε ζεύγος πρωτεϊνών που κωδικοποιούνται στο γονιδίωμα που μελετάται.
- γ. Η φασματομετρία μάζας, τεχνική η οποία αναλύει τη σύνθεση συμπλόκων πρωτεΐνης (Zhang, 2009).

Ωστόσο, τα δεδομένα που έχουν δημιουργηθεί από τις παραπάνω τεχνικές μπορεί να είναι συχνά ελλιπή ή μη αξιόπιστα. Επομένως, συμπληρωματικά με τις πειραματικές τεχνικές, έχουν εφαρμοστεί διάφορες υπολογιστικές μέθοδοι για την πρόβλεψη αλληλεπιδράσεων μεταξύ πρωτεϊνών που έχουν ανιχνευθεί πειραματικά. Δεδομένου ότι οι πρωτεΐνες που αλληλεπιδρούν έχουν αυξημένη πιθανότητα να είναι λειτουργικά συνδεδεμένες, τα υπολογιστικά εργαλεία στοχεύουν στη διερεύνηση δικτύων πρωτεϊνικών αλληλεπιδράσεων ώστε να μπορούν να προβλεφτούν λειτουργικά χαρακτηριστικά των πρωτεϊνών και νέες αλληλεπιδράσεις από τις ήδη γνωστές. Οι βασικότερες μεθοδολογίες μπορούν να συνοψιστούν ως ακολούθως:

- α. *Προσεγγίσεις γονιδιωματικής κλίμακας*, τυπικά πραγματοποιούν μια σύγκριση των γονιδιακών αλληλουχιών σε γονιδιώματα και συχνά δικαιολογούνται βάσει των συσχετισμένων εξελικτικών μηχανισμών των γονιδίων. Τον σκοπό αυτό εξυπηρετεί η διαθεσιμότητα πλήρων γονιδιωμάτων για διάφορους οργανισμούς, η οποία επιτρέπει την πρόβλεψη σχέσεων μεταξύ πρωτεϊνών σε γονιδιωματική κλίμακα.
- β. *Προσεγγίσεις με βάση την ακολουθία*, οι οποίες βασίζονται στην υπόθεση ότι τα ορθόλογα (orthologs) ενός πρωτεϊνικού ζεύγους τα οποία αλληλεπιδρούν σε έναν οργανισμό, θα αλληλεπιδρούν και σε έναν άλλον. Τα ορθόλογα είναι αντίστοιχα γονίδια σε διαφορετικά είδη, τα οποία εξελίσσονται από ένα κοινό γονιδιο-πρόγονο μέσω της ειδογένεσης.
- γ. *Προσεγγίσεις με βάση τη δομή*, οι οποίες αποσκοπούν στο να προβλέψουν την τρισδιάστατη φύση ενός πολυμερούς πρωτεϊνικού συμπλόκου. Η πρόσδεση (docking) είναι η κλασσική προσέγγιση για την ανίχνευση πρωτεϊνικών αλληλεπιδράσεων μέσω της πρόβλεψης της δομής των προσδεδεμένων πρωτεϊνών.
- δ. *Προσεγγίσεις με βάση τη μάθηση*, όπου δεδομένης μιας βάσης δεδομένων γνωστών αλληλεπιδρώντων ζευγαριών, ένα σύστημα μηχανικής μάθησης μπορεί να εκπαιδευτεί έτσι ώστε να αναγνωρίζει αλληλεπιδράσεις με βάση τα ειδικά βιολογικά χαρακτηριστικά τους.

- ε. Προσεγγίσεις με βάση την τοπολογία δικτύων, οι οποίες βασίζονται στην κατασκευή Πρωτεϊνικών Δικτύων για την πρόβλεψη νέων αλληλεπιδράσεων μεταξύ πρωτεϊνών ή την ταυτοποίηση αξιόπιστων αλληλεπιδράσεων (Zhang, 2009).

Στη συνέχεια της ενότητας θα συζητηθούν διάφορα προβλήματα που σχετίζονται με τα Πρωτεϊνικά Δίκτυα με προσεγγίσεις που βασίζονται στην τοπολογία δικτύων και εφαρμόζουν, ειδικότερα, μεθόδους από την Θεωρία της Πληροφορίας.

4.3.2 Αναγνώριση Πρωτεϊνικών Συμπλόκων

Τα πρωτεϊνικά σύμπλοκα αποτελούν θεμελιώδεις λειτουργικές μονάδες και έχουν βασικό ρόλο στην εκτέλεση πολλών βιολογικών διεργασιών. Η αναγνώριση ενός πρωτεϊνικού συμπλόκου μπορεί να δώσει σημαντικές πληροφορίες για το πώς οργανώνονται οι πρωτεΐνες εντός του κυττάρου και κατ' επέκταση για τις λειτουργίες τις οποίες επιτελούν. Το πεδίο μελέτης παρουσιάζει έντονο ενδιαφέρον στη βιοϊατρική έρευνα, καθώς πολυάριθμες μελέτες υποδεικνύουν ότι τα πρωτεϊνικά σύμπλοκα εμπλέκονται σε οδούς ασθενειών (Ideker & Sharan, 2008; Shen et al., 2017).

Έχουν προταθεί διάφοροι υπολογιστικοί αλγόριθμοι για την αναγνώριση πρωτεϊνικών συμπλόκων από Δίκτυα Πρωτεϊνικών αλληλεπιδράσεων. Οι μεθοδολογίες που έχουν αναπτυχθεί αναζητούν υποδίκτυα υψηλής συνδεσιμότητας σε Πρωτεϊνικά Δίκτυα, με την παραδοχή ότι οι πρωτεΐνες σε ένα σύμπλοκο πρωτεΐνης έχουν υψηλό βαθμό αλληλεπίδρασης (Lam & Chan, 2012).

4.3.2.1 Εντροπία Γράφου

Ένα Πρωτεϊνικό Δίκτυο αναπαρίσταται ως μη-κατευθυνόμενος και μη-σταθμισμένος Γράφος $G(V,E)$ με ένα σύνολο κορυφών V που αναπαριστούν τις πρωτεΐνες και ένα σύνολο ακμών E που αναπαριστούν τις αλληλεπιδράσεις μεταξύ τους. Ένας Γράφος λέγεται σταθμισμένος όταν οι ακμές του χαρακτηρίζονται από έναν αριθμό που ονομάζεται βάρος. Μια συστάδα (cluster) στην περίπτωση αυτή θεωρείται ο υπαγόμενος Υπογράφος (subgraph) $G'(V',E')$ του γράφου G , όπου $N(v)$ το σύνολο των γειτόνων μιας κορυφής v στο V . Ο βαθμός της v είναι η απόλυτη τιμή του $N(v)$, $|N(v)|$ ο οποίος αναπαριστά τον αριθμό των άμεσων συνδέσεων ή ακμών που έχει η κορυφή με άλλες κορυφές του δικτύου. Οι εσωτερικές συνδέσεις μιας κορυφής v ορίζονται ως οι ακμές από τη v στις κορυφές της V' . Τότε, ως $ri(v)$ ορίζεται η πιθανότητα της v να έχει εσωτερικές συνδέσεις. Επομένως έχουμε:

$$p_i(v) = \frac{n}{|N(v)|} \quad (i)$$

όπου n ο αριθμός των γειτονικών κορυφών της v στη V' . Παρομοίως, ως $p_0(v)$ ορίζεται η πιθανότητα της v να έχει εξωτερικές συνδέσεις το οποίο περιγράφεται ως:

$$p_0(v) = 1 - p_i(v) \quad (ii)$$

Τότε η Εντροπία $e(v)$ μιας κορυφής v με βάση την πιθανότητα κατανομής της ορίζεται ως:

$$e(v) = -p_i(v) \log_2 p_i(v) - p_0(v) \log_2 p_0(v) \quad (iii)$$

Η Εντροπία $e(G)$ ενός Γράφου $G(V,E)$ τότε, ορίζεται ως το άθροισμα της εντροπίας όλων των κορυφών στο G :

$$e(G) = \sum_{v \in V} e(v) \quad (iv)$$

Η συνδεσιμότητα μεταξύ γειτονικών σημείων σε μια συστάδα αποτελεί δείκτη της ποιότητάς του. Στην προκειμένη περίπτωση, μπορούμε να πούμε ότι η συστάδα έχει υψηλή ποιότητα εάν έχει πυκνές συνδέσεις στο εσωτερικό της συστάδας και αραιές προς το εξωτερικό αυτής. Η Εντροπία Γράφου έχει διατυπωθεί έτσι ώστε να μετρά την ποιότητα της συσταδοποίησης. Ένας Γράφος με χαμηλή Εντροπία υποδηλώνει ότι οι κορυφές της συστάδας έχουν πιο πολλούς εσωτερικούς συνδέσμους και λιγότερους εξωτερικούς.

Ο αλγόριθμος συσταδοποίησης, βρίσκει τις συστάδες με τη μικρότερη Εντροπία Γράφου όπως περιγράφεται στη συνέχεια:

1. Επιλέγεται μια τυχαία αρχική κορυφή και σχηματίζεται μια αρχική συστάδα λαμβάνοντας υπόψη την αρχική κορυφή και τους γείτονές του.
2. Οι κορυφές που μειώνουν την Εντροπία Γράφου προστίθενται ή αφαιρούνται από την συστάδα έως ότου να βρεθεί η βέλτιστη συστάδα (δηλ. η συστάδα με τη μικρότερη Εντροπία Γράφου)
3. Η διαδικασία επιλογής αρχικής κορυφής και εύρεσης της βέλτιστης συστάδας επαναλαμβάνεται μέχρις ότου να μην υπάρχουν άλλες κορυφές για να χρησιμοποιηθούν ως αρχικές (Lian, Song, & Cho, 2010).

Ο αλγόριθμος στοχεύει στην αναγνώριση συμπλόκων πρωτεϊνών και στην πρόβλεψη λειτουργικών τμημάτων, δηλαδή ομάδων πρωτεϊνών οι οποίες συμμετέχουν στις ίδιες βιολογικές διεργασίες. Η Εντροπία Γράφου είναι ένα μέτρο που εφαρμόζει τον ορισμό της Εντροπίας σε έναν Γράφο, ώστε να εξετάσει τη δομική του πολυπλοκότητα.

4.3.2.2 Ο αλγόριθμος PCIFI

Ο αλγόριθμος PCIFI (protein complex identification by discovering functional interdependence) είναι ακόμη ένας αλγόριθμος που αναγνωρίζει πρωτεϊνικά σύμπλοκα σε δίκτυα αλληλεπίδρασης πρωτεΐνης-πρωτεΐνης λαμβάνοντας υπόψη τόσο τις λειτουργικές σχέσεις μεταξύ πρωτεϊνών όσο και τη δομική τοπολογία του δικτύου. Το PCIFI μπορεί να αναγνωρίσει τα σύμπλοκα πρωτεϊνών με μεγαλύτερη ακρίβεια σε σχέση με αντίστοιχους αλγορίθμους. Ο αλγόριθμος εκτελεί τα επόμενα βήματα:

1. Στο πρώτο βήμα αναπαρίσταται το Πρωτεϊνικό Δίκτυο ως Γράφος και οι μοριακές λειτουργίες κάθε πρωτεΐνης αντιστοιχίζονται στις κορυφές.
2. Σε δεύτερο βήμα, τα αλληλεπιδρώντα ζεύγη πρωτεϊνών, τα οποία είναι λειτουργικά ανεξάρτητα, αφαιρούνται από το Γράφο.
3. Στο τρίτο βήμα, για χρησιμοποιείται το μέτρο Αμοιβαίας Πληροφορίας σε όλα τα εναπομείναντα αλληλεπιδρώντα ζεύγη πρωτεϊνών, για τον προσδιορισμό της ισχύος της σχέσης αλληλεπίδρασης μεταξύ κάθε κορυφής που έχει απομείνει και όλων των γειτόνων της.
4. Για την κανονικοποίηση των τιμών της Αμοιβαίας Εντροπίας μεταξύ δύο κορυφών χρησιμοποιείται το μέτρο της Σύνθετης Εντροπίας, για τον υπολογισμό ενός μέτρου πλεονασμού που αφορά στην λειτουργική αλληλεξάρτηση μεταξύ των κορυφών που έχουν απομείνει. Με βάση αυτό το μέτρο, ο αλγόριθμος εντοπίζει τα σύμπλοκα πρωτεϊνών επιλέγοντας τα ζεύγη πρωτεϊνών που είναι περισσότερο λειτουργικά αλληλεξαρτώμενα.

