



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ  
ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ  
ΤΟΜΕΑΣ ΣΗΜΑΤΩΝ, ΕΛΕΓΧΟΥ ΚΑΙ ΡΟΜΠΟΤΙΚΗΣ  
ΕΡΓΑΣΤΗΡΙΟ ΟΡΑΣΗΣ ΥΠΟΛΟΓΙΣΤΩΝ, ΕΠΙΚΟΙΝΩΝΙΑΣ ΛΟΓΟΥ ΚΑΙ ΕΠΕΞΕΡΓΑΣΙΑΣ ΣΗΜΑΤΩΝ  
ΖΩΓΡΑΦΟΥ 157 73, ΑΘΗΝΑ

Σημασιολογική Κατάτμηση Αντικειμένων με  
Τεχνικές Φασματικής Θεωρίας Γράφων

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

ΤΟΥ

ΕΥΑΓΓΕΛΟΥ ΧΑΤΖΗΠΙΑΝΤΑΖΗ

Επιβλέπων: Πέτρος Μαραγκός  
Καθηγητής Ε.Μ.Π.

ΕΡΓΑΣΤΗΡΙΟ ΟΡΑΣΗΣ ΥΠΟΛΟΓΙΣΤΩΝ, ΕΠΙΚΟΙΝΩΝΙΑΣ ΛΟΓΟΥ ΚΑΙ ΕΠΕΞΕΡΓΑΣΙΑΣ ΣΗΜΑΤΩΝ  
Αθήνα, Οκτώβριος 2018





Εθνικό Μετσόβιο Πολυτεχνείο  
Σχολή Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών  
Τομέας Σημάτων, Ελέγχου και Ρομποτικής  
Εργαστήριο Όρασης Υπολογιστών, Επικοινωνίας Λόγου και Επεξεργασίας  
Σημάτων  
Ζωγράφου 157 73, Αθήνα

## Σημαιολογική Κατάτμηση Αντικειμένων με Τεχνικές Φασματικής Θεωρίας Γράφων

### ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

ΤΟΥ

**ΕΥΑΓΓΕΛΟΥ ΧΑΤΖΗΠΑΝΤΑΖΗ**

**Επιβλέπων:** Πέτρος Μαραγκός  
Καθηγητής Ε.Μ.Π.

Εγκρίθηκε από την τριμελή εξεταστική επιτροπή την 11 Οκτωβρίου 2018 .

(Υπογραφή)

(Υπογραφή)

(Υπογραφή)

.....  
Πέτρος Μαραγκός  
Καθηγητής  
Ε.Μ.Π.

.....  
Γεράσιμος Ποταμιάνος  
Αναπληρωτής Καθηγητής  
Πανεπιστήμιο Θεσσαλίας

.....  
Χαράλαμπος Ψυλλάκης  
Λέκτορας  
Ε.Μ.Π.

Αθήνα, Οκτώβριος 2018





(Υπογραφή)

.....  
**Ευάγγελος Χατζηπανταζής**

Διπλωματούχος Ηλεκτρολόγος Μηχανικός και Μηχανικός Υπολογιστών Ε.Μ.Π.

Copyright ©–All rights reserved Ευάγγελος Χατζηπανταζής, 2018.

Με επιφύλαξη παντός δικαιώματος.

Απαγορεύεται η αντιγραφή, αποθήκευση και διανομή της παρούσας εργασίας, εξ' ολοκλήρου ή τμήματος αυτής, για εμπορικό σκοπό. Επιτρέπεται η ανατύπωση, αποθήκευση και διανομή για σκοπό μη κερδοσκοπικό, εκπαιδευτικής ή ερευνητικής φύσης, υπό την προϋπόθεση να αναφέρεται η πηγή προέλευσης και να διατηρείται το παρόν μήνυμα. Ερωτήματα που αφορούν τη χρήση της εργασίας για κερδοσκοπικό σκοπό πρέπει να απευθύνονται προς τον συγγραφέα.

Οι απόψεις και τα συμπεράσματα που περιέχονται σε αυτό το έγγραφο εκφράζουν το συγγραφέα και δεν πρέπει να ερμηνευθεί ότι αντιπροσωπεύουν τις επίσημες θέσεις του Εθνικού Μετσόβιου Πολυτεχνείου.





## Εθνικό Μετσόβιο Πολυτεχνείο

Σχολή Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών

Τομέας Σημάτων, Ελέγχου και Ρομποτικής

Εργαστήριο Όρασης Υπολογιστών, Επικοινωνίας Λόγου και Επεξεργασίας

Σημάτων

Ζωγράφου 157 73, Αθήνα



# Περίληψη

Οι μέθοδοι κατάτμησης δεδομένων σε γράφους έχουν τα τελευταία χρόνια παρουσιάσει μια άυξηση ενδιαφέροντος, λόγω της αποτελεσματικότητάς τους σε πληθώρα εφαρμογών και της επαρκούς αναπαράστασης της δομής των δεδομένων και των σχεσιακών ιδιοτήτων τους. Ιδιαίτερα στο χώρο της Όρασης Υπολογιστών η διαδικασία δημιουργίας γράφων από εικόνες έχει αρχίσει να λύνει πολλά προβλήματα μη επιβλεπόμενης μάθησης, με τα οποία πάσχιζαν οι κλασσικές μέθοδοι. Ένα τέτοιο παράδειγμα αλγορίθμων κατάτμησης δεδομένων στην δομή του γράφου είναι οι φασματικές γραφοτομές, που ανακαλύπτουν καθολικές ιδιότητες ενός γράφου και των συμπαγών cluster που τον απαρτίζουν αναλύοντας το φάσμα του. Μέσα από την συσχέτιση της κατάτμησης εικόνων με την συσταδοποίηση δεδομένων και την αντιληπτική ομαδοποίηση και με διαίσθηση που αφορμάται από τον νόμους Gestalt της ψυχολογίας, εφαρμόζουμε μια αυτοματοποιημένη ιεραρχική μέθοδο για τον μετασχηματισμό της αναπαράστασης της εικόνας από την δομή των pixel στην δομή του γράφου. Στην συνέχεια, θεμελιώνουμε πιά φορμαλιστικά τους αλγορίθμους φασματικής κατάτμησης, αναλύουμε και λύνουμε το πρόβλημα βελτιστοποίησης που σχηματίζουν οι σταθμισμένες γραφοτομές και επιχειρηματολογούμε για την ορθότητά της λύσης αντιστοιχίζοντάς τη ταυτόχρονα και με άλλες καλά μελετημένες τεχνικές κατάτμησης. Τέλος, ανοίγουμε το “μαύρο κουτί” των φασματικών γραφοτομών, που αποτελούν μη επιβλεπόμενη μέθοδο, με την προσθήκη εξωγενούς πληροφορίας και θεμελιώνουμε θεωρητικά ένα framework για την αποτελεσματική εισαγωγή εξωτερικών περιορισμών.

## Λέξεις Κλειδιά

φασματική θεωρία γραφοτομών, Σταθμισμένες γραφοτομές, Γράφος γειτονικών περιοχών, συστάδες εικονοστοιχείων, ημι-επιβλεπόμενη μάθηση, περιορισμοί υποχρεωτικής σύζευξης, περιορισμοί υποχρεωτικής αποσύζευξης, προσημασμένες σταθμισμένες γραφοτομές.



# Abstract

Methods on data segmentation on graphs have in recent years shown an increase in interest due to their effectiveness in a variety of applications and a successful representation of the structure of the data and their relational properties. Particularly in the area of Computer Vision, the process of creating graphs from images has begun to solve many unsupervised learning problems with which classical methods have struck. One such example of data partitioning algorithms in the structure of a graph is the spectral graph clustering family that discovers the global properties of a graph and the underlying compact structure of the clusters that compose it by analyzing its spectrum. Through the association of image segmentation with data clustering and perceptual grouping and with intuition inspired by the laws of Gestalt in psychology, we apply an automated hierarchical method to transform image representation from the pixelwise representation into a graph representation. We then formalize rigorously the spectral segmentation algorithms, analyze and solve the optimization problem formed by the Normalized Cuts and argue for its correctness by associating it simultaneously with other well-thought-out segmentation techniques. Finally, we open the "black box" of the spectral graph clustering algorithms, which are currently a non-supervised method, with the addition of extraneous information and we theoretically establish a framework for the effective introduction of external constraints.

## Keywords

spectral graph clustering, Normalized Cuts, Region Adjacency Graph, Superpixels, Semi-supervised learning, Must-link constraints, Cannot-link Constraints, Signed Normalized Cuts.





# Ευχαριστίες

Αρχικά, θα ήθελα να ευχαριστήσω τον καθηγητή μου Πέτρο Μαραγκό, για την βοήθεια που μου προσέφερε καθ'όλη την διάρκεια της διπλωματικής εργασίας και την εμπειρία που αποκόμισα στα πρώτα βήματα της ερευνητικής μου καριέρας. Ήδη από τα πρώτα χρόνια μου σαν προπτυχιακός φοιτητής είχα εντυπωσιαστεί από το αντικείμενο και τον τρόπο διδασκαλίας του, γεγονός που με οδήγησε να εκπονήσω διπλωματική έχοντάς τον σαν επιβλέποντα. Τον ευχαριστώ ιδιαίτερα για την επιλογή του να μου προτείνει αυτό το θέμα, που ταίριαζε στα ενδιαφέροντά μου και διεύρυνε τους ορίζοντες μου μέσα από την διερεύνηση ποικίλων τομέων ακόμη και έξω από την Επιστήμη των Υπολογιστών. Με την εξαιρετική του διαίσηση σε κάποια τεχνικά ζητήματα και τις προτάσεις του, με την παρακίνησή του και την κατανόησή που έδειξε, συνεισέφερε τα μέγιστα στην ολοκλήρωση της διπλωματικής. Τέλος, τον ευχαριστώ για τις συμβουλές που μου έδωσε είτε μέσα από τα μαθήματά του, είτε σε προσωπικό επίπεδο· συμβουλές πολύτιμες για την μετέπειτα ερευνητική και όχι μόνο, πορεία μου.

Ακόμη, θα ήθελα να ευχαριστήσω τους φίλους μου για όλες τις φορές που ήταν εκεί για μένα. Πολλές φορές υποτιμάμε την αξία ενός διαλείμματος, της φυγής από την καθημερινότητα και την σημασία του να έχεις κάποιον να μοιράζεσαι τις ανησυχίες σου, τα ερωτηματικά σου, αλλά και τον ενθουσιασμό σου για τα επόμενα. Το ρόλο αυτό παίζανε οι φίλοι μου, και η βοήθεια που μου προσέφεραν ήταν απεριόριστη ακόμη κι αν δεν το καταλάβαιναν πάντα.

