

## ΔΙΠΛΩΜΑΤΙΚΕΣ ΕΡΓΑΣΙΕΣ

Καθηγητής **Ε. Γαλλόπουλος**

Εργαστήριο Πληροφοριακών Συστημάτων Υψηλής Επίδοσης

### Γενική περιγραφή

- Τα θέματα των διπλωματικών αφορούν κατά κύριο λόγο στη μελέτη και στην αξιολόγηση υπαρχόντων μεθόδων και εργαλείων για τον Επιστημονικό Υπολογισμό και στην ανάπτυξη και υλοποίηση νέων.
- Βασικοί συνεργάτες στις Δ.Ε. είναι ο Δρ. Ι. Βενέτης και η κα. Ε.-Μ. Κοντοπούλου. Επιπλέον υπάρχει περίπτωση συνεργασιών με άλλους ερευνητές σε ακαδημαϊκά ιδρύματα στο εξωτερικό.
- Για περισσότερες πληροφορίες παρακαλείσθε να επικοινωνήσετε άμεσα με τον επιβλέποντα.
- Τυπικοί στόχοι στο τέλος είναι η υλοποίηση μεθόδων, η συστηματική αξιολόγησή τους, η δυνατότητα χρήσης των αποτελεσμάτων από την κοινότητα, καθώς και γραπτή και προφορική παρουσίασή τους σε επίπεδο επιστημονικού φόρουμ (συνέδριο, επιστημονικό περιοδικό, κ.λπ.) Στο παρελθόν, ορισμένες από τις εργασίες οδήγησαν σε πολύ αξιόλογες δημοσιεύσεις και λογισμικό.

*Το πιο σημαντικό είναι να είστε εσείς περήφανοι για το αποτέλεσμα της εργασίας η οποία σηματοδοτεί με δημιουργικό τρόπο την κορωνίδα του πολυετούς μόχθου σας στο Τμήμα.*

**Προϋποθέσεις για την εκπόνηση μιας καλής διπλωματικής:** α) Να σας ενδιαφέρουν οι προτεινόμενες θεματικές περιοχές. Σε γενικές γραμμές, αυτές αφορούν στη μελέτη, στο σχεδιασμό και στην υλοποίηση νέων αλγορίθμων σε βιβλιοθήκες και περιβάλλοντα για σύγχρονα υπολογιστικά συστήματα. Δεν αποτελεί έκπληξη βεβαίως ότι πολλά αφορούν σε υπολογισμούς με μητρώα!

β) Να έχετε κατανοήσει την ύλη των σχετικών μαθημάτων και να σας ενδιαφέρουν οι υλοποιήσεις. Αν έχετε πετύχει καλή επίδοση σε αυτά ή έχετε ασχοληθεί σοβαρά και σας ενδιέφεραν οι εργασίες που δόθηκαν στον Επιστημονικό Υπολογισμό, θεμελιώνεται μια πολύ καλή προοπτική για καλά αποτελέσματα. Αρκετά θέματα αφορούν στην παράλληλη επεξεργασία, και στην ανάκτηση πληροφορίας. Σε κάθε διπλωματική, σημαντικό ρόλο θα έχουν οι υλοποιήσεις.

γ) Στις περισσότερες εργασίες χρειάζεται να έχετε τη διάθεση να προγραμματίσετε. Το περιβάλλον εξαρτάται από τη Δ.Ε. αλλά συνήθως πρόκειται για MATLAB, C και Python.

**Προσοχή:** Οι παρακάτω εκφωνήσεις δίνουν ένα γενικό πλαίσιο και μπορεί να τροποποιηθούν μετά από τη συνάντηση με τον κ. Γαλλόπουλο.

## Επιστημονικοί υπολογισμοί και παράλληλη επεξεργασία

### 1 Αλγόριθμοι και παράλληλος υπολογισμός σε CUDA

Τα τελευταία χρόνια παρατηρείται μια στροφή των κατασκευαστών καρτών γραφικών να συμπεριλάβουν χαρακτηριστικά στα προϊόντα τους που θα τα καθιστά δυνατούς συνεπεξεργαστές στην επίλυση γενικότερων υπολογιστικών προβλημάτων και όχι μόνο προβλημάτων γραφικών. Χαρακτηριστικό παράδειγμα αποτελεί η κάρτα NVIDIA Tesla K20X με 2688 πυρήνες. Σημαντικές προσπάθειες έχουν γίνει και ως προς την ευκολία προγραμματισμού αυτών των συστημάτων. Το μοντέλο προγραμματισμού CUDA επεκτείνει την γλώσσα προγραμματισμού C για την διευκόλυνση του προγραμματισμού των παραπάνω καρτών.