Όταν η διαδικασία ολοκληρωθεί για κάθε πρωτεΐνη, ο αλγόριθμος αναζητά στους υπόλοιπους Γράφους του δικτύου για συνδεδεμένους Υπογράφους. Οι μεγαλύτεροι Υπογράφοι στη συνέχεια λαμβάνονται ως σύμπλοκα πρωτεϊνών (Lam & Chan, 2012).

4.3.3 Αναγνώριση Βιοδεικτών

Ως βιοδείκτες ασθένειας, αποκαλούνται τα βιομόρια τα οποία παρουσιάζουν διαφοροποίηση στη συγκέντρωσή τους κατά την εκδήλωση και πορεία μιας νόσου. Έχει αποδειχθεί ότι τα Πρωτεϊνικά Δίκτυα θα μπορούσαν να είναι χρήσιμα στην εύρεση βιοδεικτών οι οποίοι συμμετέχουν ή υποδεικνύουν παθογένεια σε νόσους με σύνθετο φαινότυπο, όπως ο καρκίνος και οι καρδιαγγειακές παθήσεις. Ως εκ τούτου, η ταυτοποίηση των κλινικά σημαντικών πρωτεϊνικών βιοδεικτών του φαινοτύπου και της βιολογικής λειτουργίας είναι ένας αναπτυσσόμενος τομέας έρευνας που υπόσχεται να επεκτείνει τις

διαγνωστικές δυνατότητες (Alterovitz, et al., 2008; Berretta & Moscato, 2010; Sechidis et al., 2018).

Συμπληρωματικά με την πρόβλεψη μεμονωμένων πρωτεϊνικών βιοδεικτών, η ανάλυση ενός Πρωτεϊνικού Δικτύου μπορεί να χρησιμοποιηθεί στην πρόγνωση ασθενειών. Η μελέτη των υποδικτύων έχει ιδιαίτερη σημασία καθώς σε αντίθεση με τις μεμονωμένες πρωτεΐνες μπορούν να παρέχουν πληροφορίες ως προς τα μοριακά σύμπλοκα, τις οδούς σηματοδότησης και άλλους μηχανισμούς που επηρεάζουν την έκβαση μιας ασθένειας. Ακόμη, η ανάλυση στο επίπεδο των δικτύων μπορεί να βελτιώσει σημαντικά την κατηγοριοποίηση των ασθενειών (πχ. διαχωρισμός "περιπτώσεων" από τους "ελέγχους", στάδια νόσου) (Ideker & Sharan, 2008).

Οι Teschendorff et al. διερεύνησαν το ρόλο την Εντροπίας Δικτύου στον καρκίνο. Ενσωματώνοντας δεδομένα γονιδιακής έκφρασης με ένα δίκτυο αλληλεπίδρασης πρωτεϊνών, έδειξαν ότι τα καρκινικά κύτταρα χαρακτηρίζονται από την αύξηση της Εντροπίας Δικτύου (συγκριτικά με τον κανονικό φαινότυπο), η οποία είναι τοπικά ορισμένη για τους κόμβους του δικτύου. Ακόμη, διαπιστώθηκε ότι γονίδια τα οποία σχετίζονται με τον πολλαπλασιασμό των κυττάρων σε καρκινικά κύτταρα και τα οποία συχνά κωδικοποιούν ογκογονίδια συνδέονται κυρίως με μειώσεις στην Εντροπία του Δικτύου. Οι ερευνητές επισημαίνουν ότι στην έρευνά αυτή όχι μόνο επαληθεύθηκε η ήδη γνωστή βιολογία του καρκίνου, αλλά με την οπτική που παρέχει η Εντροπία Δικτύου μπορούν να προκύψουν νέες γνώσεις για τα καρκινικά δίκτυα καθώς και νέοι πιθανοί στόχοι δράσης φαρμάκων (Teschendorff et al., 2012).

Οι Chuang et. al. πρότειναν μια μέθοδο εύρεσης δεικτών που βασίζεται σε δεδομένα που εξάγονται από υποδίκτυα Πρωτεϊνικών Δικτύων και όχι σε μεμονωμένα γονίδια. Τα υποδίκτυα που προέκυψαν αποκάλυψαν νέες στοιχεία για τα μονοπάτια που εμπλέκονται στην εξέλιξη και την πρόγνωση της μετάστασης σε καρκίνο του μαστού. Σύμφωνα με τη μέθοδο τους, τα προφίλ γονιδιακής έκφρασης των μεταστατικών και των μη μεταστατικών όγκων ασθενών υπερκαλύπτονται σε ένα δίκτυο αλληλεπίδρασης πρωτεΐνης-πρωτεΐνης. Εντοπίστηκαν, υποδίκτυα των οποίων τα επίπεδα έκφρασης συσχετίζονται με μετάσταση, όπου το επίπεδο έκφρασης ενός υποδικτύου ορίζεται ως συνάρτηση των επιπέδων έκφρασης των γονιδίων μελών του. Διαπιστώθηκε ότι οι δείκτες του υποδικτύου αυτού ήταν περισσότερο αναπαραγώγιμοι σε σχέση με επιμέρους γονίδια-δείκτες που επιλέχθηκαν, χωρίς την πληροφορία για το ίδιο το δίκτυο. Ως αποτέλεσμα, επετεύχθη μεγαλύτερη ακρίβεια στην ταξινόμηση των μεταστατικών έναντι των μη μεταστατικών όγκων. Το κύριο μέτρο που χρησιμοποιήθηκε για την ταξινόμηση των καρκίνων ήταν η Αμοιβαία Πληροφορία. Η Αμοιβαία Πληροφορία χρησιμοποιήθηκε για να συλλάβει την πιθανή ετερογένεια έκφρασης σε ασθενείς με καρκίνο, και να εντοπίσει διαφορές όχι μόνο στον μέσο όρο αλλά και στην διακύμανση της έκφρασης (Chuang, et al., 2007).

Κατά τους Sechidis et al. σε δεδομένα κλινικών δοκιμών, ένας καλός τρόπος επιλογής ενός συνόλου βιοδεικτών είναι η μεγιστοποίηση της Αμοιβαίας Πληροφορίας μεταξύ του στόχου Y και της κοινής τυχαίας μεταβλητής T της θεραπείας και του βέλτιστου συνόλου χαρακτηριστικών X^* . Οι ίδιοι ερευνητές ανέπτυξαν μια νέα μεθοδολογία, την INFO+ χρησιμοποιώντας πολλαπλά μέτρα βασιζόμενα στη Θεωρία της Πληροφορίας για την ταξινόμηση των βιοδεικτών βάσει της ικανότητας πρόβλεψης και πρόγνωσης. Η μέθοδος ελέγχθηκε σε δεδομένα κλινικών δοκιμών σε περιπτώσεις καρκίνου του πνεύμονα και καρδιαγγειακής πάθησης (Sechidis et al., 2018).

Η χρήση βιολογικών δεικτών από εύκολα προσβάσιμα βιορευστά (π.χ. αίμα, ούρα) είναι πλεονεκτική για την αξιολόγηση της κατάστασης ιστών και οργάνων. Τα βιορευστά δεσμεύουν πρωτεΐνες και θραύσματα πρωτεϊνών που απελευθερώνονται από τα κύτταρα στο σώμα, είτε ως απόβλητα είτε για να επικοινωνούν με άλλα κύτταρα ή ιστούς. Επιπλέον, τα βιορευστά είναι πολύ πιο εύκολα προσβάσιμα, σε αντίθεση με τις περισσότερο επεμβατικές ή μη εφαρμόσιμες τεχνικές όπως βιοψίες ιστών (π.χ. εγκεφαλικός ιστός). Ωστόσο, οι υπάρχουσες προσεγγίσεις για την πρόβλεψη βιοδεικτών αναλύουν χωριστά τους ιστούς και τα βιορευστά.

Οι Alterovitz et al. πρότειναν ένα θεωρητικό πλαίσιο για την ανακάλυψη νέων βιολογικών δεικτών που χρησιμοποιεί τη λειτουργική συνέργια μεταξύ βιορευστών και ιστών. Εννοιολογικά, ο ιστός μπορεί να λειτουργήσει ως ένας πομπός πληροφοριών και το βιορευστό ως δέκτης, καθώς επίσης και οι αλληλεπιδράσεις ιστών-βιορευστών μπορούν να θεωρηθούν ως κανάλια πληροφοριών. Έτσι, η πληροφορία που μπορεί να δώσει ένα βιορευστό για έναν ιστό μπορεί να αξιολογηθεί με τη βοήθεια της Θεωρίας της Πληροφορίας. Το πλαίσιο αυτό μπορεί να χρησιμοποιηθεί για τη διερεύνηση της φυσιολογικής λειτουργίας των ιστών και της απόκρισή τους σε φάρμακα και περιπτώσεις ασθένειας. Το κύριο μέτρο που χρησιμοποίησαν στην ανάλυσή τους οι ερευνητές αυτοί είναι η Σχετική Εντροπία. Η Σχετική Εντροπία είναι ένα μέτρο της απόστασης μεταξύ μιας μη παρατηρούμενης κατανομής T , στην περίπτωση αυτή ο ιστός και μιας παρατηρούμενη κατανομή B , όπου στην περίπτωση αυτή είναι το βιορευστό. Η χαμηλότερη Σχετική Εντροπία υποδηλώνει μεγαλύτερη αντιστοιχία μεταξύ των δύο κατανομών. Η Σχετική Εντροπία R μεταξύ ενός ιστού T και ενός βιορευστού B προσδιορίστηκε ως:

$$R(B, T) = \sum_{n=1}^N b(V_n) \log \frac{b(V_n)}{t(V_n)}$$

Όπου, $b(V_n)$ και $t(V_n)$ υποδηλώνουν τη συχνότητα επισήμανσης του όρου V_n στα B και T , αντίστοιχα, και N είναι ο συνολικός αριθμός των όρων που αφορούν στη φυσιολογική λειτουργία, τη νόσο και τα φάρμακα (Alterovitz et al., 2008).

4.3.4 Μελέτη Ευρωστίας Δικτύου

Η ανθεκτικότητα στις "αστοχίες" του δικτύου είναι μια σημαντική ιδιότητα των συστημάτων που αλληλεπιδρούν (interactoms), καθώς η αναδιοργάνωση των αλληλεπιδράσεων αυτών μπορεί να οδηγήσει σε κυτταρικό θάνατο ή ασθένεια. Οι Zitnik et al., μελετώντας δίκτυα αλληλεπίδρασης πρωτεϊνών από 1.840 είδη, διαπίστωσαν ότι τα δίκτυα έχουν την τάση να εξελίσσουν πιο ανθεκτικές αλληλεπιδράσεις. Συγκεκριμένα, έδειξαν ότι μια εξαιρετικά ανθεκτική αλληλεπίδραση έχει ευεργετική επίδραση στην επιβίωση του οργανισμού σε πολύπλοκα, μεταβλητά και ανταγωνιστικά ενδιαιτήματα (Zitnik, et al., 2019).