Τέλος, ιδιαίτερη αναφορά θα ήθελα να κάνω στα αδέρφια μου και τους γονείς μου για την κατανόηση και την απεριόριστη στήριξή τους, όχι μόνο την περίοδο της διπλωματικής, αλλά καθ'όλη την πορεία της ζωής μου. Για τους γονείς μου, για τον τρόπο που με μεγάλωσαν, τις ευκαιρίες που μου έδωσαν και την στήριξή τους στις επιλογές μου και τα αδέρφια μου, γιατί είναι και πάντα θα είναι η ομάδα μου.

Ευάγγελος Χατζηπανταζής,  
Οκτώβριος 2018



# Περιεχόμενα

Περίληψη	i
Abstract	iii
Ευχαριστίες	v
Περιεχόμενα	viii
Κατάλογος Σχημάτων	ix
<b>1 Εισαγωγή</b>	<b>1</b>
1.1 Η αναγκαιότητα επεξεργασίας των δεδομένων στην δομή του γράφου . . . .	1
1.2 Η κατεύθυνση της διπλωματικής εργασίας . . . . .	2
<b>2 Εισαγωγή στη Φασματική Θεωρία Γράφων</b>	<b>5</b>
2.1 Εισαγωγή στους Γράφους Ομοιότητας . . . . .	5
2.1.1 Ορισμός Γράφου Ομοιότητας . . . . .	5
2.1.2 Κατασκευή γράφου ομοιότητας σε point cloud data . . . . .	7
2.1.3 Κατασκευή γράφων σε εικόνες: SLIC-RAG graph . . . . .	15
2.2 Εισαγωγή στη Φασματική Θεωρία Γράφων . . . . .	25
2.2.1 Τελεστές επί των κόμβων . . . . .	25
2.2.2 Graph Laplacians . . . . .	27
<b>3 Φασματική κατάτμηση γράφων με την ελαχιστοποίηση σταθμισμένων τομών</b>	<b>35</b>
3.1 Εισαγωγή . . . . .	35
3.2 Φασματική κατάτμηση γράφου σε 2 κλάσεις . . . . .	38
3.2.1 Μοντελοποίηση προβλήματος . . . . .	38
3.2.2 Σύγκριση με άλλους αλγόριθμους φασματικής κατάτμησης . . . . .	41
3.2.3 Επαναφορά στον διακριτό χώρο . . . . .	42
3.2.4 Υπολογιστική πολυπλοκότητα και αλγόριθμοι . . . . .	44
3.2.5 Ένα σχόλιο για την δεύτερη και τις μεγαλύτερες ιδιοτιμές . . . . .	46
3.3 Επέκταση φασματικής κατάτμησης σε πολλές κλάσεις . . . . .	50

3.3.1	Μοντελοποίηση Προβλήματος σε $K$ κλάσεις . . . . .	50
3.3.2	Επίλυση για $K > 2$ κλάσεις . . . . .	53
3.3.3	Διαφορετικά <i>Rounding Scemes</i> . . . . .	57
3.4	Εφαρμογή και πρακτικά χαρακτηριστικά . . . . .	62
3.5	Προβλήματα που δημιουργούνται και μια πρόταση για να λυθούν . . . . .	65
<b>4</b>	<b>Επέκταση των σταθμισμένων τομών με τεχνικές ημειπιβλεπόμενης μάθησης</b> . . . . .	<b>69</b>
4.1	Εισαγωγή a-priori γνώσης για την βελτίωση του clustering . . . . .	69
4.2	Εισαγωγή a-priori πληροφορίας στον αλγόριθμο Normalized Cuts . . . . .	70
4.2.1	Τοποθέτηση 2 κόμβων στο ίδιο cluster (Must Link Constraint) . . . . .	71
4.2.2	Τοποθέτηση ενός κόμβου σε ένα συγκεκριμένο cluster ( <i>Unary Constraints</i> ) . . . . .	91
4.2.3	Τοποθέτηση 2 κόμβων σε 2 διαφορετικά clusters. ( <i>Cannot Link Constraint</i> ) . . . . .	99
4.2.4	Παράδειγμα επίδρασης ML και CL περιορισμών σε εικόνες . . . . .	103
4.2.5	Άλλοι Περιορισμοί . . . . .	106
<b>5</b>	<b>Επίλογος</b> . . . . .	<b>109</b>
5.1	Συνεισφορές της Διπλωματικής . . . . .	109
5.2	Μελλοντικές επεκτάσεις της διπλωματικής . . . . .	110

# Κατάλογος Σχημάτων

2.1	Απεικονίζουμε 3 κατανομές. Οι πρώτες 2 είναι μισοφέγγαφα με το κάτω να έχει μεγαλύτερη πυκνότητα από το πάνω ενώ η τρίτη είναι μια <i>Gaussian</i> .	10
2.2	Εξάρτηση από την παράμετρο: Or-Knn, And-Knn, e-radius	12
2.3	Clustering με Normalized Cut	14
2.4	Watershed Superpixels and Centroids on Image and Image Gradient	18
2.5	SLIC parameters	20
2.6	bird image. RAG-SLIC graph and Normalized Cut segmentation	21
2.7	bird image. Clusters extended to pixels	22
2.8	baby segmentation	23
2.9	Size Parameter. Smaller superpixels is not always a better choice.	24
2.10	Σύγκριση ποιότητας και χρόνου μεταξύ των δύο μεθόδων κατασκευής super-pixel. Η πρώτη μέθοδος εκτός από το ότι παρουσιάζει εμφανώς καλύτερα αποτελέσματα, είναι και πολύ πιο γρήγορη. Ο σχηματισμός του γράφου πήρε 0.3 δευτερόλεπτα και για τις δύο, ενώ στο πρώτο η τομή πήρε 0.005 δευτερόλεπτα και στο δεύτερο 0.05 δευτερόλεπτα.	25
2.11	Απεικονίζουμε το ίδιο σήμα σε 3 διαφορετικούς γράφους, για να καταδείξουμε ότι η ομαλότητα του εξαρτάται και από την υποβόσκουσα δομή. Σχήμα από [38]	29
3.1	Κατάτμηση ομόκεντρων κύκλων	36
3.2	Κατάτμηση διαπλεκόμενων σπειρών με Normalized Cuts	36
3.3	Αλγόριθμος κατωφλιοποίησης γράφου	37
3.4	Min Cut versus Min Normalized Cut	38
3.5	Ο αλγόριθμος του Kernel Kmeans	48
3.6	Embedding των σημείων σε ένα χώρο τριών διαστάσεων.	58
3.7	Προβολή σημείων σε ένα χώρο ιδιοδιανυσμάτων και κατάτμησή τους με χρήση <i>k - means</i>	58
3.8	Αλγόριθμος Yu-Shi	61
3.9	Αλγόριθμος <i>Yu, Shi</i> : Κατάτμηση σημείων	62
3.10	Κατάτμηση με χρήση <i>intervening contours</i>	63
3.11	Χρωματική κατάτμηση σε πολλές κλάσεις	63

3.12	Κατάτμηση αντικειμένων με πολύ διαφορετική δομή. Ο αλγόριθμος σταθμισμένων τομών δεν κάνει καμία υπόθεση για το σχήμα των clusters. . . . .	64
3.13	Η μεταβολή του <i>eigen gap</i> καθώς αυξάνονται τα επίπεδα θορύβου . . . . .	66
3.14	Η αδυναμία του Normalized Cuts να διαχωρίσει elongated clusters και η ευαισθησία του σε <i>outliers</i> . . . . .	67
3.15	Το <i>average gap</i> [31] . . . . .	67
3.16	Η ανάγκη για κατάτμηση αντικειμένων σε πολλαπλές κλίμακες . . . . .	68
4.1	Χρειάζεται η ενσωμάτωση εξωγενούς πληροφορίας στο πρόβλημα της κατάτμησης, γιατί μόνο έτσι μπορεί να διαχωριστεί η αντανάκλαση του βράχου και να ενοποιηθεί η κλάση του ανθρώπου. Αλλιώς, μια ιεραρχική κατάτμηση είναι αρκετά ικανοποιητική. . . . .	70
4.2	Δίκτυο Πρωτεϊνών. Ιδανικά, θα θέλαμε ένα εξωτερικό ειδήμονα ή ένα προ-στάδιο του αλγορίθμου, να επιβάλλει στους όμοια χρωματισμένους κόμβους του μακρομορίου να ανήκουν στην ίδια κλάση . . . . .	72
4.3	Γράφος περιορισμών πριν (κόκκινο) και μετά (μπλε) την ελαχιστοποίηση . . . .	84
4.4	Ελαχιστοποίηση πίνακα περιορισμών. . . . .	84
4.5	<i>Reduced</i> γράφος (ή γράφος συνεκτικών συνιστωσών) . . . . .	85
4.6	Πρώτο συνθετικό παράδειγμα. Βλέπουμε την αδυναμία του unsupervised αλγορίθμου να εντοπίσει το Ground Truth Segmentation . . . . .	86
4.7	Η αποτελεσματικότητα του αλγορίθμου φαίνεται στο γεγονός ότι μόνο 3 <i>ML</i> περιορισμοί βελτιώσαν αισθητά την κατάτμηση. . . . .	87
4.8	Με 20 περιορισμούς πετυχαίνουμε την πιό καλή κατάτμηση. . . . .	88
4.9	Έχουμε επιβάλλει 200 τυχαίους <i>ML</i> περιορισμούς ενώ έχουμε 10000 δείγματα συνολικά (Δηλαδή υπάρχουν 10000-3 γραμμικά ανεξάρτητοι περιορισμοί). . . . .	88
4.10	Επίδοση και χρόνος συναρτήσει του αριθμού των ανεξάρτητων περιορισμών. Τα ποσοστά έχουν υπολογιστεί με βάση τους μέγιστους ανεξάρτητους περιορισμούς που είναι (#Κομβων)-(#cluster). . . . .	89
4.11	Η κατάτμηση χειροτερεύει γιατί ο <i>ML</i> περιορισμός είναι πολύ αδύναμος από μόνος του για να τραβήξει όλα τα λάθος καταναμημένα δεδομένα στην πρώτη κλάση. . . . .	90
4.12	Παρότι προσθέτουμε πολλούς <i>ML</i> περιορισμούς, επειδή δεν είναι στοχευμένοι τελικά η κατάτμηση χειροτερεύει. . . . .	90
4.13	Αποτελεσματικότητα των αρνητικών βαρών στο πρόβλημα της κατάτμησης . . . . .	101
4.14	Εφαρμογή του αλγορίθμου <i>ML</i> , <i>CL</i> <i>NCut</i> σε εικόνες . . . . .	105

# Κεφάλαιο 1

## Εισαγωγή

### 1.1 Η αναγκαιότητα επεξεργασίας των δεδομένων στην δομή του γράφου

Τεράστιο είναι το στοίχημα της επιστημονικής κοινότητας να δημιουργήσει αυτοματισμούς για όλες τις ανθρώπινες λειτουργίες. Ίσως το μεγαλύτερο βάρος αυτήν την περίοδο, πέφτει στην κατανόηση και κατ'επέκταση στην αναπαραγωγή των γνωστικών λειτουργιών του ανθρώπου και πιο συγκεκριμένα στο πώς αντιλαμβάνεται αυτά που βλέπει, πώς αναγνωρίζει μοτίβα και αντικείμενα με σημασιολογικό περιεχόμενο, ιεραρχικά από μικρότερες δομές. Η ταυτόχρονη εργασία γνωστικών ψυχολόγων, νευροβιολόγων κλπ. αποτελεί εφιαλτήριο για την επιστήμη της Όρασης Υπολογιστών που ασχολείται με τη δημιουργία τεχνητών νοημοσυνών, που εκτελούν τις λειτουργίες αυτές αυτόματα. Ένα από τα λιγότερο καλά θεμελιωμένα προβλήματα, που καλείται να αντιμετωπίσει η Όραση Υπολογιστών είναι η αφηρημένη μοντελοποίηση ιδιοτήτων για μη επιβλεπόμενη κατάτμηση εικόνων, με στόχο τη μετέπειτα αναγνώριση αντικειμένων. Σε αυτό το πρόβλημα ακόμη και ο άνθρωπος παρουσιάζει ασυνέπειες, καταλήγοντας πολλές φορές με διαφορετικές κατατμήσεις ακόμη και για την ίδια εικόνα. Επιπλέον, στο πρόβλημα αυτό, δεν μας δίνεται ένα ground truth για να συγκρίνουμε την απόδοση του αλγορίθμου μας. Στα πλαίσια αυτής της προσπάθειας, η αναπαράσταση της εικόνας με χρήση pixel έχει αποδειχθεί μη επαρκής και ως προς το κόστος των αλγορίθμων που πραγματοποιούν την κατάτμηση και ως προς την απόδοσή τους.