Στη Δ.Ε. αυτή εργασία ενδιαφέρει η διερεύνηση παράλληλων αλγορίθμων για ένα από τα παρακάτω προβλήματα (υπάρχει περίπτωση εκπόνησης περισσότερων της μίας διπλωματικής):

1. Μέθοδοι για την ταχεία παράλληλη επίλυση μεγάλων γραμμικών συστημάτων με δομή ζώνης (banded). Πολλά μητρώα σε προσομοιώσεις έχουν τέτοια δομή ή μπορεί να μετατραπούν σε μητρώα με παρόμοια δομή μετά από αναδιάταξη γραμμών και στηλών,  $PAQ$ , ή να έχουν παραπλήσια δομή,  $PAQ + E$  για κάποιο «μικρό»  $E$ . Επομένως, μέθοδοι για τη διαχείριση τέτοιων μητρώων αποκτούν μεγάλο ενδιαφέρον. Ο Spike είναι ένας πολύ σημαντικός αλγόριθμος για την παράλληλη επίλυση τέτοιων συστημάτων. Χρησιμοποιείται επίσης ως προρυθμιστής (preconditioner) για την επιτάχυνση επαναληπτικών επιλυτών. Στα πλαίσια της διπλωματικής εργασίας ζητείται η ενοποίηση δύο τεχνολογιών αιχμής: η υλοποίηση αλγορίθμων με χρήση της CUDA για την επίλυση συστημάτων με μητρώα ζώνης με έμφαση στον Spike και η αξιολόγηση των ευρημάτων. Στόχος είναι να αξιοποιούνται οι δυνατότητες παραλληλισμού που παρέχουν οι κάρτες γραφικών και η συστηματική σύγκριση των επιδόσεων για μεγάλα προβλήματα. Επιπλέον, να αξιοποιηθεί η υψηλή ταχύτητα των καρτών γραφικών σε πράξεις αριθμητικής κινητής υποδιαστολής 32-bits για βέλτιστη επίδοση με περιστασιακή μόνο αρωγή αριθμητικής διπλής ακρίβειας.
2. Μία άλλη κατεύθυνση αφορά σε μεθόδους για την ταχεία εκτίμηση του ίχνους και της διαγωνίου αντιστρόφου μητρώου, η υλοποίηση της παραλληλοποίησης με CUDA και η αξιολόγηση της επίδοσης σε σύγκριση με την παραλληλοποίηση σε άλλους πολυπύρηνους επεξεργαστές και συστάδες επεξεργαστών. Ο υπολογισμός αυτών των τιμών χρησιμεύει σε πολλές σύγχρονες εφαρμογές, όπως για παράδειγμα στην περιοχή της ποσοτικοποίησης της αβεβαιότητας (uncertainty quantification) αλλά και σε πιο στοιχειώδεις αλλά δύσκολους υπολογισμούς. Για παράδειγμα, η διαγώνιος του εκθετικού μητρώου χρησιμοποιείται στην αξιολόγηση δικτύων.

### 2 Υπολογισμοί με συναρτήσεις μητρώων: Μέθοδοι, υλοποιήσεις και εφαρμογές

Μεγάλος αριθμός προβλημάτων στην επιστήμη και τεχνολογία περιέχει ως πυρήνα τον υπολογισμό  $Cf(A)B$  όπου  $A, B, C$  είναι μητρώα,  $f$  κάποια συνάρτηση και  $t$  κάποια παράμετρος. Δύο εξέχουσες περιπτώσεις είναι όταν  $f(A) = A^{-1}$  ή  $\exp(A)$ . Για παράδειγμα, σε σύγχρονες μεθόδους υπολογισμού χαρακτηριστικών δικτύων αλλά και στην επίλυση μεγάλων συστημάτων διαφορικών εξισώσεων που εξαρτώνται από το χρόνο με μεθόδους που αποκαλούνται 'εκθετικοί διαδότες'. Η διπλωματική θα έχει για στόχο τη μελέτη αυτού του προβλήματος. Μία κατεύθυνση αφορά στην εύρεση των βασικών

μεθόδων στη βιβλιογραφία και στην κατάταξή τους. Επίσης, η επιλογή ορισμένων από αυτές και η σχεδίαση και υλοποίηση του υπολογισμού σε παράλληλο υπολογιστικό σύστημα. Ως προς αυτό, ενδιαφέρει ιδιαίτερα η χρήση υπολογιστικών μοντέλων που λαμβάνουν υπόψη τους τη αριθμητική, τις μεταφορές και την παραλληλία.