Οι Manke et al. χρησιμοποίησαν για την ανάλυση δικτύων το Θεώρημα Διακύμανσης, το οποίο προέρχεται από το πεδίο των δυναμικών συστημάτων. Το θεώρημα δηλώνει ότι μακροσκοπικά, η ανθεκτικότητα ενός συστήματος (R), σχετίζεται με το επίπεδο Αβεβαιότητας ή Εντροπίας (S) των υποκείμενων μικροσκοπικών δυναμικών διεργασιών που λαμβάνουν χώρα εντός του συστήματος. Πιο συγκεκριμένα, το $\Delta S \Delta R > 0$ όπου ΔR και ΔS αντιπροσωπεύουν αντίστοιχα τις αλλαγές στην ευρωστία και την Εντροπία του συστήματος. Το θεώρημα εφαρμόστηκε σε δίκτυα αλληλεπίδρασης πρωτεϊνών σε ζυμομύκητες και στο *C. elegans*, και αποδείχθηκε ότι τα γονίδια που συμβάλλουν περισσότερο στην Εντροπία Δικτύου είναι πιο πιθανό να είναι περισσότερο "απαραίτητα" για τον οργανισμό. Το χαρακτηριστικό αυτό ενός γονιδίου που μπορεί να προβλεφθεί με την Εντροπία Δικτύου, σχετίζεται άμεσα με την ευρωστία του συστήματος υπό την αναστολή της έκφρασής του (Manke, et al. 2005)

Σε μια άλλη μελέτη, μετρήθηκε η ανθεκτικότητα ενός Πρωτεϊνικού Δικτύου σε διαταραχές, όσον αφορά τη συμβολή των πρωτεϊνών στην Εντροπία Δικτύου. Οι Manke et al. έδειξαν ότι οι πρωτεΐνες με μεγάλη συμβολή στην Εντροπία Δικτύου συσχετίζονται σε μεγάλο βαθμό με την ευρωστία του δικτύου. Σύμφωνα με την ανάλυσή τους, αυτές οι πρωτεΐνες μπορεί να είναι θανατηφόρες καθώς η τυχαία απομάκρυνσή τους οδήγησε στην αποσύνθεση κυτταρικών δικτύων. Η μελέτη αυτή υπογραμμίζει τη σημασία της Εντροπίας Δικτύου ως θεμελιώδη μεταβλητή ενός δυναμικού δικτύου και ως μέτρο της δομής και των δυναμικών ιδιοτήτων των Γενετικών Δικτύων (Manke, Demetrius, & Vingron, 2006).

4.4 Σηματοδοτικά Δίκτυα

4.4.1 Εισαγωγή

Όλα τα είδη κυττάρων διαθέτουν μηχανισμούς επικοινωνίας οι οποίοι τους επιτρέπουν τόσο την επικοινωνία με άλλα κύτταρα όσο και την αλληλεπίδραση με το περιβάλλον τους. Αν και οι μηχανισμοί αυτοί διαφέρουν σημαντικά ανάμεσα σε μονοκύτταρους και πολυκύτταρους οργανισμούς, υπάρχουν ωστόσο ορισμένες κοινές αρχές. Οι μηχανισμοί της επικοινωνίας διέπονται από α. την *αρχή της εξειδίκευσης*, δηλαδή αναγνωρίζουν το είδος και την ένταση των σημάτων ή ερεθισμάτων που δέχονται και β. περιλαμβάνουν τη δυνατότητα άμεσης (ή και μακροπρόθεσμης) τροποποίησης κυτταρικών λειτουργιών ως απόκριση στο αρχικό σήμα. Χαρακτηριστικά παραδείγματα άμεσης και ταχείας κυτταρικής απόκρισης σε ένα εξωκυττάριο σήμα είναι η αύξηση της ταχύτητας κίνησης των μαστιγίων του *E. coli* παρουσία θρεπτικών συστατικών στο περιβάλλον, αλλά και η έκκριση ινσουλίνης από τα κύτταρα Β των νησιδίων του Langerhans παρουσία γλυκόζης στο αίμα.

Ειδικά για τα κύτταρα ενός πολυκύτταρου οργανισμού η επικοινωνία έχει εξέχουσα σημασία για ένα πλήθος ζωτικών διεργασιών, όπως είναι η ρύθμιση της ανάπτυξης και οργάνωσης σε ιστούς, τη διαφοροποίησή τους και εν τέλη τη συντονισμένη λειτουργία τους στο πλαίσιο του ιστού. Η κυτταρική επικοινωνία στους ζωικούς οργανισμούς μπορεί να ταξινομηθεί με βάση την απόσταση δράσης ως εξής:

1. Κύτταρα που αλληλεπιδρούν από απόσταση, όπου η επικοινωνία γίνεται μέσω χημικών ουσιών που εκκρίνονται από τα κύτταρα και δρουν ως σηματοδοτικά μόρια.
2. Κύτταρα σε γειτονία, όπου η επικοινωνία γίνεται μέσω μεμβρανικών πρωτεϊνών, που δρουν ως ειδικά σημεία πρόσδεσης μεταξύ δυο κυττάρων.
3. Κύτταρα σε επαφή, όπου η επικοινωνία μπορεί να γίνει μέσω χασμοσυνδέσμων (gap junctions), οι οποίοι επιτρέπουν την απευθείας επικοινωνία των κυτοπλασμάτων των κυττάρων, επιτρέποντας την άμεση ανταλλαγή μικρών μορίων.

Ο όρος σηματοδοτική πορεία ή μονοπάτι (signaling pathway) χρησιμοποιείται για την περιγραφή των ενδοκυττάρων βιοχημικών αντιδράσεων που ενεργοποιούνται από τους υποδοχείς. Τα μονοπάτια αυτά καθορίζουν σε μεγάλο βαθμό και την τελική κυτταρική απόκριση. Γενικά, η κυτταρική σηματοδότηση ακολουθεί την πορεία:

ΣΗΜΑ → ΥΠΟΔΟΧΕΑΣ → ΤΕΛΕΣΤΗΣ → ΚΑΤΑΡΡΑΚΤΗΣ ΕΝΔΟΚΥΤΤΑΡΙΩΝ ΣΗΜΑΤΩΝ → ΕΝΙΣΧΥΣΗ ΣΗΜΑΤΟΣ → ΤΕΛΙΚΑ ΠΡΟΪΟΝΤΑ → ΑΠΟΚΡΙΣΗ → ΠΡΟΣΑΡΜΟΓΗ/ΑΠΕΥΑΙΣΘΗΤΟΠΟΙΗΣΗ

Το σήμα (προσδέτης, ligand) συνήθως προσδένεται στον υποδοχέα διεγείροντάς τον, π.χ. στην περίπτωση που ο υποδοχέας είναι ιοντικός διάυλος, η πρόσδεση θα έχει ως αποτέλεσμα την διάνοιξη του διαύλου και τη μεταφορά ιόντων. Τα σήματα μπορεί να είναι:

- α. Μικρά χημικά μόρια με διαφορετικές χημικές ιδιότητες όπως π.χ. πτητικές ενώσεις, λιπόφιλες ενώσεις, παράγωγα λιπαρών οξέων, υδρόφιλα αμινοξέα, νευροδιαβιβαστές, μικρά πεπτίδια, ακόμη και μεγάλες πρωτεΐνες.
- β. Μη χημικής φύσης σήματα όπως η μηχανική πίεση, η θερμοκρασία και η ακτινοβολία. Αέρια ή μικρά λιπόφιλα μόρια, όπως π.χ. οι στεροειδείς ορμόνες (οιστρογόνα, ανδρογόνα, κορτικοστεροειδή), τα οποία εισέρχονται ελεύθερα στα κύτταρα και επομένως μπορούν να δρουν σε ενδοκυττάρια υποδοχείς.

Ένα σήμα μπορεί να αλληλεπιδρά με διαφορετικού τύπου υποδοχείς από τους οποίους εξαρτάται η τελική απόκριση του κυττάρου σε ένα σήμα. Η σύνδεση του σηματοδοτικού μορίου στον υποδοχέα συνήθως πραγματοποιείται με μια βιοχημική αντίδραση εντός του κυττάρου, με άμεσο αποτέλεσμα την ενδοκυττάρια παραγωγή συγκεκριμένων μικρών μορίων που ονομάζονται β' αγγελιοφόροι ή β' μεσολαβητές, (second messengers). Οι β' αγγελιοφόροι είναι μόρια τα οποία μεταδίδουν σήματα ορμονών, αυξητικών παραγόντων και άλλων ουσιών από τους υποδοχείς προς τα στοχευόμενα μόρια μέσα στο κύτταρο. Έχουν την ικανότητα να επιδρούν σε άλλες πρωτεΐνες, ένζυμα ή διαύλους ιόντων, αλλάζοντας τη δράση τους. Στη συνέχεια, ενεργοποιούν έναν καταρράκτη αντιδράσεων (signaling cascade) στο εσωτερικό του κυττάρου, το οποίο μεταφράζεται σε ενίσχυση του αρχικού εξωκυττάρια σήματος. Συχνά αναφέρεται και ο όρος τελεστής (effector) για τα ένζυμα που καθορίζουν την τελική κυτταρική απόκριση.

Οι αντιστρεπτές αντιδράσεις φωσφορυλίωσης και υδρόλυσης του GTP, που καταλύονται από ένζυμα, χρησιμοποιούνται εκτενώς στα σηματοδοτικά μονοπάτια για την άμεση αλλαγή των ιδιοτήτων των πρωτεϊνών στόχων. Σημαντικό ρόλο παίζουν και οι πρωτεΐνες χωρίς ενζυμική ενεργότητα, οι οποίες συνήθως λειτουργούν ως πρωτεΐνες προσαρμογής (adaptors) και καθοδηγούν την οργάνωση σηματοδοτικών πολυπρωτεϊνικών συμπλόκων (signaling complexes). Ακόμη, σημαντική θέση σε αυτούς τους μηχανισμούς έχει μια ομάδα από πρωτεΐνες που μπορούν να αλληλεπιδρούν με πολλές άλλες πρωτεΐνες, οι λεγόμενες ικριωματικές πρωτεΐνες (scaffold proteins).

Η διεπικοινωνία ή διασταυρούμενη επικοινωνία (crosstalk) είναι μια ρυθμιστική αλληλεπίδραση μεταξύ δύο διαφορετικών σηματοδοτικών μονοπατιών και συμβαίνει σε περιπτώσεις που:

α. Ένα συστατικό ενός μονοπατιού ρυθμίζεται από κάποιο συστατικό ενός άλλου.

β. Τα δύο μονοπάτια συγκλίνουν σε έναν κοινό στόχο.

γ. Ο μεσολαβητής ενός μονοπατιού μετατρέπεται σε μεσολαβητή ενός άλλου.

Κάθε κύτταρο αντιλαμβάνεται και αποκρίνεται στη μεταβολή της συγκέντρωσης ενός εξωτερικού σήματος, κάτι το οποίο συμβαίνει για μεγάλο εύρος εντάσεων του σήματος. Στη περίπτωση όμως που το κύτταρο-στόχος εκτεθεί για μεγάλο χρονικό διάστημα στο ερέθισμα, χάνει την ικανότητα να αποκριθεί στην αρχική ένταση του ερεθίσματος. Αυτό το φαινόμενο ονομάζεται προσαρμογή (adaptation) ή απευαισθητοποίηση (desensitization) και υποδηλώνει ότι το κύτταρο μπορεί να προσαρμόζει την ευαισθησία ανίχνευσης των επιπέδων κάθε ερεθίσματος με τρόπο αντιστρεπτό.

Αξίζει ακόμη να σημειωθεί ότι μόνο ορισμένα μονοπάτια σηματοδότησης οδηγούν σε άμεσες μεταβολές κυτταρικών λειτουργιών (έκκριση, χημειοταξία, μυϊκή συστολή, μεταβολή μεμβρανικού δυναμικού), λειτουργούν δηλαδή σε μικρή χρονική κλίμακα. Ωστόσο, σχεδόν όλα τα μονοπάτια μπορούν να οδηγήσουν σε μακροπρόθεσμες μεταβολές των κυτταρικών λειτουργιών μέσω επίδρασης στη γονιδιακή μεταγραφή.