Η πιο διαδεδομένη δομή για να υλοποιήσει κανείς τέτοιου είδους σχεσιακές ιδιότητες, είναι η δομή του γράφου, που δίνει τεράστια ευελιξία και συνδέεται με μια πολύ καλά θεμελιωμένη θεωρία. Στα προβλήματα μάλιστα της κατάτμησης, οι γράφοι ανακύπτουν σε πολλούς τομείς και εφαρμογές, όπως το community detection σε κοινωνικά δίκτυα ή η διάσπαση πρωτεϊνών σε αντίστοιχα δίκτυα, η αυτόματη ομαδοποίηση κειμένων, ταινιών κλπ., ανάλογα με το περιεχόμενό τους και άλλα. Στον αντίποδα της κατασκευής γράφων για την κατάτμηση δεδομένων βρίσκεται η κατασκευή ανθεκτικών σε κατατμήσεις γράφων, τομέας με τον οποίο ασχολούνται πολύ η επιστήμη των Δικτύων Υπολογιστών και η Επιστήμη των Τηλεπικοινωνιών. Η κατασκευή αραιών γράφων, που κρατάνε την περισσότερη πληροφο-

ρία από το πρόβλημα, είναι τελικά ένα υπό έρευνα ζήτημα, που αφορά όλους σχεδόν τους κλάδους της Επιστήμης των Υπολογιστών. Από την άλλη, η πληθώρα νέων εφαρμογών και η ραγδαία αύξηση της πολυπλοκότητας και της διάστασης των υπαρχόντων σε μεγάλο εύρος επιστημών, οδηγεί σε μια γενική τάση αυτοματισμού της επεξεργασίας όλων αυτών των δεδομένων, συμπίεσής τους, οπτικοποίησης τους και τελικά συλλογής μόνο των σημαντικότερων από αυτά για αποθήκευσή τους. Δεδομένου ότι οι προϋπάρχοντες αλγόριθμοι δεν εκμεταλλεύονται τις σημασιολογικές σχέσεις των αντικειμένων ή την υποβόσκουσα δομή τους και χρησιμοποιούν διανύσματα χαρακτηριστικών για να τα περιγράψουν, συχνά δεν έχουν το επιθυμητό αποτέλεσμα σε εφαρμογές κατάτμησης. Συνεπώς, συμπεραίνουμε την αναγκαιότητα της δημιουργίας αλγορίθμων για συσταδοποίηση δεδομένων, που βρίσκονται πάνω σε έναν γράφο.

Παρόλη την προσπάθεια για δημιουργία μη επιβλεπόμενων αλγορίθμων κατάτμησης, αρκετές φορές τα αποτελέσματα δεν είναι ικανοποιητικά. Συνήθως, σε εικόνες χρησιμοποιούνται ως προστάδιο για μεταγενέστερη κατάτμηση ή αναγνώριση των αντικειμένων της εικόνας. Ωστόσο, επειδή είναι αδύνατο να κωδικοποιηθεί η σημασιολογία όταν υπάρχουν πολλές λογικές κατατμήσεις, φαινόμενο το οποίο στην βιβλιογραφία απαντάται ως *cluster ambiguity*, για να βελτιώσουμε την απόδοση των αλγορίθμων θα πρέπει να ενσωματώσουμε στο πρόβλημα της μη-επιβλεπόμενης κατάτμησης, κάποιο είδος αυτόματης μάθησης. Καθώς είμαστε στην εποχή της πληροφορίας θα ήταν σπατάλη να μην εκμεταλλευτούμε τα δισεκατομμύρια εικόνων και βίντεο που ανεβαίνουν καθημερινά σε πλατφόρμες όπως το Youtube και το Facebook. Αυτή τη στιγμή ο πυρετός της μηχανικής μάθησης και των νευρωνικών δικτύων χρησιμοποιεί πληθώρα αυτών των δεδομένων για αποτελεσματική εκπαίδευση και ρύθμιση των παραμέτρων, με πολύ επαναστατικά αποτελέσματα. Ωστόσο σε άλλου τύπου εφαρμογές, τα επιβλεπόμενα παραδείγματα είναι δυσεύρετα, αλλά ακόμη και για εικόνες και βίντεο η διαδικασία του *annotation* είναι πολύ χρονοβόρα, κοστοβόρα και ακυρώνει το στόχο για συνολικό αυτοματισμό της διαδικασίας. Αρκετές φορές μάλιστα ακόμη και για τις πιο συνηθισμένες εφαρμογές, τα δεδομένα δεν επαρκούν, ειδικά όταν πρέπει να ρυθμιστούν ταυτόχρονα εκατομμύρια παράμετροι. Για αυτό το λόγο, μεγάλη σημασία δίνεται τα τελευταία χρόνια σε τεχνικές ημιεπιβλεπόμενης μάθησης, όπου τα δεδομένα δεν χρησιμοποιούνται άμεσα για την εκπαίδευση, γιατί είναι πολύ λίγα συγκριτικά με τα δεδομένα ελέγχου, αλλά μέσα από τα παρατηρημένα και μη παρατηρημένα δεδομένα εξάγουμε πληροφορία απευθείας για τις κλάσεις και την μορφή των τελευταίων.

## 1.2 Η κατεύθυνση της διπλωματικής εργασίας

Στην συγκεκριμένη διπλωματική εργασία θα ασχοληθούμε με μια μέθοδο με εφαρμογές στην κατάτμηση γράφων, που αφορμάται από τις ιδιότητες του φάσματος ενός γράφου. Δεδομένης της ραιικής κωδικοποίησης που μπορεί να επιβάλλει η δομή του γράφου, θα ελέγξουμε την αποτελεσματικότητα και την ταχύτητα αλγορίθμων φασματικών γραφοτομών σε προβλήματα εικόνων και *point cloud* δεδομένων.

Στο δεύτερο κεφάλαιο θα εισάγουμε την έννοια του φάσματος του γράφου και των ε-



φαρμογών του σε συσταδοποίηση δεδομένων και κατάτμηση εικόνων. Θα ασχοληθούμε ιδιαίτερα με τον μετασχηματισμό της αναπαράστασης των εικόνων από την κανονική δομή του ορθοκανονικού πλέγματος και των pixel, στην προγενέστερη ομαδοποίησή τους σε superpixel και την πρόσδεση των τελευταίων στις κορυφές ενός γράφου. Μετά από αρκετό πειραματισμό, παρατηρήσαμε ότι η ιεραρχική αυτή απεικόνιση των δεδομένων είναι πιο αποτελεσματική και γενικεύεται καλύτερα σε πλήθος εικόνων αλλά και σε point cloud δεδομένα. Ακόμη, θα αναλύσουμε τη συνδεσιμότητα αυτών των κόμβων, που αποτελεί παράμετρο με τεράστια επίδραση στην τελική κατάτμηση.

Στο επόμενο κεφάλαιο, θα θεμελιώσουμε καλύτερα το πρόβλημα των σταθμισμένων γραφοτομών και θα το λύσουμε με μεθόδους φασματικής κατάτμησης γράφων. Θα αναλύσουμε την γενικότερη μορφή του προβλήματος, τον προσεγγιστικό αλγόριθμο για την επίλυσή του και θα τον αντιστοιχίσουμε με γνωστές μεθόδους κατάτμησης για να κερδίσουμε διαίσθηση πάνω στην λειτουργία του. Εφαρμόζοντάς τον σε προβλήματα εικόνων, θα βελτιώσουμε την ταχύτητα και την απόδοση του κλασσικού αλγορίθμου φασματικών γραφοτομών μέσα από την επιβολή της ιεραρχίας του κεφαλαίου 2.

Τέλος, στο κεφάλαιο 4 θα αναπτύξουμε θεωρητικά και πρακτικά ένα ημι-επιβλεπόμενο framework για την εσωμάτωση εξωτερικής πληροφορίας στο πρόβλημα των φασματικών γραφοτομών. Η ανάγκη για λειτουργίες μάθησης από λίγα δεδομένα θα παρουσιαστεί μέσα από εφαρμογές must link και cannot link περιορισμών αλλά και άλλων unary περιορισμών που θα ενισχύσουν την ποιότητα της κατάτμησης και θα δώσουν ευελιξία στον χρήστη ανάλογα με την εφαρμογή.

Στο τελευταίο κεφάλαιο αναλύουμε τις μελλοντικές επεκτάσεις της παρούσας διπλωματικής.



