



ΕΘΝΙΚΟ ΚΑΙ ΚΑΠΟΔΙΣΤΡΙΑΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

**ΣΧΟΛΗ ΘΕΤΙΚΩΝ ΕΠΙΣΤΗΜΩΝ
ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ ΚΑΙ ΤΗΛΕΠΙΚΟΙΝΩΝΙΩΝ**

**ΔΙΑΤΜΗΜΑΤΙΚΟ ΠΡΟΓΡΑΜΜΑ ΜΕΤΑΠΤΥΧΙΑΚΩΝ ΣΠΟΥΔΩΝ
"ΤΕΧΝΟΛΟΓΙΕΣ ΠΛΗΡΟΦΟΡΙΚΗΣ ΣΤΗΝ ΙΑΤΡΙΚΗ ΚΑΙ ΤΗ ΒΙΟΛΟΓΙΑ"**

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

**W-DIAMOND:
Ανίχνευση συσχετιζόμενων γονιδίων σε ασθένειες**

Ιωάννης Κ. Βαξεβανάκης

Υπότροφος του Ιδρύματος Ωνάση



**ONASSIS
ΚΟΙΝΩΦΕΛΕΣ
ΙΔΡΥΜΑ**

Επιβλέποντες: **Σταύρος Περαντώνης**, Διευθυντής Ερευνών του Τμήματος Πληροφορικής και Τηλεπικοινωνιών, ΕΚΕΦΕ "Δημόκριτος"

Γεώργιος Παλιούρας, Διευθυντής Ερευνών του Τμήματος Πληροφορικής και Τηλεπικοινωνιών, ΕΚΕΦΕ "Δημόκριτος"

ΑΘΗΝΑ

Μάρτιος 2020



NATIONAL AND KAPODISTRIAN UNIVERSITY OF ATHENS

**SCHOOL OF SCIENCE
DEPARTMENT OF INFORMATICS & TELECOMMUNICATIONS**

**POSTGRADUATE PROGRAM "INFORMATION TECHNOLOGIES IN MEDICINE
AND BIOLOGY"**

MASTER THESIS

**W-DIAMOnD:
Detection of Diseases Pathways**

John K. Vaxevanakis

Scholar of Onassis Foundation



Supervisors: **Stavros Perantonis**, Research Director, Institute of Informatics and Telecommunications, NCSR "Demokritos"

George Paliouras, Research Director, Institute of Informatics and Telecommunications, NCSR "Demokritos"

**Athens
March 2020**

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

W-DIAMOnD:
Ανίχνευση συσχετιζόμενων γονιδίων σε ασθένειες

Ιωάννης Κ. Βαξεβανάκης

A.M.: ΠΙΒ 0169

Υπότροφος του Ιδρύματος Ωνάση



ΕΠΙΒΛΕΠΟΝΤΕΣ: **Σταύρος Περαντώνης**, Διευθυντής Ερευνών του Τμήματος Πληροφορικής και Τηλεπικοινωνιών, ΕΚΕΦΕ "Δημόκριτος"

Γεώργιος Παλιούρας, Διευθυντής Ερευνών του Τμήματος Πληροφορικής και Τηλεπικοινωνιών, ΕΚΕΦΕ "Δημόκριτος"

ΕΞΕΤΑΣΤΙΚΗ ΕΠΙΤΡΟΠΗ: **Σταύρος Περαντώνης**, Διευθυντής Ερευνών του Τμήματος Πληροφορικής και Τηλεπικοινωνιών, ΕΚΕΦΕ "Δημόκριτος"

Ζωή Κούρνια, Κύρια Ερευνήτρια, Ίδρυμα Ιατροβιολογικών Ερευνών Ακαδημίας Αθηνών (ΙΙΒΕΑΑ)

Γεώργιος Παλιούρας, Διευθυντής Ερευνών του Τμήματος Πληροφορικής και Τηλεπικοινωνιών, ΕΚΕΦΕ "Δημόκριτος"