Τα σηματοδοτικά μονοπάτια περιγράφονται ως δίκτυα ευθύγραμμης μετάδοσης πληροφοριών μέσα στο κύτταρο με τελικό στόχο την απόκριση στο αρχικό σήμα. Στην πραγματικότητα, οι ιδιότητες που αναφέρθηκαν παραπάνω, όπως η διασταυρούμενη επικοινωνία, η προσαρμογή, η απευαισθητοποίηση, δίνουν στις σηματοδοτικές πορείες μια εικόνα πολύπλοκου δικτύου με πολλαπλές διακλαδώσεις μεταξύ ορισμένων βασικών γραμμικών οδών που συγκλίνουν σε κεντρικά σημεία και κόμβους. (Berg MJ, Tymoczko LJ, 2011; Lekka, et al., 2015; Lodish et al., 2007)

Τα Σηματοδοτικά Δίκτυα είναι εξαιρετικά δυναμικά, με χρόνους απόκρισης που μπορεί να είναι τάξεις μεγέθους μικρότεροι από τους αντίστοιχους των Ρυθμιστικών και Μεταβολικών Δικτύων και για το λόγο αυτό αποτελούν ίσως την πιο δύσκολη προς μελέτη κατηγορία. Από την άλλη πλευρά η σημασία τους για τη μελέτη της απόκρισης σε εξωτερικά ερεθίσματα και της κυτταρικής επικοινωνίας, λειτουργιών που σχετίζονται στενά με παθολογικές καταστάσεις τα κάνει ιδιαίτερα ενδιαφέροντα κυρίως για τομείς όπως είναι π.χ. η μοριακή ανοσολογία (Νικολάου & Χουβαρδάς, 2015). Σε παθολογικές καταστάσεις, μία ή περισσότερες σηματοδοτικές πορείες αναδιοργανώνονται (ή να αναπροσαρμόζονται) π.χ. έλλειψη ενός σήματος συχνά μπορεί να οδηγήσει σε υπερενεργοποίηση μιας εναλλακτικής σηματοδοτικής πορείας μέσα στα κύτταρα (Lekka et al., 2015; Νικολάου, 2015)

Τα σηματοδοτικά μονοπάτια θεωρούνται ως το κύριο μέσο για τη μετάδοση πληροφοριών, στο οποίο οι χημικοί αγγελιοφόροι μεταδίδουν πληροφορίες σχετικά με το εξωτερικό περιβάλλον προς τα κέντρα αποφάσεων του κυττάρου (Rhee, Cheong, & Levchenko, 2012; Waltermann & Klipp, 2011). Η Θεωρία της Πληροφορίας του Shannon

παρέχει ένα μαθηματικό πλαίσιο που επιτρέπει την ανάλυση των στοιχείων της κυτταρικής σηματοδότησης χωρίς να απαιτείται η λεπτομερής γνώση των Σηματοδοτικών Δικτύων (Brennan & Levchenko, 2012). Στην συνέχεια της ενότητας αυτής θα περιγραφούν κάποιες βασικές εφαρμογές της Θεωρίας της Πληροφορίας στην κυτταρική σηματοδότηση.

4.4.2 Μέτρηση της ποσότητας της μεταφερόμενης πληροφορίας

Η Θεωρία της Πληροφορίας του Shannon παρέχει ένα μαθηματικό πλαίσιο ανεξάρτητο από την πολυπλοκότητα του δικτύου που μπορεί να ποσοτικοποιήσει τις πληροφορίες που μπορούν να μεταδοθούν υπό την παρουσία βιοχημικού θορύβου (Rhee et al., 2012). Σύμφωνα με τους Brennan et al. στα συστήματα της κυτταρικής σηματοδότησης, η πληροφορία ποσοτικοποιείται ως η αβεβαιότητα σχετικά με το περιβάλλον η οποία αφαιρείται από τη δραστηριότητα της σηματοδότησης (η οποία είναι ισοδύναμη με τη γνώση που αποκτά το σύστημα σηματοδότησης). Έτσι η ποσότητα της πληροφορίας εξαρτάται τόσο από τη μεταβλητότητα του περιβάλλοντος (το αρχικό επίπεδο αβεβαιότητας) όσο και από τον θόρυβο που υπάρχει εγγενώς στη διαδικασία σηματοδότησης, επηρεάζοντας την ποσότητα αβεβαιότητας που απομένει (Brennan et al., 2012).

Σύμφωνα με τη Θεωρία της Πληροφορίας, ένα σύστημα σηματοδότησης μπορεί να θεωρηθεί ως μια κατανομή πιθανοτήτων $P(Y|X = x)$, όπου για ένα δεδομένο επίπεδο εισόδου x , λαμβάνεται μια έξοδος Y . Για παράδειγμα, η είσοδος μπορεί να είναι η συγκέντρωση του προσδέτη που ενεργοποιεί έναν υποδοχέα και η έξοδος είναι η δραστηριότητα ενός τελεστή σηματοδότησης, που μπορεί να είναι η συγκέντρωση ενός ενεργοποιημένου μεταγραφικού παράγοντα. Η έξοδος Y , μεταφέρει πληροφορίες σχετικά με το επίπεδο εισόδου x . Η ποσότητα των πληροφοριών που μεταφέρονται εξαρτάται από το ίδιο το σύστημα σηματοδότησης, το οποίο περιλαμβάνει ένα σύνολο χαρακτηριστικών όπως τα επίπεδα θορύβου, η ευαισθησία στις μεταβολές των τιμών εισόδου και η συχνότητα μετάδοσης διαφορετικών τιμών εισόδου. Η Αμοιβαία Πληροφορία, $I(X,Y)$, ποσοτικοποιεί τη μεταφορά πληροφοριών ενός δεδομένου συστήματος σηματοδότησης $P(Y|X = x)$, που συναντά τιμές εισόδου μετά από μια δεδομένη κατανομή, $P(X)$. Οι Jetka et al. μέτρησαν την Αμοιβαία Πληροφορία μεταξύ εισόδου και εξόδου, ως τη μέση μείωση της αβεβαιότητας όσον αφορά την είσοδο που προκύπτει από την παρατήρηση μιας εξόδου (Jetka et al., 2018).

Οι Cheong et al. ανέπτυξαν ένα ολοκληρωμένο θεωρητικό και πειραματικό πλαίσιο, για να προβλέψουν και να μετρήσουν την ποσότητα της πληροφορίας που μεταφέρεται από μοριακά και κυτταρικά δίκτυα. Τα πειραματικά δεδομένα προήλθαν από την ανάλυση της σηματοδότησης του παράγοντα νέκρωσης όγκου (TNF) και μετρώντας την Αμοιβαία

Πληροφορία για ένα ή περισσότερα σηματοδοτικά μονοπάτια, έδειξαν ότι η σηματοδότηση μέσω ενός δικτύου (αντί για ένα μονοπάτι) περιόριζε την απώλεια πληροφοριών λόγω θορύβου (Cheong R. et al., 2011).

Σε μία άλλη μελέτη οι Lenaerts et al. μέτρησαν την ανταλλαγή πληροφοριών σε μονοπάτια σηματοδότησης μέσω του σχηματισμού μακρομοριακών συμπλοκών. Οι συγγραφείς βασίστηκαν στην υπόθεση ότι κάθε σύμπλοκο πρωτεϊνών ABC μπορεί να θεωρηθεί ως θορυβώδες κανάλι επικοινωνίας, στο οποίο ο ρυθμός της μεταγωγής της πληροφορίας μέσω του συμπλόκου μπορεί να προβλεφθεί από τον ρυθμό αποδόμησής του. Η Αμοιβαία Πληροφορία χρησιμοποιήθηκε για να ποσοτικοποιήσει τη μεταφορά πληροφοριών που σχετίζονται με την συνεργιστικότητα (cooperativity), ένα εγγενές χαρακτηριστικό μακρομοριακών συμπλόκων, το οποίο στην μελέτη αυτή σχετίζεται με τον σχηματισμό σηματοδοτικών πολυπρωτεϊνικών συμπλόκων. Η Αμοιβαία Πληροφορία ποσοτικοποιεί το βαθμό συνεργιστικότητας του συστήματος, μέσω της ποσότητας της πληροφορίας που ανταλλάσσεται ανάμεσα στα στοιχεία του συμπλόκου. Ως εκ τούτου, μετράται η Αμοιβαία Πληροφορία μεταξύ ενός ανοδικού προσδέτη A και ενός καθοδικού τελεστή C μέσω μίας ενδιάμεσης πρωτεΐνης B η οποία παρέχει τον δίαυλο επικοινωνίας για την ανταλλαγή πληροφορίας μεταξύ των A και C. Η Αμοιβαία Πληροφορία, μέσω του σχηματισμού του μακρομοριακού συμπλόκου ABC μπορεί να είναι θεωρηθεί ως ένα εναλλακτικό μέτρο για το βαθμό συνεργιστικότητας στη διαδικασία δημιουργίας πρωτεϊνικών συμπλόκων (Lenaerts et al., 2009).

Η συνδυαστική ανάλυση ομικών δεδομένων μπορεί να βοηθήσει στη καλύτερη κατανόηση λειτουργιών και την ταυτοποίηση βιοχημικών μονοπατιών. Οι Locasale και Wolf-Yadlin χρησιμοποίησαν την Αρχή της Μέγιστης Εντροπίας για τον προσδιορισμό ενός δικτύου αλληλεπιδράσεων μεταξύ θέσεων φωσφορυλίωσης, λαμβάνοντας υπόψη μόνο τους περιορισμούς που προκύπτουν από τις συσχετίσεις ανά ζεύγη που παρατηρούνται σε ποσοτικά δεδομένα φασματομετρίας μάζας. Ως αποτέλεσμα, κατάφεραν να δημιουργήσουν από πρωτεομικά δεδομένα φασματομετρίας μάζας ένα βιολογικά ερμηνεύσιμο Δίκτυο Σηματοδότησης το οποίο προβλέπει τις αλληλεπιδράσεις που εμπλέκονται σε προηγουμένως μη χαρακτηρισμένους κόμβους σηματοδότησης. Τα ευρήματά της μελέτης αυτής υποδεικνύουν ότι τα μοντέλα Δικτύων Μέγιστης Εντροπίας είναι ένα σημαντικό εργαλείο για την ερμηνεία ποσοτικών δεδομένων Πρωτεομικής (Locasale & Wolf-Yadlin, 2009).

4.4.3 Εκτίμηση της Χωρητικότητας Καναλιού στη μετάδοση σήματος

Στις μοριακές επικοινωνίες, το μέσο στο οποίο μεταδίδονται σηματοδοτικά μόρια από τους πομπούς προς τους δέκτες μπορεί να θεωρηθεί ως *κανάλι επικοινωνίας*. Σύμφωνα

με τη Θεωρία Πληροφορίας του Shannon, ο μέγιστος ρυθμός με τον οποίο μπορούν να μεταδοθούν οι πληροφορίες μέσω ενός καναλιού επικοινωνίας αναφέρεται ως *Χωρητικότητα Καναλιού* (§2.7) και μπορεί να χρησιμοποιηθεί για τη μέτρηση του μέγιστου αριθμού ερεθισμάτων που μπορεί να διακρίνει ένα κύτταρο, με βάση τις θορυβώδεις αποκρίσεις των συστημάτων σηματοδότησής του. Οι Cheong et al. θεωρώντας τα συστήματα σηματοδότησης ως θεωρητικά κανάλια επικοινωνίας, έδειξαν ποσοτικά ότι ο θόρυβος σε ένα και μόνο κύτταρο μπορεί να περιορίσει σημαντικά την ποσότητα μεταδιδόμενων πληροφοριών σχετικά με την ένταση της εισόδου, ιδιαίτερα μέσα σε μεμονωμένες διαδρομές σηματοδότησης (Cheong, et al., 2011).

Οι Liu και Nakano μοντελοποίησαν τον μηχανισμό σηματοδότησης των δισθενών κατιόντων ασβεστίου (Ca^{2+}) τα οποία θεωρούνται ως ενδοκυττάριοι β' αγγελιοφόροι που εμπλέκονται στα Σηματοδοτικά Δίκτυα. Ο μηχανισμός τους θεωρείται ένας από τους βασικότερους μηχανισμούς σηματοδότησης, καθώς είναι γνωστό ότι ρυθμίζει διάφορες βιολογικές λειτουργίες όπως η διαφοροποίηση, ο πολλαπλασιασμός και η γονιμοποίηση. Η Χωρητικότητα Καναλιού στη μετάδοση σήματος στην περίπτωση αυτή, μπορεί να εκτιμηθεί με βάση το μέγιστο αριθμό ενεργοποιημένων β' αγγελιοφόρων στο Σηματοδοτικό Δίκτυο (Liu & Nakano, 2007).