## Κεφάλαιο 2

# Εισαγωγή στη Φασματική Θεωρία Γράφων

### 2.1 Εισαγωγή στους Γράφους Ομοιότητας

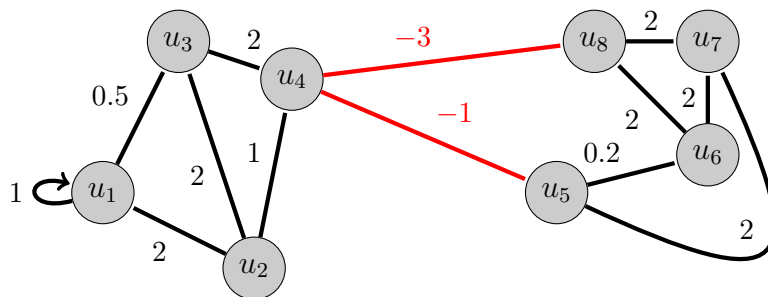
#### 2.1.1 Ορισμός Γράφου Ομοιότητας

Συνήθως οι εφαρμογές της κατάτμησης (σε εικόνες, βίντεο ή γενικότερα δεδομένα όπως ένα δίκτυο πρωτεϊνών) έχουν ως στόχο τη δημιουργία συστάδων με μεγάλη εσωτερική ομοιότητα και μεγάλη εξωτερική ανομοιότητα. Η *ομοιότητα* εξ' ορισμού υποδηλώνει μια σχέση μεταξύ δύο ή περισσότερων αντικειμένων και ποσοτικοποιείται, χρησιμοποιώντας τα μεμονωμένα χαρακτηριστικά (data attributes) κάθε αντικειμένου και ένα μέτρο της απόστασης μεταξύ τους. Εναλλακτικά, μπορεί να μην γνωρίζουμε τα χαρακτηριστικά κάθε ξεχωριστού αντικειμένου, αλλά να έχουμε πρόσβαση σε πληροφορία που δηλώνει, το κατά πόσο αυτό το αντικείμενο σχετίζεται με κάποιο άλλο (pairwise relations). Τέτοιο παράδειγμα είναι η πληροφορία που παίρνουμε συνήθως από ένα κοινωνικό δίκτυο. Στην εργασία αυτή θα ασχοληθούμε με μεθόδους που χρησιμοποιούν πληροφορία για ζεύγη αντικειμένων. Η απεικόνιση αυτής της πληροφορίας μπορεί να γίνει ισοδύναμα μέσα από την δομή του γράφου, μέσα από πίνακες ομοιότητας ή μέσα από διμελείς σχέσεις. Άλλοι αλγόριθμοι κατάτμησης, όπως ο *k-means* εφαρμόζονται απευθείας σε διανύσματα, αλλά εδώ θα πρέπει να βρούμε έναν τρόπο να ορίσουμε την απόσταση των χαρακτηριστικών, ώστε να απεικονίζουμε τα διανύσματα σαν ομοιότητα ζευγών.

**Ορισμός 2.1. (Graph)** Ως κατευθυνόμενος γράφος ορίζεται το ζεύγος  $\mathcal{G} = (V, E)$ , όπου  $V = \{u_1, \dots, u_N\}$  είναι ένα σύνολο από κορυφές ή κόμβους, ενώ  $E \subseteq V \times V$  είναι ένα σύνολο από διατεταγμένα ζεύγη κόμβων, που ονομάζονται ακμές.

**Ορισμός 2.2. (Affinity Matrix)** Σε κάθε γράφο  $N$  κόμβων αντιστοιχεί ένας τετραγωνικός πίνακας (γειτνίασης)  $A$  μεγέθους  $N \times N$  που έχει την τιμή 1 στην θέση  $(i, j)$  αν η ακμή  $(u_i, u_j)$  ανήκει στο σύνολο  $E$  και 0, αλλιώς.

Τώρα που ορίσαμε την έννοια του γράφου, μπορούμε να απεικονίσουμε διμελείς σχέσεις, χρησιμοποιώντας τους κόμβους για τα δεδομένα και τις ακμές για τις μεταξύ τους συνδέσεις. Οι κατευθυνόμενοι γράφοι υπονοούν, ότι η σχέση των κόμβων μπορεί να μην είναι συμμετρική (π.χ. σχέση εξάρτησης). Επιπλέον, μπορούμε να εισάγουμε ένα βάρος για κάθε ακμή που θα κωδικοποιεί το μέγεθος της ομοιότητας ή της απόστασης κλπ. μεταξύ δύο κόμβων. Αυτό μπορεί να είναι και αρνητικό. Οι τιμές αυτές θα αποθηκεύονται σε ένα πίνακα βαρών  $W = [w(u_i, u_j)]$  αντίστοιχο του πίνακα γειτνίασης. Ως γράφους ομοιότητας θα ορίσουμε τους γράφους, των οποίων οι ακμές κωδικοποιούν μια συμμετρική σχέση και τα βάρη αντικατοπτρίζουν την ομοιότητα των κόμβων και συνεπώς είναι μη-αρνητικά. Οι γράφοι αυτοί είναι μη-κατευθυνόμενοι, που σημαίνει ότι  $W^T = W$ . Επιτρέπουμε και ακμές ανακύκλωσης στους γράφους αυτούς. Μια ακόμη έννοια που θα συναντήσουμε, είναι αυτή της *συνεκτικής συνιστώσας* ενός γράφου, που είναι το σύνολο των κόμβων που μπορούμε να επισκεφτούμε ξεκινώντας από έναν και ακολουθώντας διαδοχικά τις ακμές του γράφου. Αν μπορούμε να προσπελάσουμε όλους τους κόμβους του γράφου, τότε ο γράφος λέγεται *συνεκτικός*. Τέλος, ορίζουμε και τον *βαθμό* ενός κόμβου, σαν το άθροισμα των βαρών όλων των εξερχόμενων (ισοδ. εισερχόμενων) ακμών του (και της ακμής ανακύκλωσης). Αν  $d_i = \sum_{j \sim i} w_{ij}$ , τότε ονομάζουμε τον διαγώνιο πίνακα  $D = \text{diag}(d_1, \dots, d_N)$ , πίνακα βαθμών. Τα παραπάνω οπτικοποιούνται στο επόμενο παράδειγμα.



Οι πίνακες που περιγράφουν τον συνεκτικό γράφο του σχήματος είναι:

$$W = \begin{bmatrix} 1 & 2 & 0.5 & 0 & 0 & 0 & 0 & 0 \\ 2 & 0 & 2 & 1 & 0 & 0 & 0 & 0 \\ 0.5 & 2 & 0 & 2 & 0 & 0 & 0 & 0 \\ 0 & 1 & 2 & 0 & -1 & 0 & 0 & -3 \\ 0 & 0 & 0 & -1 & 0 & 0.2 & 2 & 0 \\ 0 & 0 & 0 & 0 & 0.2 & 0 & 2 & 2 \\ 0 & 0 & 0 & 0 & 2 & 2 & 0 & 2 \\ 0 & 0 & 0 & -3 & 0 & 2 & 2 & 0 \end{bmatrix}, \quad D = \begin{bmatrix} 3.5 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 5 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 4.5 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1.2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 4.2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 6 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

### 2.1.2 Κατασκευή γράφου ομοιότητας σε point cloud data

Συχνά στην πράξη δεν μας δίνονται τα δεδομένα με την μορφή γράφου, αλλά σαν σημεία στο χώρο. Σε αυτή την περίπτωση πρέπει να κατασκευάσουμε εμείς το γράφο ομοιότητας, αποφασίζοντας ποιοί κόμβοι θα συνδέονται με ακμή και τί βάρος θα έχει αυτή η ακμή. Προφανώς, η κατασκευή ενός καλού γράφου είναι άρρηκτα συνυφασμένη με την εκάστοτε εφαρμογή, ωστόσο μπορούμε να πούμε με σιγουριά, ότι για προβλήματα συσταδοποίησης, θα θέλαμε ο γράφος να έχει πολλές ακμές μεγάλου βάρους μέσα σε κάθε κλάση και λίγες (ει δυνατόν καθόλου) ακμές μεταξύ των κλάσεων. Επιπλέον, θα θέλαμε ο γράφος να είναι συνεκτικός, έστω σε κάθε κλάση, και οι ακμές να είναι τόσο πυκνές, ώστε να εκμεταλλεύονται πλήρως την τοπική πληροφορία, αλλά και αρκετά αραιές για να μην εισάγουν θόρυβο και χρονική πολυπλοκότητα στο πρόβλημα. Προφανώς, ο ιδανικός γράφος θα ήταν ένα σύνολο από ανεξάρτητα μονοπάτια, ένα σε κάθε κλάση. Το πρόβλημα είναι, ότι και η κατασκευή του γράφου είναι *unsupervised*, οπότε δεν μπορούμε να ξέρουμε εξ' αρχής την ιδανική μορφή του γράφου. Θα προσπαθήσουμε όμως να κατασκευάσουμε γράφους, που θα προσεγγίζουν όσο γίνεται τις παραπάνω απαιτήσεις. Βιβλιογραφικά αναφέρουμε, ότι υπάρχουν learning τεχνικές για κατασκευή του γράφου από δεδομένα ([18],[51]), αλλά δεν θα ασχοληθούμε με αυτό το ζήτημα στη συγκεκριμένη εργασία.

#### Πώς θα υπολογίσουμε το βάρος μια ακμής;

Έστω, ότι έχουμε δύο κόμβους  $u_i, u_j$  με διανύσματα χαρακτηριστικών  $f_i, f_j$  αντίστοιχα. Θέλουμε να ορίσουμε την ομοιότητα των κόμβων από τις αποστάσεις των χαρακτηριστικών τους. Προφανώς περιμένουμε, ότι όσο μειώνεται η απόσταση των  $f_i, f_j$ , τόσο θα αυξάνεται η ομοιότητα τους. Κάποιες δημοφιλείς συναρτήσεις ομοιότητας είναι:

- **RBF kernel** που ορίζεται ως:

$$\text{rbf}(u_i, u_j) = \exp\left(-\frac{\|f_i - f_j\|^2}{2\sigma^2}\right), \quad (2.1)$$

όπου  $\|f_i - f_j\|$  είναι η Ευκλείδεια απόσταση των χαρακτηριστικών (γί αυτό χρησιμοποιείται αποκλειστικά για δεδομένα που ανήκουν σε έναν ευκλείδιο χώρο), ενώ το  $\sigma$  μια ελεύθερη παράμετρος. Οι τιμές της συνάρτησης αυτής, ανήκουν στο διάστημα  $[0,1]$  και φθίνουν εκθετικά με την απόσταση των χαρακτηριστικών. Ανάλογα με την τιμή της ελεύθερης παραμέτρου, μπορεί να μοντελοποιήσει μικρότερες ή μεγαλύτερες γειτονιές και τελικά μπορεί να ερμηνευθεί σαν μέτρο ομοιότητας. Στην πράξη, η

ελεύθερη παράμετρος είναι δύσκολο να υπολογισθεί και συχνά το αποτέλεσμα της συσταδοποίησης είναι ευεπηρεάστο σε μικρές αλλαγές της. Επίσης, δεν είναι απαραίτητο να είναι σταθερή για όλους τους κόμβους (*adaptive*). Έτσι, μπορούμε να μοντελοποιήσουμε γειτονιές μεταβλητού μεγέθους στον ίδιο γράφο ([16]). Σε εικόνες, που τα *pixel* έχουν εκτός από χρωματική πληροφορία και μία θέση στην εικόνα, συχνά χρησιμοποιείται σαν μέτρο ομοιότητας:

$$w(i, j) = \exp\left(-\frac{\|g_i - g_j\|^2}{2\sigma_g^2} - \frac{\|x_i - x_j\|^2}{2\sigma_x^2}\right), \quad (2.2)$$

Τέλος, ο χώρος χαρακτηριστικών αυτού του πυρήνα είναι άπειρης διάστασης, πράγμα που τον καθιστά κατάλληλο για σύνολα δεδομένων με πολύ μη-γραμμικά όρια κλάσεων.