### **3 Μελέτη και ανάπτυξη εργαλείοθκών για επιστημονικούς υπολογισμούς**

Στόχος της Δ.Ε. είναι η μελέτη υπαρχόντων εργαλείων σε σύγχρονα υπολογιστικά περιβάλλοντα όπως η MATLAB και η δημιουργία νέων. Μία κατεύθυνση αφορά στην αυτοματοποίηση της αξιολόγησης επίδοσης αλγορίθμων με βάση τον χρόνο επίλυσης και την ταχύτητα. Είναι γνωστό ότι αυτά τα χαρακτηριστικά υπολογίζονταν με τεχνικές όπως το tic-toc και πιο πρόσφατα με το timeit το οποίο έχει πρόσφατα ενσωματώθηκε στη MATLAB καθώς και με το profile. Είναι επίσης γνωστό ότι στο παρελθόν η Mathworks κατήργησε την εντολή flops. Στη διπλωματική αυτή αναμένεται :

- Να γίνουν εκτενείς πειραματισμοί με υπάρχουσες συναρτήσεις και αξιολόγησή τους.
- Να σχεδιαστούν και να υλοποιηθούν προγράμματα για την αυτόματη κατασκευή μοντέλων επίδοσης προγραμμάτων.
- Να διερευνηθεί η δυνατότητα εκτίμησης του πλήθους πράξεων αριθμ. κινητής υποδιαστολής που εκτελούνται στο πρόγραμμα και να υλοποιηθεί εργαλείο με αυτό το στόχο.
- Μία διαφορετική κατεύθυνση αφορά στη διερεύνηση του πολύ ενδιαφέροντος θέματος των δυνατοτήτων πολλών υπολογισμών να διεξαχθούν σε συνδυασμό διαφορετικών επιπέδων ακρίβειας. Μεγάλο μέρος τους, οικονομικά, σε χαμηλή ακρίβεια με την περιστασιακή μόνον αρωγή διορθωτικών υπολογισμών πολύ μεγαλύτερης ακρίβειας. Ενδιαφέρει η μελέτη αυτού του θέματος και η δημιουργία εργαλείων για την επίτευξη μεγαλύτερης ταχύτητα για την ακρίβεια που είναι αποδεκτή στην εφαρμογή.

## **Εργαλεία στην Ανάκτηση Πληροφορίας**

Έχουμε θέσει στη διάθεση της επιστημονικής κοινότητας το εργαλείο TMG (Text to Matrix Generator) (πληροφορίες [εδώ](#)) το οποίο αποτελεί προϊόν έρευνας και ανάπτυξης μελών της ομάδας. Πρόκειται για μια εργαλειοθήκη (toolbox) MATLAB (τμήματα του κώδικα στη γλώσσα Perl) για την αυτόματη προετοιμασία πινάκων «όρων-κειμένων» από συλλογές κειμένων με στόχο την περαιτέρω επεξεργασία τους (ως μητρώα) με τεχνικές γραμμικής άλγεβρας. Το εργαλείο αυτό έχει ήδη διανεμηθεί σε πολλούς χρήστες σε όλο τον κόσμο και χρησιμοποιείται σε πανεπιστήμια, ερευνητικά κέντρα και στη βιομηχανία. Οι διπλωματικές που θα εκπονηθούν θα έχουν ως αποτέλεσμα την ανάπτυξη και την επέκταση των δυνατοτήτων και χρήσεων του λογισμικού. Στόχος είναι το εργαλείο αυτό αφ' ενός να εξελιχθεί ώστε να αντιμετωπίζει μεγαλύτερο εύρος προβλημάτων, ταχύτερα, και με μεγαλύτερη ποικιλία αλγορίθμων. Σε εξέλιξη βρίσκονται προσπάθειες για την ενσωμάτωση πρόσφατων αλγοριθμικών τεχνικών βασισμένων στη συνέργεια ανάλυσης μητρώων και πιθανοτικών τεχνικών.