Σε μία άλλη μελέτη, ο Horozov έδειξε ότι η Χωρητικότητα Καναλιού της οδού σηματοδότησης από τη συγκέντρωση των παραγόντων μεταγραφής (TFs) στο επίπεδο έκφρασης, θα μπορούσε να ενισχυθεί σε μια συγκεκριμένη κατάσταση θορύβου. Βασική υπόθεση ήταν ότι το Ρυθμιστικό Δίκτυο έχει σχεδιαστεί για τη βελτιστοποίηση της μεταφοράς πληροφοριών από την είσοδο (συγκεντρώσεις TF) στην έξοδο (επίπεδο γονιδιακής έκφρασης). Οι πληροφορίες στην είσοδο τυπικά κωδικοποιούνται ως μια σταθερή συγκέντρωση ενός TF, c , που δεσμεύεται στη θέση του υποκινητή του επιθυμητού γονιδίου απόκρισης και ενισχύει ή αναστέλλει τη μεταγραφή του. Σε μοριακό επίπεδο, η διαδικασία δέσμευσης υπόκειται εγγενώς σε θόρυβο ο οποίος αναπαρίσταται μέσω μιας πιθανοτικής σχέσης μεταξύ της συγκέντρωσης του TF, c , και του επιπέδου γονιδιακής έκφρασης g , ως $P(g/c)$, όπου η λεπτομερής μορφή της σχέσης εξαρτάται από φυσικές παραμέτρους. Η διαδικασία αυτή μπορεί να παρομοιαστεί με την επικοινωνία μέσω ενός θορυβώδους καναλιού. Έτσι, γνωρίζοντας τη φύση του θορύβου ενός καναλιού $P(g/c)$, είναι εφικτό να υπολογιστεί η πιθανότητα κατανομής του σήματος εισόδου $P_{TF}^*(c)$ που μεγιστοποιεί το ρυθμό μετάδοσης σήματος. Ουσιαστικά, η κατανομή αυτή πληροφορεί τον πομπό για το πόσο συχνά πρέπει να χρησιμοποιείται μια συγκεκριμένη συγκέντρωση TF για τη βέλτιστη μετάδοση πληροφοριών, οι οποίες κωδικοποιούνται υπό μορφή συγκέντρωσης. Τα αποτελέσματα της μελέτης μπορούν να συνοψιστούν ως εξής:

- α. Η αλληλεπίδραση στο επίπεδο της δέσμευσης (παρεμπόδιση-interference) σχετίζεται με το ανοδικό δίκτυο των μεταγραφικών παραγόντων (TFs) που αλληλεπιδρούν μεταξύ τους.

- β. Ένα μονοπάτι σηματοδότησης με παρεμβολή, δηλαδή συσχετισμούς όσον αφορά στον θόρυβο εισόδου λόγω μικροσκοπικών αλληλεπιδράσεων των μορίων σηματοδότησης, μπορούν να βελτιστοποιήσουν τη μετάδοση πληροφοριών με την εφαρμογή διασταυρούμενης ρύθμισης μεταξύ των αλληλεπιδρώντων μορίων, έτσι ώστε η συγκέντρωση μιας εισόδου να εξαρτάται από την άλλη.
- γ. Σε σύγκριση με τους μη αλληλεπιδρώντες μεταγραφικούς παράγοντες, οι μεταγραφικοί παράγοντες που αλληλεπιδρούν σε επικαλυπτόμενες θέσεις δέσμευσης και οι οποίοι εμφανίζουν ανοδική διασταυρούμενη ρύθμιση (upstream cross-regulation) μπορούν να ενισχύσουν τη μετάδοση πληροφοριών (Hormoz, 2013).

4.4.4 Μελέτη της κατευθυντικής απόκρισης χημειοτακτικών κυττάρων

Η χημειοταξία είναι ένας μηχανισμός ο οποίος εμφανίζεται τόσο σε μονοκύτταρους όσο και σε πολυκύτταρους οργανισμούς, οι οποίοι κινούνται ως απόκριση σε κάποια ετερογένεια της συγκέντρωσης μορίων στο εξωκυττάριο περιβάλλον τους. Διαφορές συγκεντρώσεων του είδους αυτού χαρακτηρίζονται από τον όρο *Κλίση* (gradient). Ενδεικτικά η χημειοταξία παίζει σημαντικό ρόλο σε διάφορες βιολογικές διαδικασίες όπως η ανάπτυξη, η ανοσολογική λειτουργία, η αποκατάσταση ιστών, η μετάσταση και η φλεγμονή (Iglesias & Devreotes, 2008). Μια κοινή πρόκληση των συστημάτων κυτταρικής σηματοδότησης είναι να λάβουν αποφάσεις οι οποίες βασίζονται σε θορυβώδη χημικά σήματα τα οποία προσφέρουν ελλιπείς πληροφορίες για το περιβάλλον τους. Για την παρακολούθηση των χημικών κλίσεων, τα χημειοτακτικά κύτταρα επιτελούν ορισμένες εργασίες, κοινές σε όλες τις χημειοτακτικές συμπεριφορές όπως η ανίχνευση κλίσης, η επεξεργασία σήματος και ο έλεγχος κινητικού συστήματος.

Σε αρκετές μελέτες έχουν διερευνηθεί οι δυνατότητες επεξεργασίας της πληροφορίας σε χημειοτακτικά κύτταρα με τη βοήθεια θεωρητικών προσεγγίσεων που έχουν ως βάση τη Θεωρία της Πληροφορίας (Andrews & Iglesias, 2007; Fuller et al., 2010; Hu et al., 2011). Οι Andrews et al. μελέτησαν την αποτελεσματικότητα των στρατηγικών της χημειοταξίας ώστε να εντοπίσουν τη θέση μιας εξωτερικής χημειοελκυστικής ουσίας, υπό την παρουσία βιοχημικού θορύβου, με τη βοήθεια της Θεωρίας Ρυθμού-Παραμόρφωσης (Rate-Distortion Theory) (§2.8). Το σύστημα ανίχνευσης της κλίσης αφού ανιχνεύσει ένα σήμα ερεθίσματος, θα επεξεργαστεί από το Δίκτυο Σηματοδότησης υποδεικνύοντας την κατεύθυνση της κίνησης του κυττάρου. Η Θεωρία Ρυθμού-Παραμόρφωσης παρέχει τα όρια στο ρυθμό με τον οποίο μεταδίδεται η πληροφορία για να επιτευχθεί ένα συγκεκριμένο κριτήριο απόδοσης. Στην περίπτωση αυτή, βρίσκει τη στρατηγική ανίχνευσης κλίσης η οποία ελαχιστοποιεί την Αμοιβαία Πληροφορία μεταξύ του ερεθίσματος και του

συστήματος απόκρισης. Η βέλτιστη στρατηγική χρησιμοποιείται ως σημείο αναφοράς προκειμένου να συγκριθούν οι αποκρίσεις που παρατηρούνται σε χημειοτακτικά κύτταρα (Andrews & Iglesias, 2007).

Σε μια άλλη μελέτη οι Hu et al. χρησιμοποίησαν την Αμοιβαία Πληροφορία προκειμένου να μελετήσουν το ρόλο των στοχαστικών διακυμάνσεων που επηρεάζουν την χημειοταξία των ευκαρυωτικών οργανισμών και να εντοπίσουν τα φυσικά όρια για την ανίχνευση της κλίσης και των βέλτιστων στρατηγικών χημειοταξίας. Τα βήματα που ακολουθήθηκαν είναι τα εξής:

1. Αρχικά, υπολογίστηκε η εξωτερική Αμοιβαία Πληροφορία για την ανίχνευση της κλίσης στην επιφάνεια των κυττάρων. Η ποσότητα αυτή αποτελεί το ανώτερο όριο των πληροφοριών που μπορούν να μεταδοθούν στο ενδοκυτταρικό σύστημα λήψης αποφάσεων.
2. Έπειτα υπολογίστηκε η Αμοιβαία Πληροφορία χημειοτακτισμού, η οποία έχει οριστεί ως η συνολική πληροφορία που αποκτήθηκε από ένα χημειοτακτικό κύτταρο.
3. Στη συνέχεια γίνεται σύγκριση της εξωτερικής και της χημειοτακτικής Αμοιβαίας Πληροφορίας προκειμένου να καθοριστεί η σχέση μεταξύ των εσωτερικών διεργασιών σηματοδότησης και της χημειοτακτικής απόδοσης.

Τα αποτελέσματα της μελέτης αυτής έδειξαν ότι μια μεγάλη ποσότητα πληροφορίας σχετική με την κλίση μπορεί να χαθεί, εξαιτίας είτε του βιοχημικού θορύβου, είτε λόγω της πολύπλοκης διαδικασίας επεξεργασίας σήματος (Hu et al., 2011).

Σε μία άλλη μελέτη, οι Yu et al. μελέτησαν τη σηματοδότηση φερομόνης σε ζυμομύκητες. Ο απλοειδής ευκαρυωτικός οργανισμός *Saccharomyces cerevisiae* επιλέγει τα πιθανά κύτταρα για ζευγάρωμα ανιχνεύοντας την εξωκυτταρική συγκέντρωση μιας εκκρινόμενης φερομόνης που παράγεται από άλλα κύτταρα ζύμης. Η ικανότητα των κυττάρων να αντιδρούν διαφορετικά σε διαφορετικές συγκεντρώσεις φερομόνης, εξαρτάται από την ποσότητα της πληροφορίας που μπορεί να μεταδώσει το σύστημα για τη συγκέντρωση της φερομόνης. Για τη διερεύνηση της σχέσης μεταξύ της δόσης (συγκέντρωση του εκκρινόμενου μορίου πρόσδεσης και δέσμευση υποδοχέα) και της απόκρισης (καθοδικοί ενδοκυτταρικοί μηχανισμοί που ανταποκρίνονται στη δέσμευση του μορίου πρόσδεσης στον υποδοχέα) χρησιμοποιήθηκε η Αμοιβαία Πληροφορία. Η ανάλυση έδειξε ότι η ευθυγράμμιση της σχέσης δόσης-απόκρισης έχει ως αποτέλεσμα την καλύτερη μετάδοση πληροφοριών, προς το κύτταρο δέκτη σχετικά με κύτταρα με τα οποία (δέκτης) μπορεί να εκτελέσει αμφιγονική αναπαραγωγή (Mian & Rose, 2011; Yu et al., 2008).

Λόγω της βιοχημικής φύσης των Σηματοδοτικών Δικτύων, ο μοριακός θόρυβος θα περιορίσει αναπόφευκτα την πιστότητα οποιωνδήποτε μηνυμάτων που λαμβάνονται και

επεξεργάζονται από τα δίκτυα μεταγωγής σημάτων ενός κυττάρου (Rhee et al., 2012). Η πιστότητα της μετάδοσης πληροφοριών είναι ιδιότητα της δομής του καναλιού επικοινωνίας και καθορίζεται σημαντικά από τη Χωρητικότητα του Καναλιού (Brennan et al., 2012). Η Αμοιβαία Πληροφορία έχει χρησιμοποιηθεί για να ποσοτικοποιήσει την πιστότητα ενός σήματος ως τον μέγιστο αριθμό τιμών εισόδου που ένα κύτταρο μπορεί να επεξεργαστεί, υπό την παρουσία θορύβου (Cheong et al., 2011).

Την σημασία της πιστότητας στη μεταφορά της πληροφορίας εξέτασαν οι Tostevin & Wolde. Οι ερευνητές μελέτησαν το πόσο αξιόπιστα το δίκτυο χημειοτακτισμού του *E. Coli* μπορεί να μεταδώσει πληροφορίες σχετικά με τη συγκέντρωση των μορίων πρόσδεσης που βρίσκονται στο μαστίγιο και να βρει το φάσμα ισχύος του σήματος εισόδου που μεγιστοποιεί το ρυθμό μετάδοσης πληροφοριών (Tostevin & Wolde, 2009).

Κεφάλαιο 4. Επίλογος

Η Βιολογία Συστημάτων (η κατανόηση δηλαδή σε επίπεδο συστήματος της βιολογικής επιστήμης), από φυσικοχημικής άποψης, ασχολείται με την μη αναστρέψιμη Θερμοδυναμική των ανοιχτών συστημάτων που έχουν θεμελιωθεί από το πεδίο της Στατιστικής Μηχανικής εκτός ισορροπίας. Ένα από τα πιο σημαντικά χαρακτηριστικά των έμβιων συστημάτων είναι η ικανότητά τους να επεξεργάζονται, να ανταλλάσσουν και να αποθηκεύουν μεγάλες ποσότητες πληροφοριών σε πολλαπλά επίπεδα της οργάνωσής τους. Οι έμβιοι οργανισμοί αποτελούν ανοιχτά συστήματα που ανταλλάσσουν συνεχώς ύλη και την ενέργεια με το περιβάλλον τους υπακούοντας στους νόμους της Θερμοδυναμικής. Η σχέση μεταξύ Θερμοδυναμικής και Πληροφορίας έχει κατανοηθεί σε μεγάλο βαθμό τα τελευταία χρόνια, προσελκύοντας όλο και περισσότερους ερευνητές που εφαρμόζουν τη νέα γνώση σε διάφορους τομείς της επιστήμης επιτυγχάνοντας σημαντική πρόοδο. Το θεωρητικό πλαίσιο που παρέχει η Θεωρία της Πληροφορίας του Shannon μπορεί να προσεγγίσει ένα ευρύ φάσμα προβλημάτων.