- **Cosine Similarity** που ορίζεται ως:

$$\cos(u_i, u_j) = \frac{f_i \cdot f_j}{\|f_i\| \|f_j\|} \quad (2.3)$$

Οι τιμές της κυμαίνονται μεταξύ -1 και 1 και συχνά χρησιμοποιείται σε εφαρμογές, που δεν μας νοιάζει το μέτρο των χαρακτηριστικών και θέλουμε να μοντελοποιήσουμε την ανομοιότητα δύο κομβών. Τέτοιο παράδειγμα είναι η *συσταδοποίηση κειμένων*, όπου τα χαρακτηριστικά των κόμβων-κειμένων, είναι ένα ιστόγραμμα των λέξεων που περιέχει. Από την άλλη, ο χώρος χαρακτηριστικών του πυρήνα αυτού δεν έχει παραπάνω διαστάσεις από τους κόμβους του γράφου. Επιπλέον, αυτή η μετρική δεν μοντελοποιεί γειτονιές και γι' αυτό στην πράξη, η μοντελοποίηση των γειτονιών επαφίεται στον σχεδιασμό των ακμών.

- **Logistic regression.** Αν έχουμε ένα διάνυσμα χαρακτηριστικών  $z_{ij}$  για κάθε ζευγάρι κόμβων  $(u_i, u_j)$ , τότε κατασκευάζουμε τον γράφο ομοιότητας ως:

$$\log r(z_{ij}) = \frac{1}{1 + \exp(-(\beta + \theta^T z_{ij}))} \quad (2.4)$$

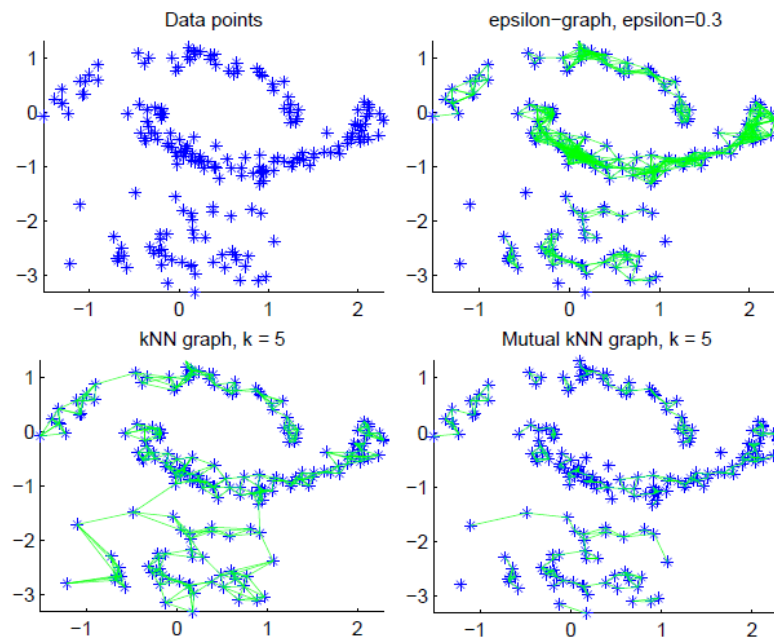
όπου  $\beta, \theta$  είναι ελεύθερες παράμετροι. Οι τιμές αυτής της συνάρτησης κυμαίνονται από 0 (ανόμοιοι κόμβοι) σε 1 (όμοιοι κόμβοι).

Υπάρχουν πολλές ακόμη συναρτήσεις και η λίστα μεγαλώνει ανάλογα με την εφαρμογή. Ειδικά σε εικόνες, αν τα χαρακτηριστικά είναι χρωματική πληροφορία, πληροφορία ακμών, πληροφορία υψής κλπ. αλλάζει ο τρόπος που ορίζουμε τις αποστάσεις μεταξύ των κόμβων. Η άλλη ερώτηση που πρέπει να απαντήσουμε για να ολοκληρώσουμε την κατασκευή του γράφου είναι **ποιές ακμές πρέπει να συμπεριλάβουμε**. Μια συγκριτική μελέτη κάποιων δημοφιλών γράφων για προβλήματα clustering έχει γίνει στο [25].

- **$\epsilon$ -radius graph.** Στον γράφο αυτόν, συνδέονται με ακμή κόμβοι, που έχουν μεταξύ τους απόσταση (ευκλείδεια ή άλλη) μικρότερη από ένα κατώφλι  $\epsilon$ . Κατ' αυτόν τον τρόπο κατασκευάζουμε έναν μη-κατευθυνόμενο γράφο, που μοντελοποιεί την γειτονιά με βάση την παράμετρο  $\epsilon$ . Συχνά, ο γράφος αυτός είναι χωρίς βάρη. Στην πράξη, αν

έχουμε δεδομένα σε διαφορετικές κλίμακες (μεγάλες και μικρές γειτονιές) τότε είναι πολύ δύσκολο να επιλέξουμε την παράμετρο  $\epsilon$ . Όπως φαίνεται στο παρακάτω σχήμα, άλλες περιοχές είναι πολύ συνδεδεμένες και άλλες αραιά συνδεδεμένες για δεδομένη τιμή της παραμέτρου. Αυτός ο ορισμός της γειτονιάς με βάση μια ενιαία γεωμετρική απόσταση, είναι πολύ αδύναμος και γι' αυτό, αν πρόκειται για βεβαρημένο γράφο, συνήθως χρησιμοποιείται μια *RBF* συνάρτηση. Επιπλέον, πρέπει να διαλέξουμε την παράμετρο αυτή, ώστε ο τελικός γράφος να είναι συνεκτικός (ή να έχει πολύ λιγότερες συνεκτικές συνιστώσες από κλάσεις) και καθόλου απομονωμένους κόμβους. Αν τα δεδομένα λαμβάνονταν ανεξάρτητα από μια κατανομή, τότε είναι γνωστό, ότι αν θέσουμε το  $\epsilon = \mathcal{O}((\log(n)/n)^d)$ , όπου  $d$  η διάσταση του χώρου χαρακτηριστικών, εξασφαλίζουμε την συνεκτικότητα, όταν το πλήθος των δειγμάτων τείνει στο άπειρο. Για πεπερασμένα δείγματα όμως, στο [25] προτείνεται, να διαλέξουμε το  $\epsilon$  σαν την μεγαλύτερη ακμή στο ελάχιστο συνεκτικό δέντρο του πλήρους γράφου. Ωστόσο, κατ' αυτόν τον τρόπο, σε περίπτωση που έχουμε *outliers* ή πολύ καλά διαχωρισμένες κλάσεις, η παράμετρος θα γίνει πολύ μεγάλη για να αντικατοπτρίζει την κλίμακα του σημαντικότερου μέρους των δεδομένων. Αυτό φαίνεται και στο σχήμα 2.1

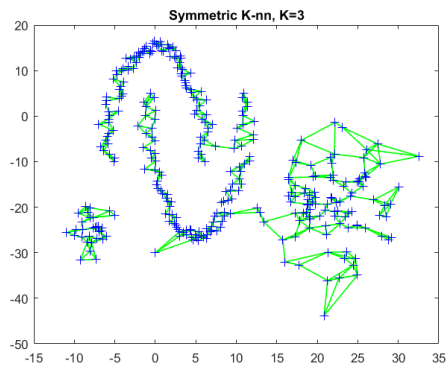
- **K-nn graph.** Στον γράφο αυτόν συνδέουμε τον κόμβο  $u_i$  με τον κόμβο  $u_j$  μόνο αν ο δεύτερος βρίσκεται στους  $K$  πλησιέστερους γείτονες του πρώτου. Εν γένει ο γράφος που δημιουργείται κατ' αυτόν τον τρόπο είναι κατευθυνόμενος γιατί η σχέση της κοτινότερης γειτνίασης δεν είναι συμμετρική. Επίσης, ο γράφος αυτός κωδικοποιεί με *adaptive* τρόπο την έννοια της γειτονιάς αφού η σύνδεση δύο κόμβων δεν σημαίνει απαραίτητα και ισχυρή ομοιότητά τους. Μπορεί με αυτό τον τρόπο, σε αντίθεση με τον  $\epsilon$ -radius graph, να συνδέσει δεδομένα διαφορετικών κλιμάκων. Το παραπάνω είναι εμφανές στο σχήμα. Επιπλέον, όταν έχουμε καλά διαχωρισμένες κλάσεις σπάει σε συνεκτικές συνιστώσες που έχουν μεγάλη σημασία. Και σε αυτή την περίπτωση η συνεκτικότητα του τελικού γράφου εξαρτάται από την τιμή της παραμέτρου  $K$ . Αν δειγματοληπούμε από μια κατανομή ανεξάρτητα δείγματα τότε όταν  $K = \mathcal{O}(\log(n))$  γνωρίζουμε ότι ο γράφος θα είναι συνεκτικός όταν το πλήθος των κόμβων τείνει στο άπειρο. Για να φτιάξουμε τον συμμετρικό γράφο μπορούμε να ακολουθήσουμε 2 δρόμους. Ο ένας είναι να αγνοήσουμε την φορά των ακμών. Αυτός λέγεται (*symmetric*) *K-nn graph*. Ο δεύτερος τρόπος είναι να συμπεριλάβουμε μια ακμή μόνο όταν βρίσκεται και στις δύο κατευθύνσεις στον γράφο. Ο νέος γράφος λέγεται *mutual K-nn graph*. Η πρώτη επιλογή δημιουργεί πιά πυκνό γράφο και ενσωματώνει περισσότερες θορυβώδεις ακμές, όπως φαίνεται στο σχήμα, αλλά κάποιες φορές είναι προτιμότερη από την δεύτερη γιατί η δεύτερη σπάει το γράφο σε πάρα πολλές μικρές συνεκτικές συνιστώσες ομαδοποιώντας μόνο εκείνους τους κόμβους που ανήκουν σε περιοχές πολύ όμοιας πυκνότητας. Σίγουρα, για να εξασφαλίσουμε συνεκτικότητα πρέπει η παράμετρος να είναι πολύ μεγαλύτερη από την αντίστοιχη του (*symmetric*) *K-nn*. Το πλεονέκτημά της όμως είναι, ότι απομονώνει εντελώς *outliers*, πράγμα εμφανές στο σχήμα 2.1.



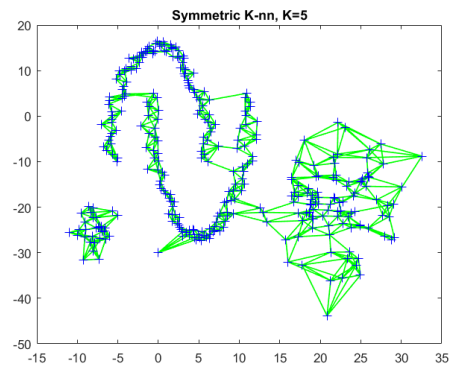
Σχήμα 2.1: Απεικονίζουμε 3 κατανομές. Οι πρώτες 2 είναι μισοφέγγαρα με το κάτω να έχει μεγαλύτερη πυκνότητα από το πάνω ενώ η τρίτη είναι μια *Gaussian*.