### **4 Επέκταση του TMG σε υπολογισμούς με γραφήματα και εφαρμογές σε κείμενα**

Στόχος της Δ.Ε. είναι η επέκταση του TMG ώστε να μπορεί να χρησιμοποιηθεί για υπολογισμούς σε γραφήματα σε συνδυασμό με την διερεύνηση μετρικών για την αξιολόγηση των γραφημάτων μέσω χαρακτηριστικών τους όπως η κεντρικότητα (centrality). Για την βελτιστοποίηση των υπολογισμών θα απαιτηθεί η χρήση επαναληπτικών τεχνικών καθώς και η αξιοποίηση των χαρακτηριστικών των γραφημάτων. Προβλέπεται διερεύνηση και αξιολόγηση άλλων υπαρχόντων προγραμμάτων με παρόμοιο στόχο (όπως τα BGL, GrTheory, Brain Connectivity Toolbox, igraph) και αξιοποίηση των ιδιοτήτων των γραφημάτων και των αντίστοιχων αραιών μητρώων για την επίτευξη της βέλτιστης επίδοσης των λειτουργιών αυτών. Η διεπαφή του συστήματος θα πρέπει να είναι στα πλαίσια του TMG, δηλ. υλοποιημένη σε MATLAB, αλλά μέρος της υλοποίησης αναμένεται να είναι σε C για καλύτερη επίδοση.

### **5 Επέκταση του TMG σε υπολογισμούς μεγάλου όγκου δεδομένων και με αναβάθμιση λειτουργιών**

Το TMG προς το παρόν έχει γραφτεί κυρίως σε MATLAB με ένα τμήμα Perl. Στόχος μιας διπλωματικής είναι να επαναπρογραμματιστούν βασικά τμήματα του TMG με στόχο τη βελτίωση της επίδοσης και την αποτελεσματική αντιμετώπιση προβλημάτων πολύ μεγάλης διάστασης διατηρώντας όμως τη διεπαφή MATLAB. Θα εξεταστούν θέματα που η χρήση της μνήμης καθώς και δομών που επιτρέπουν την αποθήκευση και τη διαχείριση πολύ μεγάλων συλλογών. Θα εξεταστούν επίσης τρόποι αναβάθμισης της διεπαφής και των οπτικοποιήσεων.

### **6 Τεχνικές διαστατικής μείωσης και ομαδοποίησης στο TMG με χρήση νέων και παράλληλων τεχνικών**

Στόχος είναι η παραλληλοποίηση αλγορίθμων που βασίζονται σε νέες τεχνικές (ενίοτε πιθανοτικές) για την προσέγγιση μητρώων μέσω διαστατικής μείωσης και η ενσωμάτωσή τους στο TMG. Το θέμα αυτό μπορεί επίσης να συνδυαστεί με τη μελέτη μεθόδων αντιμετώπισης από το TMG προβλημάτων με όγκο που δεν χωρούν στην κύρια μνήμη και με παράλληλη επεξεργασία.

## Υπολογισμοί σε μητρώα και γραφήματα

### 7 Γραφήματα, μητρώα σε εργοδικές ιστορίες

Σε πρόσφατες εργασίες μας (βλ. [NHW'13](#) και [\(2012 Ted Nelson award, HT'12\)](#) ) διερευνήσαμε υπολογιστικές τεχνικές για εργοδική λογοτεχνία. Αυτή έχει οριστεί (από τον Aarseth<sup>1</sup>) ως το είδος της λογοτεχνίας στην οποία απαιτείται από τον αναγνώστη να λάβει κάποιαν απόφαση (έργο) για να πάει στην «επόμενη σελίδα» ώστε να ακολουθήσει την εξέλιξη της ιστορίας (οδό). Ένα καλό παράδειγμα λογοτεχνίας αυτού του τύπου είναι ορισμένα βιβλία του Ευγένιου Τριβιζά.

Στόχος της διπλωματικής είναι αφ' ενός να συλλέξει δεδομένα από on-line «εργοδικές» πηγές, η κατασκευή συλλογής μητρώων από τα δεδομένα αυτά και η διεξοδική ανάλυσή τους με σύγχρονες τεχνικές υπολογιστικής γραμμικής άλγεβρας και γραφοθεωρίας ως συνέχεια των αναλύσεων που αναφέρονται στις εργασίες που αναφέρθηκαν. Μία επιπλέον κατεύθυνση αφορά στο συνδυασμό πληροφοριών από τις διασυνδέσεις με πληροφορίες που περιέχονται στις σελίδες.