Η Βιολογία Συστημάτων μελετά τα βιολογικά συστατικά σε επίπεδο συστήματος και η Θεωρία της Πληροφορίας παρέχει τα θεωρητικό πλαίσιο για τη μελέτη των σχέσεων μεταξύ των συστατικών αυτών. Η Βιολογία Συστημάτων είναι ένας αναδυόμενος τομέας που στοχεύει στην κατανόηση των πολύπλοκων βιολογικών συστημάτων με τη χρήση της υπολογιστική και μαθηματική μοντελοποίηση. Ωστόσο, ο τομέας είναι γεμάτος ανοιχτά προβλήματα και προκλήσεις τις οποίες η Θεωρία της Πληροφορίας έχει τη δυνατότητα να διαμορφώσει θεμελιωδώς. Η παρούσα διατριβή αναδεικνύει την ευρεία εφαρμοσιμότητα

της Θεωρία της Πληροφορίας σε προβλήματα της Βιολογίας Συστημάτων. Συνοψίζει ορισμένες από τις πιο πρόσφατες εφαρμογές της Θεωρίας της Πληροφορίας που επικεντρώνονται στα Βιολογικά Δίκτυα.

Η Βιολογία Συστημάτων δεν θα είχε σημειώσει πρόοδο χωρίς την ταυτόχρονη εξέλιξη των Ομικών τεχνολογιών. Οι τεχνολογίες αυτές έχουν συμβάλουν στον εντοπισμό μεγάλων ποσοτήτων γονιδίων, μεταγραφικών παραγόντων, πρωτεϊνών, μεταβολιτών κ.α. Ταυτόχρονα, η ανάλυση στο επίπεδο του δικτύου έχει συμβάλει σημαντικά στην κατανόηση των βιολογικών συστημάτων. Στο πλαίσιο αυτό, η ανάλυση δικτύων θεωρείται κατάλληλη για την διαχείριση αντιμετώπιση της πολυπλοκότητας και αποσκοπεί στην ανίχνευση αρχών και μοτίβων στα βιολογικά συστήματα, παραβλέποντας σκόπιμα π.χ. μηχανιστικές βιοχημικές λεπτομέρειες.

Η μοντελοποίηση μεταγραφικών αλληλεπιδράσεων από μεγάλα δίκτυα αλληλεπιδρώντων στοιχείων έχει αποτελέσει σημαντικό ζήτημα για τη Βιολογία Συστημάτων. Για την ανακατασκευή αυτών των γενετικών συστημάτων ελέγχου είναι απαραίτητο να κατανοηθεί πλήρως η ροή της αιτιότητας στα Γενετικά Ρυθμιστικά Δίκτυα. Μια τέτοια κατανόηση απαιτεί μια διεπιστημονική σύνθεση θεωρίας και πειραμάτων, υπολογιστικών μεθόδων και τεχνολογιών υψηλής απόδοσης. Οι αλγόριθμοι για τα Γενετικά Ρυθμιστικά Δίκτυα μπορούν να προβλέπουν και να αναπαριστούν δίκτυα γονιδίων από χιλιάδες δεδομένα γονιδιακής έκφρασης με ολοένα και περισσότερες εφαρμογές στη βιοϊατρική έρευνα. Στην παρούσα ανασκόπηση παρουσιάστηκαν οι πιο διαδεδομένες μέθοδοι επαγωγής σε Γενετικά Ρυθμιστικά Δίκτυα που βασίζονται στη Θεωρία της Πληροφορίας, REVEAL, ARACNE, CLR, MRNET, C3NET. Το βασικό μέτρο που χρησιμοποιούν οι αλγόριθμοι αυτοί προκειμένου να προβλέψουν τις ρυθμιστικές σχέσεις μεταξύ γονιδίων είναι η Αμοιβαία Πληροφορία. Κύριο πλεονέκτημα των μεθόδων που χρησιμοποιούν την Αμοιβαία Πληροφορία είναι η ταχύτητα και ακρίβεια υπολογισμών, δεδομένου ότι μπορούν να λάβουν έως και χιλιάδες μεταβλητές από σχετικά περιορισμένο αριθμό δειγμάτων.

Για την ανάλυση του κυτταρικού μεταβολισμού έχουν αναπτυχθεί σημαντικά υπολογιστικά εργαλεία που προσεγγίζουν προβλήματα όπως ο προσδιορισμός της δραστηριότητας μεταβολικών ροών, η αντίστροφη μηχανική Μεταβολικών Δικτύων και η μελέτη ευρωστίας δικτύου. Περιγράφηκαν οι αλγόριθμοι ADEMA MIRA MIDER οι οποίοι δέχονται ως δεδομένο εισόδου τις ποσότητες των συγκεντρώσεων των μεταβολιτών που συμμετέχουν σε ένα Μεταβολικό Δίκτυο. Ως βασικό μέτρο χρησιμοποιείται η Αμοιβαία Πληροφορία και η Υπό Συνθήκη Εντροπία, για τον προσδιορισμό των μη-γραμμικών σχέσεων μεταξύ πολλαπλών μεταβολιτών. Συζητήθηκαν ακόμη μελέτες που έχουν ως αντικείμενο την εξέλιξη Μεταβολικών Δικτύων καθώς και την ανάλυση της πολυπλοκότητας και της ευρωστία δικτύων. Οι μέθοδοι ανάλυσης που προτάθηκαν βασίζονται σε μέτρα όπως η Εντροπία Kolmogorov-Sinai και η Αρχή Μέγιστης Εντροπίας.

Θεωρητικές προσεγγίσεις συζητήθηκαν, επίσης, για διάφορα προβλήματα που σχετίζονται με τα Πρωτεϊνικά Δίκτυα και έχουν μελετηθεί στη βιβλιογραφία με τη βοήθεια μεθοδολογιών από την Θεωρία της Πληροφορίας. Τα προβλήματα αυτά εστιάζουν σε τομείς όπως η αναγνώριση πρωτεϊνικών συμπλόκων που εμπλέκονται σε οδούς ασθενειών, η αναγνώριση βιοδεικτών που συνδέονται με ασθένειες και η μελέτη ιδιοτήτων του δικτύου, όπως η ευρωστία και η πολυπλοκότητα και η σχέση τους με καταστάσεις ασθένειας. Στο πλαίσιο αυτό σχολιάστηκαν οι εφαρμογές της Εντροπίας Γράφου και του αλγόριθμου PCIFI στην αναγνώριση πρωτεϊνικών συμπλόκων σε δίκτυα αλληλεπίδρασης πρωτεΐνης-πρωτεΐνης. Μεθοδολογίες με βάση την Εντροπία Δικτύου την Αμοιβαία Πληροφορία και τη Σχετική Εντροπία χρησιμοποιήθηκαν για να προσεγγίσουν το πρόβλημα της αναγνώρισης βιοδεικτών σε διάφορους τύπους ασθενειών. Επιπλέον, η συμβολή της Εντροπίας Δικτύου συζητήθηκε στο πλαίσιο της ανάλυσης της ανθεκτικότητας Πρωτεϊνικών Δικτύων.

Τα Σηματοδοτικά Δίκτυα περιγράφουν τις διεργασίες της κυτταρικής σηματοδότησης μέσω της ενεργοποίησης ή καταστολής πρωτεϊνών και ενζύμων. Αρκετές από τις υπάρχουσες μελέτες μετρούν το ρυθμό μετάδοσης πληροφοριών σε Δίκτυα Σηματοδότησης με τη χρήση της Θεωρίας της Πληροφορίας. Στις περιπτώσεις αυτές καθώς και σε περιπτώσεις που διερευνάται η αξιοπιστία μεταβίβασης πληροφοριών το βασικό μέτρο που έχει χρησιμοποιηθεί είναι η Αμοιβαία Πληροφορία. Η Χωρητικότητα Καναλιού μπορεί να χρησιμοποιηθεί για την κατανόηση της μετάδοσης σημάτων καθώς πρακτικά βρίσκει το μέγιστο αριθμό ερεθισμάτων που μπορεί να διακρίνει ένα κύτταρο, λαμβάνοντας υπόψη τις θορυβώδεις αποκρίσεις των συστημάτων σηματοδότησης. Άλλες μελέτες που στηρίζονται σε θεωρητικές προσεγγίσεις της Θεωρίας της Πληροφορίας αναλύουν χημειοτακτικές αποκρίσεις σε διάφορα ερεθίσματα. Πολλές από αυτές στοχεύουν στον προσδιορισμό των βέλτιστων στρατηγικών της χημειοταξίας και το κύριο μέτρο που χρησιμοποιούν για τις στις αναλύσεις είναι η Αμοιβαία Πληροφορία.

Στην παρούσα διατριβή παρουσιάστηκαν μελέτες που διερευνούν τις δυνατότητες στατιστικών εργαλείων της Θεωρίας της Πληροφορίας για την ανάλυση και την ερμηνεία πολύπλοκων βιολογικών συστημάτων. Οι θεωρητικές μέθοδοι με βάση τη Θεωρία της Πληροφορίας έχουν αποδειχθεί ως πολύτιμα εργαλεία για την προσέγγιση ενός μεγάλου αριθμού βιολογικών προβλημάτων. Σε μία εποχή που τα τεχνολογικά μέσα συνεχώς βελτιώνονται, είναι επόμενο ότι τα βιολογικά δεδομένα θα συνεχίσουν να γίνονται όλο και πιο περισσότερα, με υψηλότερη ποιότητα και ολόένα πιο ποικίλα, η ανάγκη για νέες βελτιωμένες τεχνικές και μεθοδολογίες πιθανόν να κλιμακωθεί. Οι εφαρμογές της Θεωρίας της Πληροφορίας έχουν αυξηθεί διαδοχικά τα τελευταία χρόνια, ενώ παρατηρείται συνεχώς βελτίωση των υπάρχοντων τεχνικών για τις ανάγκες της εκάστοτε μελέτης. Το πεδίο αυτό αναμένεται να εξαπλωθεί περαιτέρω τα επόμενα χρόνια με νέες κατευθύνσεις για εφαρμογές στη Βιολογία και την Ιατρική.