Μάρτιος 2020

MASTER THESIS

**W-DIAMOnD:
Detection of Diseases Pathways**

John K. Vaxevanakis

S.R.N: ΠΙΒ 0169

Scholar of Onassis Foundation



SUPERVISORS: **Stavros Perantonis**, Research Director, Institute of Informatics and Telecommunications, NCSR "Demokritos"
George Paliouras, Research Director, Institute of Informatics and Telecommunications, NCSR "Demokritos"

EXAMINATION COMITEE: **George Paliouras**, Research Director, Institute of Informatics and Telecommunications, NCSR "Demokritos"
Stavros Perantonis, Research Director, Institute of Informatics and Telecommunications, NCSR "Demokritos"
Zoe Kournia, Principal Researcher, Biomedical Research Foundation Academy of Athens (BRFAA)

March 2020

ΠΕΡΙΛΗΨΗ

Στην παρούσα Διπλωματική Εργασία πραγματοποιήθηκε η δημιουργία ενός νέου αλγορίθμου, του (Weighted) W-DIAMOnD, ο οποίος βασίζεται στον αλγόριθμο DIseAse MOdule Detection (DIAMOnD). Ο DIAMOnD επιλύει το πρόβλημα Εύρεσης Συσχετιζόμενων Γονιδίων με Ασθένειες. Με δεδομένο ένα δυαδικό βιολογικό δίκτυο κάποιας ασθένειας αναφοράς και ένα σύνολο γονιδίων που είναι γνωστό ότι συσχετίζονται με την ασθένεια, ο αλγόριθμος βρίσκει επαναληπτικά ένα αριθμό γονιδίων βασισμένος σε μια συνάρτηση κόστους. Ο W-DIAMOnD έχει την ίδια λειτουργία με τον DIAMOnD, με την διαφορά ότι η συνάρτηση κόστους του νέου αλγορίθμου μπορεί να δεχτεί και συνεχείς τιμές. Με αυτόν τον τρόπο υπάρχει η δυνατότητα ο αλγόριθμος να εκμεταλλεύεται την πληροφορία των συνδέσεων στα αντίστοιχα δίκτυα με βάρη. Στην συνέχεια, συγκρίθηκαν οι δύο αλγόριθμοι και εξετάστηκε η αποτελεσματικότητά τους τόσο σε συνθετικά όσο και σε πραγματικά βιολογικά δίκτυα. Τέλος πραγματοποιήθηκε η εφαρμογή των αλγορίθμων σε ένα βιολογικό δίκτυο βασισμένο στην ασθένεια Μυϊκή Δυστροφία Duchenne και βρέθηκαν κάποια πιθανά γονίδια να σχετίζονται με την ασθένεια.

ΘΕΜΑΤΙΚΗ ΠΕΡΙΟΧΗ: Υπολογιστική Βιολογία

ΛΕΞΕΙΣ ΚΛΕΙΔΙΑ: βιολογικά δίκτυα, μονοπάτια ασθενειών, αλληλεπίδραση γονιδίων, υπεργεωμετρική κατανομή, μυϊκή δυστροφία.

ABSTRACT

In this thesis, a new Weighted W-DIAMOnD algorithm was developed, based on the DIseAse MOdule Detection (DIAMOnD) algorithm. DIAMOnD solves the problem of Finding Disease Pathways. Given a binary biological network of a reference disease and a set of genes that are known to be associated with the disease, the algorithm repeatedly finds a new set of genes based on a cost function. W-DIAMOnD works similarly to DIAMOnD, except that W-DIAMOnD's cost function can accept continuous values. In this manner, the algorithm can exploit the information of the edges in the corresponding weighted networks. Furthermore, the two algorithms were compared and tested for their effectiveness in both synthetic and real biological networks. Finally, the algorithms were implemented in a biological network based on Duchenne's Muscular Dystrophy (DMD) disease and some possible genes related to the disease were found.

SUBJECT AREA: Computational Biology

KEYWORDS: biological networks, diseases pathways, gene interaction, hypergeometric distribution, muscular dystrophy