### 8 Αλγόριθμοι για link-based ranking σε γραφήματα και εφαρμογές

Η περιοχή αυτή αφορά γενικά τις ιδιότητες μεθόδων βαθμολόγησης κόμβων δικτύων και γενικότερα στον υπολογισμό χαρακτηριστικών τιμών όπως centrality. Αλγόριθμοι αυτού του τύπου χρησιμοποιούνται ευρύτατα σήμερα για βαθμολόγηση σε γραφήματα που προέρχονται από δίκτυα κ.λπ. Οι αλγόριθμοι αυτοί χρησιμοποιούν πυρήνες γραμμικής άλγεβρας σε συνδυασμό με πιθανοτικές τεχνικές. Για παράδειγμα, ο PageRank που ήταν ο βασικός αλγόριθμος που χρησιμοποιούσε η Google για την «βαθμολόγηση» ιστοσελίδων βασίζεται στις τιμές του κυρίαρχου ιδιοδιανύσματος του μητρώου  $\alpha S + (1 - \alpha)H$  όπου το μητρώο  $H$  είναι πρώτης τάξης και το  $S$  προέρχεται από το μητρώο γειτνίασης του γραφήματος. Μία επέκταση της αξιολόγησης αυτής προέρχεται από την προσέγγιση του διανύσματος  $f(S)^n$  για επιλεγμένη συνάρτηση  $f$  και διάνυσμα  $v$ . Στόχος της διπλωματικής είναι η διερεύνηση των μεθόδων προσέγγισης των τιμών αυτών εκκινώντας από την κλασική μέθοδο καθώς και την θεμελιώδη εργασία του Katz, η ευαισθησία των μεθόδων αυτών ως προς τα δεδομένα του προβλήματος, η επέκταση των μετρικών και των μεθόδων, η ερμηνεία τους και ο υπολογισμός τους. Για παράδειγμα, ενδιαφέρει ιδιαίτερα η επέκταση και εφαρμογή των μεθόδων που αναπτύχθηκαν σε [πρόσφατη εργασία μας](#).

Μία άλλη κατεύθυνση, σε συνεργασία με τον καθηγητή κ. Μεγαλοοικονόμου αφορά στη διερεύνηση τεχνικών βαθμολόγησης για χρονοεξαρτώμενα γραφήματα ώστε να λαμβάνεται υπόψη ο παράγοντας του χρόνου στον υπολογισμό των χαρακτηριστικών τιμών του γραφήματος.

### 9 Υπολογισμοί με μη αρνητικά μητρώα και εφαρμογές στην ανάκτηση πληροφορίας

Η (προσεγγιστική) [μη αρνητική παραγοντοποίηση \(μη αρνητικών\) μητρώων](#),  $A = WH$ , όπου  $A, W, H \geq 0$ , είναι ένα σημαντικό εργαλείο στην περιοχή της ανάκτησης πληροφορίας. Το μαθηματικό πρόβλημα είναι πολύ πιο δύσκολο από τις κλασικές παραγοντοποιήσεις (π.χ. SVD) και στη βιβλιογραφία έχουν προταθεί πολλές μέθοδοι για τον υπολογισμό των παραγόντων. Στόχος είναι η διερεύνηση του

<sup>1</sup>βλ. [http://en.wikipedia.org/wiki/Espen\\_Aarseth](http://en.wikipedia.org/wiki/Espen_Aarseth)

προβλήματος, των αλγορίθμων επίλυσης, των βιβλιοθηκών που παρέχουν αυτή τη δυνατότητα και η χρήση σε εφαρμογές. Ενδιαφέρουν ιδιαίτερα η παραγοντοποίηση μεγάλων μητρώων οπότε προέχει η διατήρηση του χαρακτηριστικού της αραιότητας στους παράγοντες καθώς και η χρήση του σε ανάκτηση πληροφορίας από κείμενα καθώς και παράλληλες τεχνικές για την υλοποίηση του NMF σε κάρτες γραφικών.

## **10 Ειδικά θέματα**

Συνεπιβλέπων: Καθηγητής κ. Β. Μεγαλοοικονόμου. Δείτε σχετική ανακοίνωση του συνεπιβλέποντος.

## **11 Ειδικά θέματα**

Σχεδιασμός ειδικού μαθήματος 'Studio 1ου έτους'. Περισσότερες λεπτομέρειες από τον επιβλέποντα. Το θέμα αυτό απαιτεί ενδιαφέρον για εκπαιδευτικά θέματα τμημάτων σαν το ΤΜΗΥΠ.

## **12 Άλλα θέματα**

Όπως αναφέραμε, οι παραπάνω εκφωνήσεις δίνουν ένα γενικό πλαίσιο και μπορεί να τροποποιηθούν ή να δοθεί και άλλη δυνατότητα. Παρακαλώ επικοινωνήστε με τον κ. Γαλλόπουλο.