Βιβλιογραφία

- A.A. Margolin, I. Nemenman, K. Basso, C. Wiggins, G. Stolovitzky, R. Dalla Favera, A. C. (2006). ARACNE: An Algorithm for the Reconstruction of Gene Regulatory Networks in a Mammalian Cellular Context. *BMC Bioinformatics*. <https://doi.org/10.1186/1471-2105-7-S1-S7>
- Agresti, A. (2007). *An Introduction to Categorical Data Analysis*. Hoboken, NJ, USA: John Wiley & Sons, Inc. <https://doi.org/10.1002/0470114754>
- Albert, R., Jeong, H., & Barabási, A.-L. (2000). Error and attack tolerance of complex networks. *Nature*, *406*, 378. Retrieved from <https://doi.org/10.1038/35019019>
- Altay, G., & Emmert-Streib, F. (2011). Structural influence of gene networks on their inference: Analysis of C3NET. *Biology Direct*, *6*(Mi), 1–16. <https://doi.org/10.1186/1745-6150-6-31>
- Alterovitz, G., Xiang, M., Liu, J., Chang, A., & Ramoni, M. F. (2008). System-wide peripheral biomarker discovery using information theory. *Pacific Symposium on Biocomputing. Pacific Symposium on Biocomputing*, *242*, 231–242. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/18229689>
- Andrews, B. W., & Iglesias, P. A. (2007). An information-theoretic characterization of the optimal gradient sensing response of cells. *PLoS Computational Biology*, *3*(8), 1489–1497. <https://doi.org/10.1371/journal.pcbi.0030153>
- Aslam, B., Basit, M., Nisar, M. A., Khurshid, M., & Rasool, M. H. (2017). Proteomics: Technologies and Their Applications. *Journal of Chromatographic Science*, *55*(2), 182–196. <https://doi.org/10.1093/chromsci/bmw167>
- Bar-Yam, Y. (1999). *Dynamics of Complex Systems (Studies in Nonlinearity)*. CRC Press.
- Basso, K., Margolin, A. A., Stolovitzky, G., Klein, U., Dalla-Favera, R., & Califano, A. (2005). Reverse engineering of regulatory networks in human B cells. *Nature Genetics*, *37*(4), 382–390. <https://doi.org/10.1038/ng1532>
- Berg MJ, Tymoczko LJ, S. L. (2011). Βιοχημεία. In *Τόμος Ι. Πανεπιστημιακές Εκδόσεις Κρήτης*.
- Berretta, R., & Moscato, P. (2010). Cancer biomarker discovery: The entropic hallmark. *PLoS ONE*, *5*(8). <https://doi.org/10.1371/journal.pone.0012262>
- Boeing, G. (2016). Visual Analysis of Nonlinear Dynamical Systems: Chaos, Fractals, Self-Similarity and the Limits of Prediction. *Systems*, *4*(4), 37. <https://doi.org/10.3390/systems4040037>
- Brennan, M. D., Cheong, R., & Levchenko, A. (2012). How Information Theory Handles Cell Signaling and Uncertainty. *Science*, *338*(6105), 334 LP – 335. <https://doi.org/10.1126/science.1227946>
- Brooks, D. R., Leblond, P. H., & Cumming, D. D. (1984). Information and entropy in a simple evolution model. *Journal of Theoretical Biology*, *109*(1), 77–93. [https://doi.org/10.1016/S0022-5193\(84\)80112-5](https://doi.org/10.1016/S0022-5193(84)80112-5)
- Butte, A. J., & Kohane, I. S. (1999). Mutual Information Relevance Networks: Functional Genomic Clustering Using Pairwise Entropy Measurements. *Biocomputing 2000*, *426*, 418–429. https://doi.org/10.1142/9789814447331_0040
- Chalancon, G., Kruse, K., & Babu, M. M. (2013). Metabolic Networks, Structure and Dynamics. In W. Dubitzky, O. Wolkenhauer, K.-H. Cho, & H. Yokota (Eds.), *Encyclopedia of Systems Biology* (pp. 1263–1267). New York, NY: Springer New York. https://doi.org/10.1007/978-1-4419-9863-7_561
- Chuang, H. Y., Lee, E., Liu, Y. T., Lee, D., & Ideker, T. (2007). Network-based classification of

- breast cancer metastasis. *Molecular Systems Biology*, 3(140), 1–10.
<https://doi.org/10.1038/msb4100180>
- Cicek, A. E., Bederman, I., Henderson, L., Drumm, M. L., & Ozsoyoglu, G. (2013). ADEMA: An Algorithm to Determine Expected Metabolite Level Alterations Using Mutual Information. *PLoS Computational Biology*, 9(1).
<https://doi.org/10.1371/journal.pcbi.1002859>
- Cicek, A. E., Roeder, K., & Ozsoyoglu, G. (2015). MIRA: Mutual information-based reporter algorithm for metabolic networks. *Bioinformatics*, 31(7), 1160.
<https://doi.org/10.1093/bioinformatics/btv081>
- Costanzo, M., Kuzmin, E., van Leeuwen, J., Mair, B., Moffat, J., Boone, C., & Andrews, B. (2019). Global Genetic Networks and the Genotype-to-Phenotype Relationship. *Cell*, 177(1), 85–100. <https://doi.org/10.1016/j.cell.2019.01.033>
- Davidson, E., & Levin, M. (2005). Gene regulatory networks. *Proceedings of the National Academy of Sciences of the United States of America*, 102(14), 4935.
<https://doi.org/10.1073/pnas.0502024102>
- Davis, N., Favero, T., de Hoog, C., & others. (1995). Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science*.
- De Badyn, M. H., Chapman, A., & Mesbahi, M. (2015). Network entropy: A system-theoretic perspective. *Proceedings of the IEEE Conference on Decision and Control, 54rd IEEE(Cdc)*, 5512–5517. <https://doi.org/10.1109/CDC.2015.7403083>
- de la Fuente, I. M. (2010). *Quantitative analysis of cellular metabolic dissipative, self-organized structures*. *International Journal of Molecular Sciences* (Vol. 11).
<https://doi.org/10.3390/ijms11093540>
- Demetrius, L., & Manke, T. (2005). Robustness and network evolution - An entropic principle. *Physica A: Statistical Mechanics and Its Applications*, 346(3–4), 682–696.
<https://doi.org/10.1016/j.physa.2004.07.011>
- Dittrich, T. (2015). “The concept of information in physics”: An interdisciplinary topical lecture. *European Journal of Physics*, 36(1). <https://doi.org/10.1088/0143-0807/36/1/015010>
- Dolev, S., & Elitzur, A. C. (2000). Biology and Thermodynamics: Seemingly-Opposite Phenomena in Search of a Unified Paradigm. *Journal of Biology*, 15, 16. Retrieved from <http://arxiv.org/abs/physics/0012036>
- Eddington, A. (1948). *The Nature of the Physical World: Gifford Lectures (1927)* (Electronic). Retrieved from <http://books.google.ca/books?id=DDohLwEACAAJ>
- Faith, J. J., Hayete, B., Thaden, J. T., Mogno, I., Wierzbowski, J., Cottarel, G., ... Gardner, T. S. (2007). Large-scale mapping and validation of Escherichia coli transcriptional regulation from a compendium of expression profiles. *PLoS Biology*, 5(1), 0054–0066.
<https://doi.org/10.1371/journal.pbio.0050008>
- Fuller, D., Chen, W., Adler, M., Groisman, A., Levine, H., Rappel, W. J., & Loomis, W. F. (2010). External and internal constraints on eukaryotic chemotaxis. *Proceedings of the National Academy of Sciences of the United States of America*, 107(21), 9656–9659.
<https://doi.org/10.1073/pnas.0911178107>
- Grimbs, S., Selbig, J., Bulik, S., Holzhütter, H. G., & Steuer, R. (2007). The stability and robustness of metabolic states: Identifying stabilizing sites in metabolic networks. *Molecular Systems Biology*, 3(146). <https://doi.org/10.1038/msb4100186>
- Haynie, D. T. (2008). *Biological Thermodynamics*. Cambridge University Press (Vol. 39).
- Heylighen, F. (2002). The science of self-organization and adaptativity. In (L. D.) Kiel (Ed.), *Knowledge Management, Organizational Intelligence and Learning, and Complexity : v. 3*. EOLSS Publishers Co Ltd. <https://doi.org/10.1.1.38.7158>

- Hormoz, S. (2013). Cross talk and interference enhance information capacity of a signaling pathway. *Biophysical Journal*, *104*(5), 1170–1180. <https://doi.org/10.1016/j.bpj.2013.01.033>
- Hu, B., Chen, W., Levine, H., & Rappel, W. J. (2011). Quantifying Information Transmission in Eukaryotic Gradient Sensing and Chemotactic Response. *Journal of Statistical Physics*, *142*(6), 1167–1186. <https://doi.org/10.1007/s10955-011-0156-4>
- Ideker, T., & Sharan, R. (2008). Protein networks in disease. *Genome Research*, *18*(4), 644–652. <https://doi.org/10.1101/gr.071852.107.Freely>
- Iglesias, P. A., & Devreotes, P. N. (2008). Navigating through models of chemotaxis. *Current Opinion in Cell Biology*, *20*(1), 35–40. <https://doi.org/10.1016/j.ceb.2007.11.011>
- Jetka, T., Nienałowski, K., Filippi, S., Stumpf, M. P. H., & Komorowski, M. (2018). An information-theoretic framework for deciphering pleiotropic and noisy biochemical signaling. *Nature Communications*, *9*(1). <https://doi.org/10.1038/s41467-018-07085-1>
- Jordan, F., Nguyen, T.-P., & Liu, W. -c. (2012). Studying protein-protein interaction networks: a systems view on diseases. *Briefings in Functional Genomics*, *11*(6), 497–504. <https://doi.org/10.1093/bfpg/els035>
- Karlebach, G., & Shamir, R. (2008). Modelling and analysis of gene regulatory networks. *Nature Reviews Molecular Cell Biology*, *9*(10), 770–780. <https://doi.org/10.1038/nrm2503>
- Kauffman, S. A. (1969). Metabolic stability and epigenesis in randomly constructed genetic nets. *Journal of Theoretical Biology*, *22*(3), 437–467. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/5803332>
- Klir, G. J. (2005). *Uncertainty and Information: Foundations of Generalized Information Theory*. Wiley-Interscience. <https://doi.org/10.1002/0471755575>
- Lam, W. W. M., & Chan, K. C. C. (2012). Discovering Functional Interdependence Relationship in PPI Networks for Protein Complex Identification. *IEEE TRANSACTIONS ON BIOMEDICAL ENGINEERING*, *59*(4), 899–908.
- Lee, S. Y., Park, J. M., & Kim, T. Y. (2011). Application of Metabolic Flux Analysis in Metabolic Engineering. *Methods in Enzymology*, *498*, 67–93. <https://doi.org/10.1016/B978-0-12-385120-8.00004-8>
- Lekka, M., Leondaritis, G., Galanopoulou, K., Kitsioulis, E. (2015). Βασικές αρχές μεταγωγής σήματος. Μεμβρανικοί υποδοχείς. In *Βιολογικές μεμβράνες: από τη δομή στις λειτουργίες*. Athens: Hellenic Academic Libraries. Retrieved from <http://hdl.handle.net/11419/4313>
- Lenaerts, T., Ferkinghoff-Borg, J., Schymkowitz, J., & Rousseau, F. (2009). Information theoretical quantification of cooperativity in signalling complexes. *BMC Systems Biology*, *3*, 1–12. <https://doi.org/10.1186/1752-0509-3-9>
- Leng, J., & Sharrock, W. (Eds.). (2012). *Handbook of Research on Computational Science and Engineering*. IGI Global. <https://doi.org/10.4018/978-1-61350-116-0>
- Lian, H., Song, C., & Cho, Y. R. (2010). Decomposing protein interactome networks by graph entropy. *Proceedings - 2010 IEEE International Conference on Bioinformatics and Biomedicine, BIBM 2010*, 585–589. <https://doi.org/10.1109/BIBM.2010.5706633>
- Liang, S., Fuhrman, S., & Somogyi, R. (1998). Reveal, a general reverse engineering algorithm for inference of genetic network architectures. In *Pacific Symposium on Biocomputing*. (pp. 18–29). Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/9697168>
- Lin, D. (1998). An Information-Theoretic Definition of Similarity. In *Proceedings of the Fifteenth International Conference on Machine Learning* (pp. 296–304). San Francisco, CA, USA: Morgan Kaufmann Publishers Inc. Retrieved from <http://dl.acm.org/citation.cfm?id=645527.657297>