Βλέπουμε στο σχήμα 2.1 τη δυσκολία επιλογής ακτίνας στον  $\epsilon$ -radius, αφού με  $\epsilon=0.3$  η μεσοαία κατανομή είναι ήδη πολύ διασυνδεδεμένη, ενώ η *Gaussian* σχεδόν καθόλου. Αυτό το φαινόμενο καταδεικνύει την αδυναμία του  $\epsilon$ -radius να συνδέσει περιοχές διαφορετικής κλίμακας. Από την άλλη, βλέπουμε στον  $K$ -nn, ότι συνδέει επιτυχώς κατανομές που είναι σε άλλη κλίμακα, αλλά μπορεί να σπάσει σε ανεξάρτητες συνεκτικές συνιστώσες, αν πρόκειται για περιοχές μεγάλης πυκνότητας μακριά η μία από την άλλη. Τέλος, βλέπουμε, πως ο mutual  $K$ -nn συνδέει σημεία που ανήκουν σε παρόμοια περιοχή πυκνότητας. Θα λέγαμε ότι βρίσκεται μεταξύ των δύο προηγούμενων, αφού ναι μεν δρα σε κατανομές διαφορετικών κλιμάκων, αλλά δεν τις συνδέει μεταξύ τους.

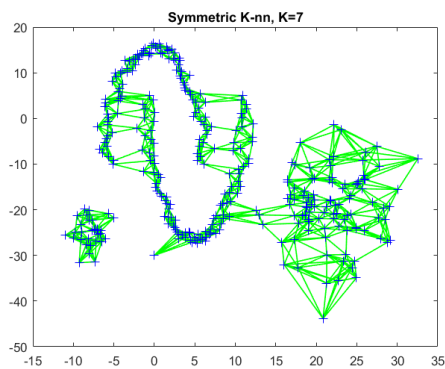




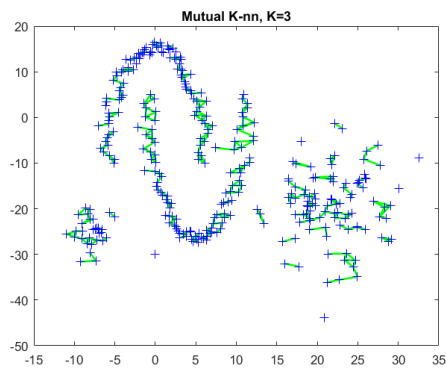
(α) Symmetric 3-nn



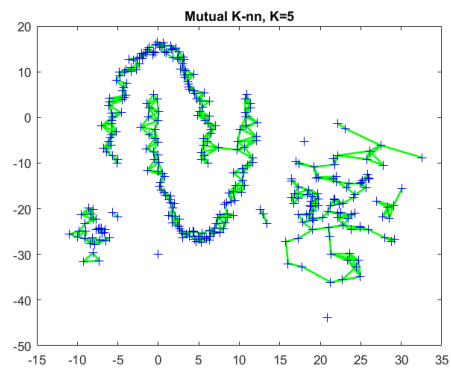
(β) Symmetric 5-nn



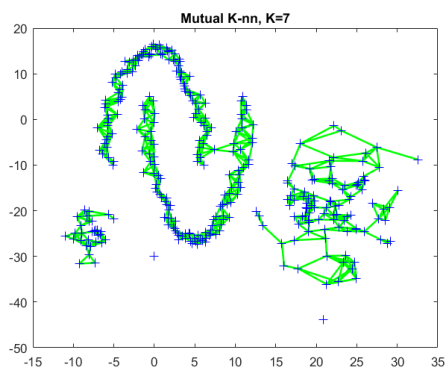
(γ) Symmetric 7-nn



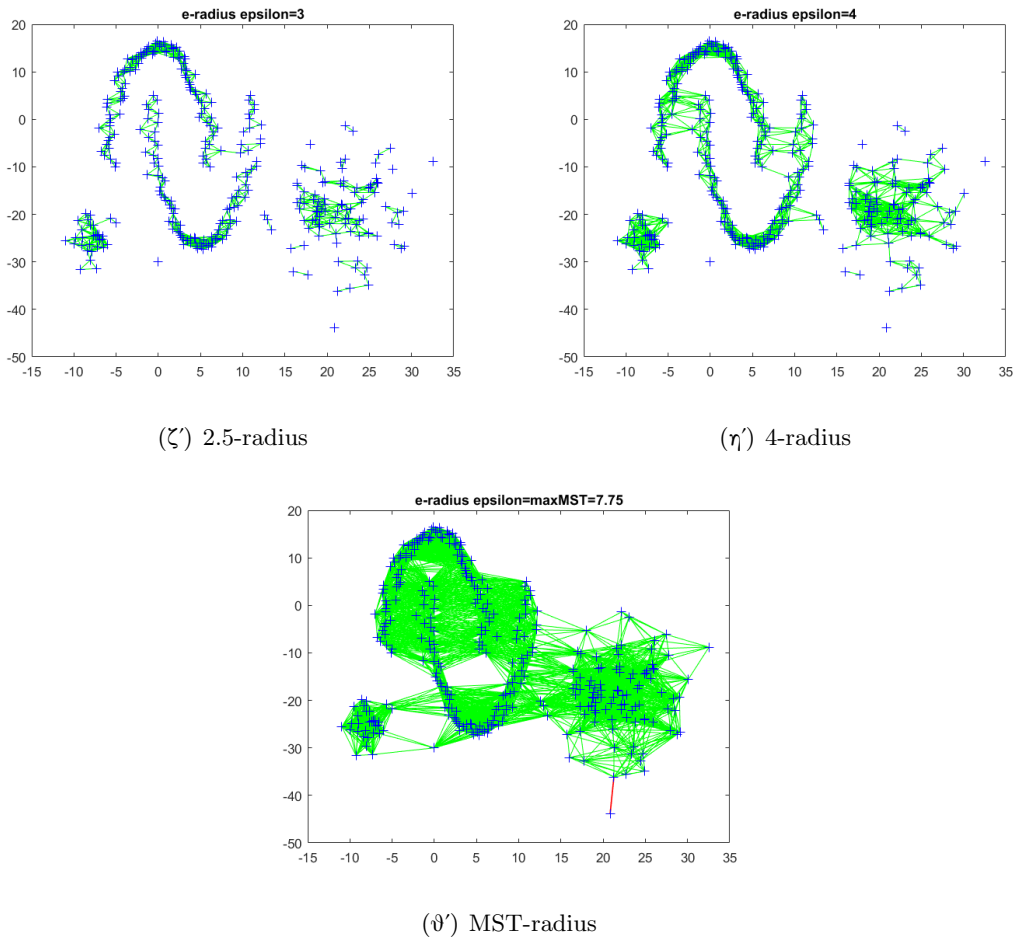
(δ) Mutual 3-nn



(ε) Mutual 5-nn



(ζ) Mutual 7-nn

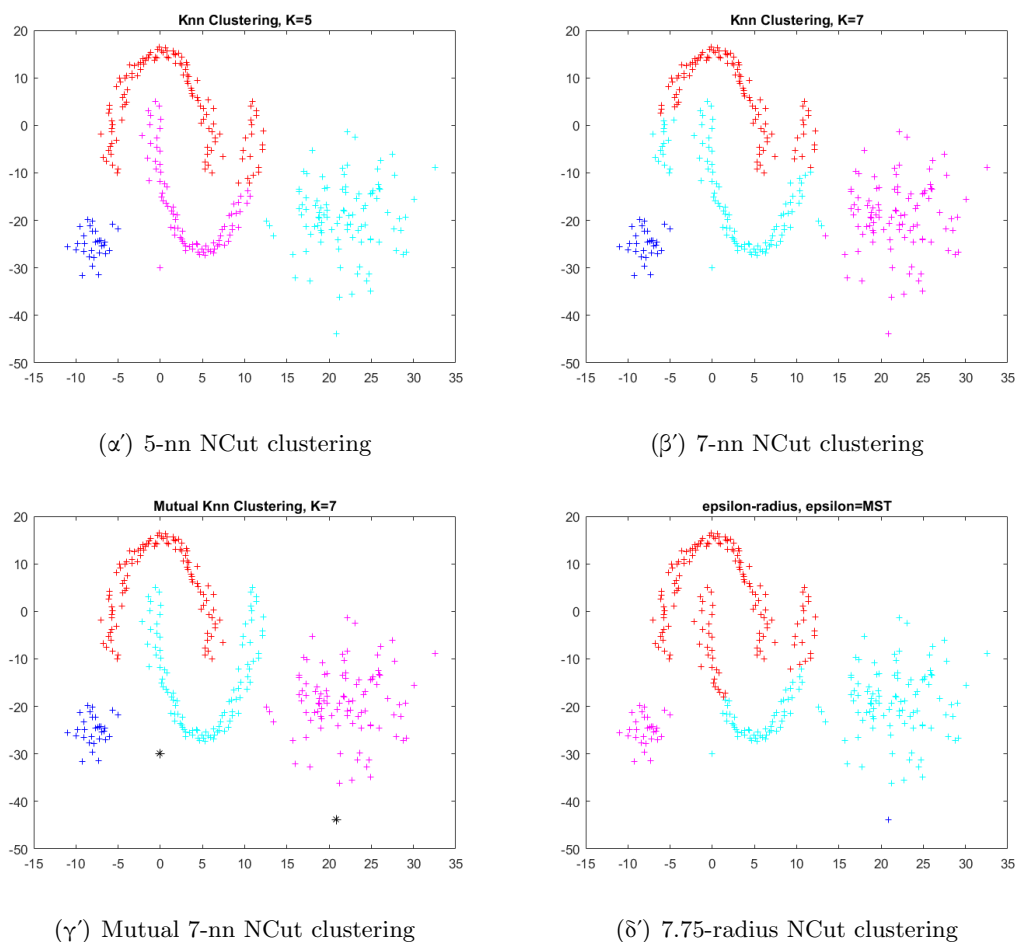


Σχήμα 2.2: Εξάρτηση από την παράμετρο: Or-Knn, And-Knn,  $\epsilon$ -radius

Στο παραπάνω σχήμα (2.2) φαίνονται 4 κατανομές. Οι πρώτες δύο είναι μισοφέγγαρα ίδιας περίπου πυκνότητας και αριθμού δειγμάτων, ενώ η τρίτη και η τέταρτη είναι Gaussian blobs με την τελευταία να έχει μεγαλύτερη διασπορά. Παρατηρούμε, πως ο Symmetric K-nn και ο Mutual K-nn δημιουργούν γειτονιές μεταβλητού μεγέθους συνδέοντας περιοχές διαφορετικών κλιμάκων, ενώ ο  $\epsilon$ -radius συνδέει μόνο κόμβους, που ανήκουν σε σταθερή ακτινική απόσταση. Το μειονέκτημα αυτό φαίνεται στην τελευταία σειρά. Το ένα πρόβλημα που δημιουργείται, είναι η δυσκολία ρύθμισης της παραμέτρου  $\epsilon$ , ώστε ο γράφος να μην έχει πολλούς απομονωμένους κόμβους και να είναι τουλάχιστον συνεκτικός πάνω σε κάθε κατανομή. Το δεύτερο πρόβλημα είναι, ότι ακόμη και όταν το πετυχαίνουμε αυτό (π.χ. όταν διαλέγουμε  $\epsilon$  ίσο με την μέγιστη ακμή του ελάχιστου συνεκτικού δέντρου (με κόκκινο στο σχήμα), η πιθανή ύπαρξη outliers ή αρκετά διαχωρισμένων κατανομών ή μεγαλύτερων και μικρότερων διασπορών, μεγαλώνει τόσο την παράμετρο, ώστε κάποιες γειτονιές είναι πολύ πυκνά συνδεδεμένες μεταξύ τους, ενώ άλλες σχεδόν καθόλου.

