



**ΕΘΝΙΚΟ ΚΑΙ ΚΑΠΟΔΙΣΤΡΙΑΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ**

**ΣΧΟΛΗ ΘΕΤΙΚΩΝ ΕΠΙΣΤΗΜΩΝ  
ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ ΚΑΙ ΤΗΛΕΠΙΚΟΙΝΩΝΙΩΝ**

**ΔΙΑΤΜΗΜΑΤΙΚΟ ΠΡΟΓΡΑΜΜΑ ΜΕΤΑΠΤΥΧΑΙΚΩΝ ΣΠΟΥΔΩΝ  
”ΤΕΧΝΟΛΟΓΙΕΣ ΠΛΗΡΟΦΟΡΙΚΗΣ ΣΤΗΝ ΙΑΤΡΙΚΗ ΚΑΙ ΤΗ ΒΙΟΛΟΓΙΑ”**

**ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ**

**Υπολογιστικές μέθοδοι βελτιστοποίησης τυχαίων περιπάτων  
σε γονιδιακά δίκτυα**

**Δημητριανός Α. Σάββα**

**Επιβλέπων: Γεώργιος Σπύρου, Ειδικός Λειτουργικός Επιστήμονας Α΄**

**ΑΘΗΝΑ**

**ΟΚΤΩΒΡΗΣ 2016**

## **ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ**

Υπολογιστικές μέθοδοι βελτιστοποίησης τυχαίων περιπάτων σε γονιδιακά δίκτυα

**Δημητριανός Α. Σάββα**

**A.M.: M1370**

**ΕΠΙΒΛΕΠΩΝ: Γεώργιος Σπύρου, Ειδικός Λειτουργικός Επιστήμονας Α΄**

## ΠΕΡΙΛΗΨΗ

Τα τελευταία χρόνια τα διαθέσιμα βιολογικά δίκτυα έχουν πολλαπλασιαστεί λόγω των αυξανόμενων δυνατοτήτων που προσφέρουν οι νέες τεχνολογίες που αναπτύσσονται καθημερινά. Λόγω του μεγάλου όγκου πληροφορίας, πολλές συστημικές προσεγγίσεις έχουν αναπτυχθεί με σκοπό την αποσαφήνιση και εκμετάλλευση των πληροφοριών που κρύβονται πίσω από τα βιολογικά δίκτυα αυτά. Μεταξύ αυτών των προσεγγίσεων, μέθοδοι ομαδοποίησης καθώς και μέθοδοι που βασίζονται στους τυχαίους περιπάτους, έχουν δείξει ελπιδοφόρα αποτελέσματα στην ανάδειξη και ιεράρχηση σημαντικών γονιδίων και γονιδιακών συσχετίσεων για διάφορες ασθένειες. Ωστόσο, λόγω της πολυπλοκότητας και του μεγέθους των βιολογικών δικτύων, υπάρχει η ανάγκη για βελτιστοποίηση των προσεγγίσεων αυτών, με σκοπό τη μείωση του υπολογιστικού χρόνου και κόστους ώστε να υπάρχει η δυνατότητα βαθύτερης εξερεύνησης των βιολογικών δικτύων.

Στη παρούσα μελέτη αναπτύχθηκε το XInfoWalk, ένα λογισμικό για την εφαρμογή μοντέλου τυχαίων περιπάτων με πληροφόρηση. Πιο συγκεκριμένα στόχος του λογισμικού αυτού ήταν η μείωση του υπολογιστικού χρόνου του αλγορίθμου Informed Walks, με την εισαγωγή αποδοτικών τεχνικών βελτιστοποίησης. Το μοντέλο Informed Walks χρησιμοποιεί πληροφορημένη αναζήτηση σε βιολογικά δίκτυα με τη χρήση τυχαίων περιπάτων και έχει χρησιμοποιηθεί για την ανάδειξη σημαντικών γονιδίων και γονιδιακών συσχετίσεων σε 7 τύπους καρκίνου. Το XInfoWalk χρησιμοποιήθηκε σε δεδομένα μικροσυστοιχιών από ασθενείς με έμφραγμα του μυοκαρδίου για την ανάδειξη σημαντικών γονιδίων και γονιδιακών σχέσεων. Τέλος διερευνήθηκε η ικανότητα των γονιδίων αυτών στην ανάδειξη υποψήφιων φαρμάκων και μοριακών μηχανισμών.

**ΘΕΜΑΤΙΚΗ ΠΕΡΙΟΧΗ:** Συστημική Βιοπληροφορική, Βιολογία Συστημάτων

**ΛΕΞΕΙΣ ΚΛΕΙΔΙΑ:** βελτιστοποίηση τυχαίων περιπάτων, βιολογικά δίκτυα, επανατοποθέτηση φαρμάκων, βιολογικοί μηχανισμοί, στατιστική ανάλυση

## **ABSTRACT**

Huge biological networks are being published everyday as a result of high throughput technologies. Because of the huge volume of data that has been made available, many systemic approaches have been developed to highlight hidden information within these biological networks. Among such approaches, clustering and random walks-based techniques have shown hopeful results about the discovery and ranking of important genes and genetic correlations for numerous diseases. However, there is a need for improvement of those approaches because of the increasing size and complexity of the biological networks. The main goal is the reduction of the computing time and cost to allow for deeper exploration of biological networks. In this particular study, we developed the XInfoWalk, a tool for exploring biological networks using informed random walk algorithms. More specifically, the goal of this software is the reduction of the computing time of the Informed Walks algorithms already developed in our group, by employing efficient techniques. The Informed Walks model explores gene networks using biologically informed random walks. It has been used for the discovery and ranking of important genes and genetic correlations for seven types of cancer. XInfoWalk was used in microarray data from patients with myocardial infarction for the discovery of important genes. XInfoWalk is an optimized version of Informed Walks that reduces dramatically the required running time and makes feasible longer runs. A case study for XinfoWalk has been performed using microarray data from patients with myocardial infarction for the discovery of important genes. The discovered gene network signatures, have been used in the investigation of possible repurposed drugs and related molecular mechanisms.

**SUBJECT AREA:** Systems Bioinformatics, Systems Biology

**KEYWORDS:** random walk optimization, biological networks, drug repositioning, biological mechanisms, statistical analysis