- Liu, J. Q., & Nakano, T. (2007). An information theoretic model of molecular communication based on cellular signaling. *Proceedings of the Bio-Inspired Models of Network, Information, and Computing Systems, Bionetics 2007. IEEE*, 316–321. <https://doi.org/10.1109/BIMNICS.2007.4610136>
- Lloyd, G. E. R. (1966). Polarity and Analogy: Two types of argumentation in early Greek Thought.
- Locasale, J. W., & Wolf-Yadlin, A. (2009). Maximum entropy reconstructions of dynamic signaling networks from quantitative proteomics data. *PLoS ONE*, 4(8). <https://doi.org/10.1371/journal.pone.0006522>
- Lodish H, Berk A, Kaiser CA, Krieger M, Scott MP, Bretscher A, Ploegh H, M. P. (2007). *Molecular Cell Biology* (6th editio). WH Freeman.
- Manke, T., Demetrius, L., & Vingron, M. (2005). Lethality and entropy of protein interaction networks. *Genome Informatics. International Conference on Genome Informatics*, 16(1), 159–163. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/16362918>
- Manke, T., Demetrius, L., & Vingron, M. (2006). An entropic characterization of protein interaction networks and cellular robustness. *Journal of the Royal Society Interface*, 3(11), 843–850. <https://doi.org/10.1098/rsif.2006.0140>
- Margolin, A. A., Wang, K., Lim, W. K., Kustagi, M., Nemenman, I., & Califano, A. (2006). Reverse engineering cellular networks. *Nature Protocols*, 1(2), 662–671. <https://doi.org/10.1038/nprot.2006.106>
- Meyer, P. E., Kontos, K., Lafitte, F., & Bontempi, G. (2007). Information-theoretic inference of large transcriptional regulatory networks. *Eurasip Journal on Bioinformatics and Systems Biology*, 2007(i). <https://doi.org/10.1155/2007/79879>
- Meyer, P., Marbach, D., Roy, S., & Kellis, M. (2010). Information-Theoretic Inference of Gene Networks Using Backward Elimination. In *Conference on Bioinformatics & Computational Biology (BIOCOMP'10)* (pp. 700–705). Las Vegas/USA. <https://doi.org/10.1093/nar/gks762>
- Mian, I. S., & Rose, C. (2011). Communication theory and multicellular biology. *Integrative Biology*, 3(4), 350–367. <https://doi.org/10.1039/c0ib00117a>
- Mitleton-Kelly, E. (2004). *Complex Systems and Evolutionary Perspectives on Organizations: The Application of Complexity Theory to Organisations. Ten principles of complexity and enabling infrastructures*. <https://doi.org/10.1177/1060028018757497>
- Morzy, M., Kajdanowicz, T., & Kazienko, P. (2017). On measuring the complexity of networks: Kolmogorov complexity versus entropy. *Complexity*, 2017. <https://doi.org/10.1155/2017/3250301>
- Mousavian, Z., Díaz, J., & Masoudi-Nejad, A. (2016). Information theory in systems biology. Part II: Protein-protein interaction and signaling networks. *Seminars in Cell and Developmental Biology*. <https://doi.org/10.1016/j.semcd.2015.12.006>
- Mousavian, Z., Kavousi, K., & Masoudi-Nejad, A. (2016a). Information theory in systems biology. Part I: Gene regulatory and metabolic networks. *Seminars in Cell & Developmental Biology*, 51, 3–13. <https://doi.org/10.1016/j.semcd.2015.12.007>
- Mousavian, Z., Kavousi, K., & Masoudi-Nejad, A. (2016b). Information theory in systems biology. Part I: Gene regulatory and metabolic networks. *Seminars in Cell and Developmental Biology*, 51, 3–13. <https://doi.org/10.1016/j.semcd.2015.12.007>
- Müller-Linow, M., Weckwerth, W., & Hütt, M. T. (2007). Consistency analysis of metabolic correlation networks. *BMC Systems Biology*, 1, 1–12. <https://doi.org/10.1186/1752-0509-1-44>
- Nemenman, I., Escola, G. S., Hlavacek, W. S., Unkefer, P. J., Unkefer, C. J., & Wall, M. E. (2007). Reconstruction of metabolic networks from high-throughput metabolite

- profiling data: In silico analysis of red blood cell metabolism. *Annals of the New York Academy of Sciences*, 1115, 102–115. <https://doi.org/10.1196/annals.1407.013>
- Nicolis, G., & Nicolis, C. (2007). *Foundations of Complex Systems: Nonlinear Dynamics, Statistical Physics, Information and Prediction*. World Scientific.
- Nicolis, G., & Nicolis, C. (2009). Foundations of Complex Systems. *European Review*, 17(02), 237. <https://doi.org/10.1017/S1062798709000738>
- Numata, J., Ebenhoh, O., & Knapp, E.-W. (2009). Measuring Correlations in Metabolomic Networks With Mutual Information, 112–122. https://doi.org/10.1142/9781848163003_0010
- Olsen, C., Meyer, P. E., & Bontempi, G. (2009). On the impact of entropy estimation on transcriptional regulatory network inference based on mutual information. *Eurasip Journal on Bioinformatics and Systems Biology*, 2009. <https://doi.org/10.1155/2009/308959>
- Padmanabhan, S., Burgess, K., Rankin, N., & Weidt, S. (2014). Metabolomics. *Handbook of Pharmacogenomics and Stratified Medicine*, 181–205. <https://doi.org/10.1016/B978-0-12-386882-4.00010-4>
- Patil, K. R., & Nielsen, J. (2005). Uncovering transcriptional regulation of metabolism by using metabolic network topology. *Proceedings of the National Academy of Sciences*, 102(8), 2685–2689. <https://doi.org/10.1073/pnas.0406811102>
- Pólya, G. (1990). *Mathematics and plausible reasoning*. Princeton University Press.
- Prigogine, I. (1987). Exploring complexity. *European Journal of Operational Research*.
- Prigogine, I., Nicolis, G., & Babloyantz, A. (1972). Thermodynamics of evolution. *Physics Today*, 25(11). <https://doi.org/10.1063/1.3071090>
- Radovic, M., Ghalwash, M., Filipovic, N., & Obradovic, Z. (2017). Minimum redundancy maximum relevance feature selection approach for temporal gene expression data. *BMC Bioinformatics*, 18(1), 9. <https://doi.org/10.1186/s12859-016-1423-9>
- R.Cheong, A. Rhee, C. Wang, I. Nemenman, A. L. (2011). Information transduction capacity of noisy biochemical signaling networks. *Science*, 6054, 354–358.
- Rhee, A., Cheong, R., & Levchenko, A. (2012). The application of information theory to biochemical signaling systems. *Physical Biology*, 9(4). <https://doi.org/10.1088/1478-3975/9/4/045011>
- Rosato, A., Tenori, L., Cascante, M., De Atauri Carulla, P. R., Martins dos Santos, V. A. P., & Saccenti, E. (2018). From correlation to causation: analysis of metabolomics data using systems biology approaches. *Metabolomics*, 14(4), 1–20. <https://doi.org/10.1007/s11306-018-1335-y>
- Rourleston, M. (1997). Significance testing of information theoretic functionals. *Physica D: Nonlinear Phenomena*, 110, 62–66.
- Samal, S. S., Krishnan, J., Esfahani, A. H., Lüders, C., & Weber, A. (2018). Metastable regimes and tipping points of biochemical networks with potential applications in precision medicine. *BioRxiv*. <https://doi.org/10.1101/466714>
- Sanguinetti, G., & Huynh-Thu, V. A. (2019). *Gene Regulatory Networks*.
- Sechidis, K., Papangelou, K., Metcalfe, P. D., Svensson, D., Weatherall, J., & Brown, G. (2018). Distinguishing prognostic and predictive biomarkers: An information theoretic approach. *Bioinformatics*, 34(19), 3365–3376. <https://doi.org/10.1093/bioinformatics/bty357>
- Shen, X., Yi, L., Jiang, X., He, T., Yang, J., Xie, W., ... Hu, X. (2017). Identifying protein complex by integrating characteristic of core-attachment into dynamic PPI network. *PLoS ONE*, 12(10), 1–15. <https://doi.org/10.1371/journal.pone.0186134>
- Silow, E. a, Mokry, A. V, & Jørgensen, S. E. (2011). Some Applications of Thermodynamics for

- Ecological Systems. In J. C. Moreno-Pirajan (Ed.), *Thermodynamics - Interaction Studies - Solids, Liquids and Gases*. InTech. <https://doi.org/10.5772/19611>
- Sinai, Y. (2009). Kolmogorov-Sinai entropy. In *Scholarpedia*.
- Srienc, F., & Unrean, P. (2010). A statistical thermodynamical interpretation of metabolism. *Entropy*, 12(8), 1921–1935. <https://doi.org/10.3390/e12081921>
- Stumpf, M. P. H., Balding, D. J., & Girolami, M. (2011). *Handbook of Statistical Systems Biology. Handbook of Statistical Systems Biology*. <https://doi.org/10.1002/9781119970606>
- Tostevin, F., & Ten Wolde, P. R. (2009). Mutual information between input and output trajectories of biochemical networks. *Physical Review Letters*, 102(21), 1–4. <https://doi.org/10.1103/PhysRevLett.102.218101>
- Trinh, C. T., Wlaschin, A., & Srienc, F. (2009). Elementary mode analysis: a useful metabolic pathway analysis tool for characterizing cellular metabolism. *Applied Microbiology and Biotechnology*, 81(5), 813–826. <https://doi.org/10.1007/s00253-008-1770-1>
- Tsimring, L. S. (2015). Noise in biology. *Reports on Progress in Physics*, 77(2), 1–62. <https://doi.org/10.1088/0034-4885/77/2/026601.Noise>
- Unrean, P., & Srienc, F. (2011). Metabolic networks evolve towards states of maximum entropy production. *Metabolic Engineering*, 13(6), 666–673. <https://doi.org/10.1016/j.ymben.2011.08.003>
- Van Helden, J., Wernisch, L., Gilbert, D., & Wodak, S. J. (2002). Graph-Based Analysis of Metabolic Networks. In H.-W. Mewes, H. Seidel, & B. Weiss (Eds.), *Bioinformatics and Genome Analysis* (pp. 245–274). Berlin, Heidelberg: Springer Berlin Heidelberg.
- Vazquez, A., Flammini, A., Maritan, A., & Vespignani, A. (2003). Global protein function prediction from protein-protein interaction networks. *Nature Biotechnology*, 21(6), 697–700. <https://doi.org/10.1038/nbt825>
- Villaverde, A. F., Ross, J., Morán, F., & Banga, J. R. (2014). MIDER: Network inference with mutual information distance and entropy reduction. *PLoS ONE*, 9(5), 1–15. <https://doi.org/10.1371/journal.pone.0096732>
- Von Bertalanffy, L. (1952). *Problems of life; an evaluation of modern biological thought*.
- Von Bertalanffy, L. (1962). *Modern Theories of Development : An introduction to Theoretical Biology*. <https://doi.org/10.2307/2179864>
- Von Bertalanffy, L. (1968). *General System Theory*. Georg. Braziller New York (Vol. 1).
- Wagner, A. (2008). Robustness and evolvability: A paradox resolved. *Proceedings of the Royal Society B: Biological Sciences*, 275(1630), 91–100. <https://doi.org/10.1098/rspb.2007.1137>
- Walhout, A. J. M., Vidal, M., & Dekker, J. (2012). *Handbook of systems biology : concepts and insights*. Elsevier/Academic Press.
- Waltermann, C., & Klipp, E. (2011). Information theory based approaches to cellular signaling. *Biochimica et Biophysica Acta - General Subjects*, 1810(10), 924–932. <https://doi.org/10.1016/j.bbagen.2011.07.009>
- Weber, B. H., Depew, D. J., Dyke, C., Salthe, S. N., Schneider, E. D., Ulanowicz, R. E., & Wicken, J. S. (1989). Evolution in thermodynamics perspective: an ecological approach. *Biology and Philosophy*, 4, 373–405.
- West, J., Bianconi, G., Severini, S., & Teschendorff, A. E. (2012). Differential network entropy reveals cancer system hallmarks. *Scientific Reports*, 2. <https://doi.org/10.1038/srep00802>
- Yu, R. C., Pesce, C. G., Colman-Lerner, A., Lok, L., Pincus, D., Serra, E., ... Brent, R. (2008). Negative feedback that improves information transmission in yeast signalling. *Nature*, 456(7223), 755–761. <https://doi.org/10.1038/nature07513>

- Zhang, A. (2009). *Protein interaction networks: Computational analysis* (Vol. 9780521888). Cambridge University Press. <https://doi.org/10.1017/CBO9780511626593>
- Zhao, Y.-Y., Vaziri, N. D., & Lin, R.-C. (2015). Lipidomics: New Insight Into Kidney Disease. *Advances in Clinical Chemistry*, 68, 153–175. <https://doi.org/10.1016/BS.ACC.2014.11.002>
- Zitnik, M., Sosič, R., Feldman, M. W., & Leskovec, J. (2019). Evolution of resilience in protein interactomes across the tree of life. *Proceedings of the National Academy of Sciences of the United States of America*. <https://doi.org/10.1073/pnas.1818013116>
- Zoppoli, P., Morganella, S., & Ceccarelli, M. (2010). TimeDelay-ARACNE: Reverse engineering of gene networks from time-course data by an information theoretic approach. *BMC Bioinformatics*. <https://doi.org/10.1186/1471-2105-11-154>
- Νικολάου, Χ. (2015). *Υπολογιστική βιολογία*.
- Νικολάου, Χ., & Χουβαρδάς, Π. (2015). *Υπολογιστική βιολογία*. Αθήνα: Σύνδεσμος Ελληνικών Ακαδημαϊκών Βιβλιοθηκών. Retrieved from <https://repository.kallipos.gr/handle/11419/1577>