Επίσης, αν θεωρήσουμε ως θορυβώδεις τις ακμές που συνδέουν διαφορετικές κατανομές, είναι εμφανές, ότι ο  $\epsilon$ -radius, όπως και ο symmetric K-nn σε μικρότερο βαθμό,

ενσωματώνουν πολύ θόρυβο στη δημιουργία του γράφου. Αντίθετα, ο mutual K-nn εν γένει απομονώνει καλά διαχωρισμένες μεταξύ τους κατανομές και απομονώνει και *outliers*, όσο κι αν αυξάνεται η παράμετρος. Παρόλα αυτά, για πιο μικρές τιμές του K, δημιουργεί πάρα πολλά μικρά *clusters* από περιοχές πολύ παρόμοιας πυκνότητας και αλγόριθμοι γραφοθεωρητικής κατάτμησης, όπως ο *Normalized Cut* δεν μπορούν να ενεργήσουν στον γράφο, αφού έχει πιο πολλές συνεκτικές συνιστώσες από το ζητούμενο αριθμό κλάσεων. Τέλος, από το επόμενο σχήμα (2.3), που δείχνει την ποιότητα της κατάτμησης για κάποιους από αυτούς τους γράφους παρατηρούμε, ότι η ενσωμάτωση θορυβωδών ακμών αποβαίνει καθοριστική για την ποιότητα της κατάτμησης ενώ, μετά την απομόνωση των *outliers* ο mutual K-nn πετυχαίνει την εύρεση των λανθανουσών κατανομών. Γενικά, στην πράξη δεν έχουμε πάντα τη δυνατότητα να απομονώσουμε *outliers* ή μικρά *cluster* και να ενεργήσουμε στα υπόλοιπα στοιχεία (ειδικά αφού ο αριθμός των μικρών *cluster* του mutual K-nn είναι τόσο μεγάλος, που αποτελεί τον μεγαλύτερο όγκο των δεδομένων), οπότε προτιμάμε γράφους, όπως ο symmetric 5-nn, που κάνουν ένα trade-off μεταξύ των inter-cluster και intra-cluster ακμών του γράφου. Στα παρακάτω σχήματα (2.3) παρατηρείται η ανωτερότητα του symmetric 5-nn ως προς τον 7-nn και ιδιαίτερα ως προς τον  $\epsilon$ -radius, που είδαμε στο προηγούμενο σχήμα πόσο περισσότερο θόρυβο ενσωματώνουν.



Σχήμα 2.3: Clustering με Normalized Cut

**Σημείωση:** Σε όλους τους προηγούμενους τρόπους κατασκευής του γράφου θεωρήσαμε ευκλείδειες απόστάσεις των σημείων. Θα μπορούσαμε να κάνουμε το ίδιο με γενικότερες *Mahalanobis* αποστάσεις ή ακόμη και πιο εξεζητημένες αποστάσεις, όπως το *commute time* μεταξύ δύο σημείων ή άλλες που θα λαμβάνουν υπόψη και την γεωμετρία των σημείων στο χώρο. Γενικότερα, θα θέλαμε ιδανικά ο τρόπος κατασκευής του γράφου, όταν πρόκειται για εφαρμογές κατάτμησης, να έχει όσο το δυνατόν μεγαλύτερη συνδεσιμότητα εσωτερικά των κλάσεων και όσο γίνεται λιγότερη εξωτερικά, πράγμα που θα έκανε τετριμμένη την δουλειά οποιουδήποτε αλγορίθμου γραφοθεωρητικής κατάτμησης. Βέβαια, καθώς το πρόβλημα είναι μη-επιβλεπόμενο, δεν γνωρίζουμε από πριν τις κατανομές και επομένως η εύρεση ενός τέτοιου γράφου θα απαιτούσε πρώτα την εύρεση των κλάσεων. Αυτός ο φαύλος κύκλος αντιμετωπίζεται με μια προσέγγιση δύο επιπέδων. Αρχικά, με κάπως αφηρημένο τρόπο απαιτούμε τη δημιουργία γράφων με τις ιδιότητες που αναφέραμε παραπάνω (*adaptive* γειτονίες, όχι μεγάλη ενσωμάτωση θορυβωδών ακμών, αραιότητα κλπ. Στη συνέχεια, με πιο έξυπνους αλγορίθμους κατάτμησης (που λαμβάνουν υπόψη και τη γεωμετρία των κατανομών), χρησιμοποιούμε τον γράφο ως ένα εφαλτήριο για την εύρεση ενός καταλληλότερου χώρου προβολής των δεδομένων, που θα διαχωρίζει καλύτερα τις κλάσεις

μεταξύ τους. Θα ήταν πολύ χρήσιμη μια θεωρητική προσέγγιση στην κατασκευή γράφων, με τη μαθηματική μοντελοποίηση των ιδιοτήτων, που περιγράψαμε παραπάνω. Επιπλέον, ένα βήμα προς την επίλυση αυτού του φαύλου κύκλου είναι, να επαναλαμβάνουμε τη διαδικασία και κάθε στάδιο (δημιουργίας γράφου ή κατάτμησης) να δανείζεται την ανανεωμένη πληροφορία από το προηγούμενο. Θα χρειαστούμε για αυτό το σκοπό ένα *framework* ημι-επιβλεπόμενης μάθησης για αλγορίθμους, όπως ο *Normalized Cut* που θα αναπτύξουμε σε επόμενο κεφάλαιο. Τέλος, η επαναληπτική εναλλαγή μεταξύ της κατασκευής γράφου και της βελτίωσης της συσταδοποίησης μας, θυμίζει σε αφαιρετικό επίπεδο την εκπαίδευση ενός νευρωνικού δικτύου με πολλά *hidden layers*. Κατά την εκπαίδευση του νευρωνικού αυτού, μαθαίνουμε εκτός από τις παραμέτρους -που έχουν ως στόχο την προσαρμογή ενός υπερ-επιπέδου για τον διαχωρισμό των κλάσεων-, και ένα *latent space*, στο οποίο προβάλλονται τα δεδομένα, με τρόπο τέτοιο ώστε να διαχωρίζονται όσο καλύτερα γίνεται από το τωρινό υπερ-επίπεδο. Ιδανικά, *baked-in* στο νευρωνικό θα πρέπει να υπάρχει κάποια παραλλαγή του αλγορίθμου *RANSAC* για απομόνωση των *outliers*. Στην περίπτωση μας, τα δεδομένα προβάλλονται σε ένα χώρο ιδιοδιανυσμάτων (θα δούμε πως ο γράφος αντιστοιχεί σ'αυτόν τον χώρο στη συνέχεια) και κατόπιν, συσταδοποιούνται με την προσαρμογή ενός υπερ-επιπέδου πάνω τους (θα δούμε και πάλι στη συνέχεια, γιατί η επιλογή ευκλείδειας απόστασης στον χώρο ιδιοδιανυσμάτων είναι η σωστή). Αν αυτή η διαδικασία επαναλαμβάνεται συνεχώς, τότε κατασκευάζουμε παράλληλα και τον γράφο και την ιδανική κατάτμηση των δεδομένων. Η ιδέα του *Neural Normalized Cut* δεν έχει μελετηθεί στη βιβλιογραφία και θα αποτελέσει μια μελλοντική επέκταση της παρούσας διπλωματικής.

### 2.1.3 Κατασκευή γράφων σε εικόνες: SLIC-RAG graph

Ανάλογα με την εφαρμογή μπορεί να υπάρχουν περισσότερο ή λιγότερο αποτελεσματικοί τρόποι να φτιάξει κανείς ένα γράφο από τα δεδομένα του. Εστιάζοντας στο πρόβλημα της γραφοθεωρητικής κατάτμησης εικόνων για εντοπισμό αντικειμένων, εκτός από τις προηγούμενες απαιτήσεις για συνεκτικότητα κλπ, καθίσταται ιδιαίτερα επιτακτική η ανάγκη για δημιουργία ενός αραιού γράφου, λόγω του μεγάλου πλήθους *pixel*. Ωστόσο, ο γράφος θα πρέπει συγχρόνως να κωδικοποιεί με τον καλύτερο δυνατό τρόπο την τοπική πληροφορία, για να ενισχύει την απόδοση του αλγορίθμου κατάτμησης. Συχνά, στην πράξη, για να ελαττωθεί η χρονική πολυπλοκότητα του προβλήματος, πραγματοποιείται σαν προεπεξεργασία μια υποδειγματοληψία των *pixel*, ενώ στην συνέχεια τα εναπομείναντα *pixel* σχηματίζουν τους κόμβους ενός 4/8 κανονικού *grid*. ([ΣημΝῦτ]). Αυτός είναι ένας αφελής τρόπος να διαχειριζόμαστε το πρόβλημα, και στην πράξη έχει ως αποτέλεσμα την απώλεια σημαντικής πληροφορίας. Μάλιστα, ακόμη και αν δεν δειγματοληπούσαμε τα *pixel* της εικόνας, εκτός από την αύξηση της χρονικής πολυπλοκότητας, θα είχαμε και πολύ μεγάλη εισαγωγή θορύβου στον γράφο από τα πολύ χαμηλού επιπέδου χαρακτηριστικά, με αποτέλεσμα την κακή ποιότητα κατάτμησης.

Η κατάτμηση εικόνων είναι γενικά ένα μη καλώς ορισμένο πρόβλημα, με την έννοια

ότι τελικός κριτής της κατάτμησης είναι ο άνθρωπος ή το επόμενο στάδιο αναγνώρισης αντικειμένων, αν υπάρχει. Επομένως, θα πρέπει να έχουμε σαν γνώμονα τις μεθόδους που χρησιμοποιεί ο άνθρωπος για την τελική κατάτμηση. Στην πραγματικότητα, αυτή δεν πραγματοποιείται σε ένα στάδιο, αλλά ιεραρχικά. Αρχικά λαμβάνεται η χαμηλού επιπέδου πληροφορία, όπως το χρώμα και οι ακμές της εικόνας. Στη συνέχεια, συνδυάζεται αυτή η πληροφορία σε μεσαίου επιπέδου χαρακτηριστικά, όπως η υφή ή η κίνηση. Για να υπερβούμε το σημασιολογικό κενό μεταξύ χαμηλού επιπέδου χαρακτηριστικών και αφηρημένων αντικειμένων υψηλού επιπέδου, σημαντικό ρόλο παίζει η ενσωμάτωση πιο αφηρημένων χαρακτηριστικών, όπως οι συμμετρίες. Κατά τον Wertheimer ([45]) -ο οποίος εφάρμοσε τους νόμους *Gestalt* στο πρόβλημα της αντιληπτικής ομαδοποίησης και οργάνωσης στην όραση-, σημαντικοί παράγοντες που οδηγούν σε οπτική ομαδοποίηση είναι η ομοιότητα, η εγγύτητα, η συνέχεια και η κλειστότητα. Με άξονα αυτούς τους παράγοντες, πρέπει λοιπόν να αναθεωρήσουμε την κατασκευή του γράφου, επαναπροσδιορίζοντας τί πρέπει να θεωρούμε ως κόμβους του γράφου και πώς πρέπει να συνδέονται μεταξύ τους. Εξάλλου, σύμφωνα με τους Ren and Malik ([33]) τα *pixel* δεν είναι φυσικές οντότητες, αλλά το αποτέλεσμα της διακριτής αναπαράστασης της εικόνας. Μπορούμε ωστόσο, για την αναπαράσταση σε προβλήματα γραφοθεωρητικής κατάτμησης, να μην αρκεστούμε στη διακριτοποίηση αυτή.

Εφαρμόζοντας τα παραπάνω θα εκμεταλλευτούμε σε πρώτο στάδιο την τοπική πληροφορία, ενώνοντας *pixel* που είναι όμοια και κοντά μεταξύ τους, σε συστάδες από *pixels* (*superpixel*). Αυτά θα αποτελούν τους κόμβους του γράφου. Στην πραγματικότητα, η θέση και το χρώμα του *superpixel* -που είναι η πληροφορία που θα εκμεταλλευτούμε στην συνέχεια για να κατασκευάσουμε τα βάρη του γράφου-, είναι η μέση τιμή της θέσης και του χρώματος αντίστοιχα των *pixel* που το απαρτίζουν. Με αυτόν τον τρόπο και μειώνουμε το πλήθος των κόμβων και δεν χανόμαστε στον θόρυβο της πολύ τοπικής πληροφορίας, που προκύπτει από χρωματικές αντιθέσεις μεταξύ *pixel*. Τελικά, ελπίζουμε ότι το επόμενο στάδιο, που στην περίπτωση μας κάνει κατάτμηση κοιτάζοντας καθολικά την εικόνα, θα ολοκληρώσει αποτελεσματικά την τοπική πληροφορία. Αν επιτρέπαμε στα *superpixel* να έχουν επικάλυψη μεταξύ τους, τότε η διαδικασία αυτή θυμίζει *mean pooling* που χρησιμοποιείται κατά κόρον στην αρχιτεκτονική των νευρωνικών, μόνο που εδώ ο πυρήνας είναι μεταβλητού μεγέθους. Η δημιουργία *superpixel* επιτυγχάνεται με υπερκατάτμηση της εικόνας, συνήθως εκτελώντας αλγόριθμους κατάτμησης σε "oversegmentation mode". ([1],[6],[33],[40])

Παραδοσιακά, η υπερκατάτμηση μιας εικόνας γίνεται με τον αλγόριθμο του *watershed* στην *grayscale* εκδοχή της εικόνας ή του *gradient* της εικόνας. Ωστόσο, όπως φαίνεται και στο παρακάτω σχήμα (2.4) αυτό δημιουργεί πολύ άνισα *superpixel* με πολύ περίεργα μη κυρτά όρια, που πλήττουν την ποιότητα της τελικής κατάτμησης, γιατί φέρνουν τα κέντρα των *superpixel* πολύ κοντά μεταξύ τους εξαιτίας του περίεργου περιγράμματός τους. Ακόμη, δεν δίνουν κανέναν έλεγχο στον χρήστη κατά τη δημιουργία των *superpixel*, πράγμα καθοριστικό όταν π.χ. υπάρχουν παραπάνω από μια αποδεκτές κατατμήσεις της εικόνας, αλλά ανάλογα με την εφαρμογή εμείς θέλουμε να σεβαστούμε μόνο μια από αυτές (*clustering*

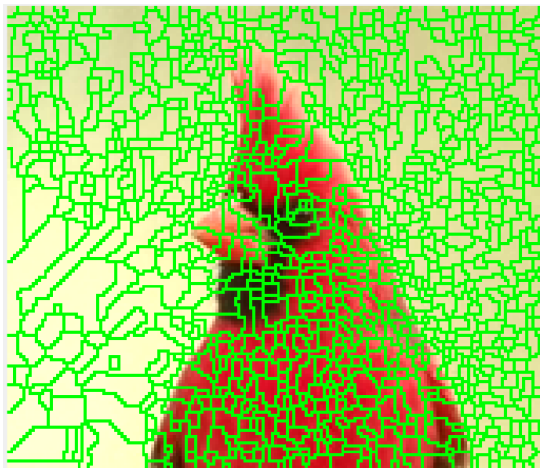
*ambiguity*). Από μια μέθοδο σχηματισμού *superpixel* θα θέλαμε ιδανικά τα εξής:

- Να είναι χρονικά αποδοτική, ειδικά αν ο σχηματισμός *superpixel* αποτελεί βήμα προεπεξεργασίας.
- Να κάνει αποδοτική διαχείριση μνήμης, για να μπορεί να επεξεργαστεί μεγαλύτερα μεγέθη εικόνων.
- Να δίνει στον χρήστη τη δυνατότητα να ελέγχει την υπερκατάτμηση, τον αριθμό ή το πόσο συμπαγή θα είναι τα *superpixel* κλπ.
- Να δημιουργεί *superpixels*, που θα σέβονται τα περιγράμματα των αντικειμένων (boundary-adherent).
- Να μπορεί να επεκταθεί σε *supervoxels*.
- Να δημιουργεί *superpixel* με απλό, ιδανικά κυρτό περίγραμμα. Στο [33] οι συγγραφείς αναφέρουν, ότι η ομοιότητα εντός μιας περιοχής συνεπάγεται καμπύλες χαμηλής ενέργειας εντός αυτής.
- Να ενισχύει εν τέλει την ποιότητα της κατάτμησης, αν πρόκειται για *presegmentation* στάδιο.

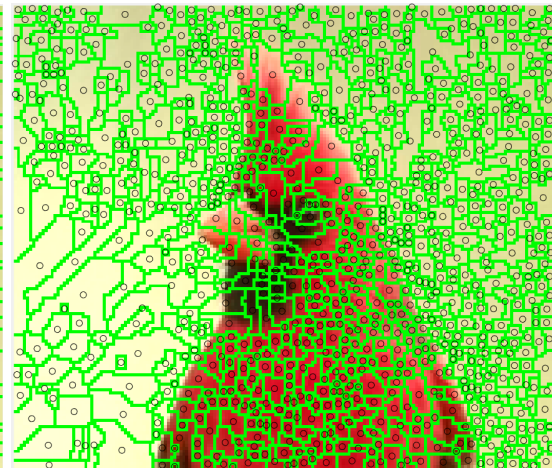




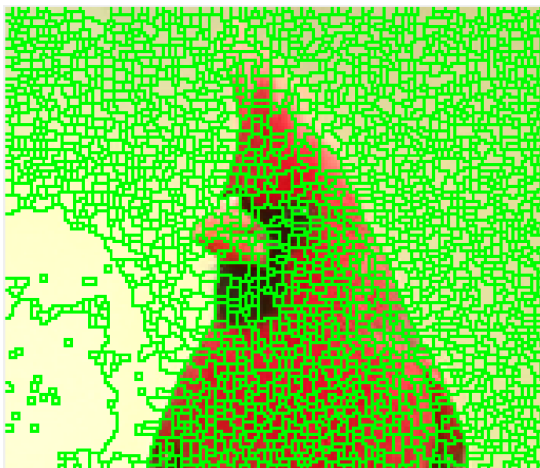
(α') Image (151x176)



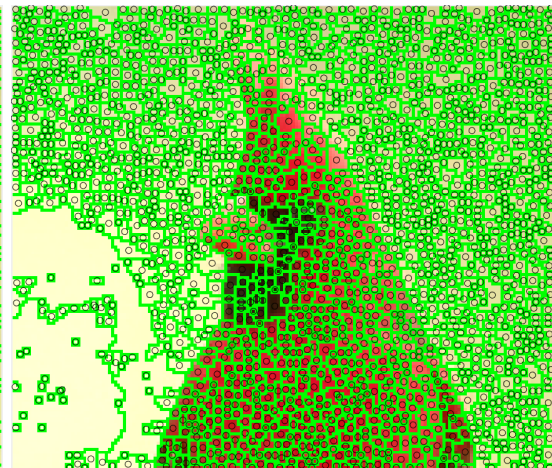
(β') Watershed Superpixels (983)



(γ') Watershed Superpixels Centroids



(δ') Watershed Superpixels on Image Gradient (2083)



(ε') Watershed Superpixels Centroids on Image Gradient

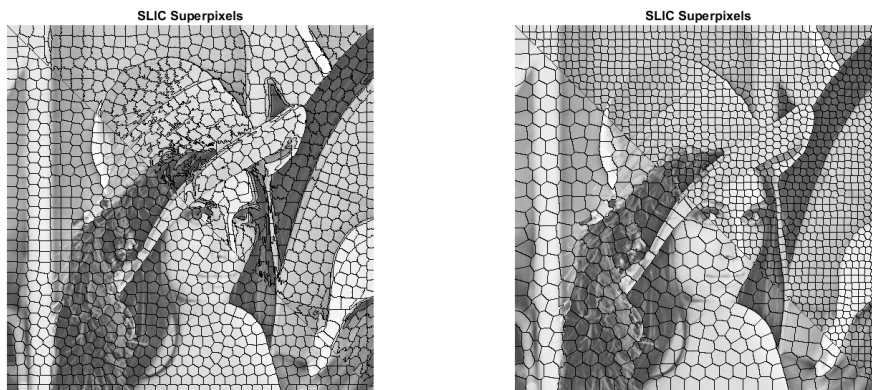
Σχήμα 2.4: Watershed Superpixels and Centroids on Image and Image Gradient



Το πρώτο πράγμα που παρατηρούμε είναι, ότι στο *gradient* της εικόνας σχηματίζονται πολύ περισσότερα *superpixels* από ότι στην κανονική εικόνα, πράγμα που όπως θα δούμε μετά, αποβαίνει καταστρεπτικό για τη χρονική πολυπλοκότητα της κατάτμησης, αλλά βελτιώνει την απόδοση σχετικά με τα *watershed superpixel* στην κανονική εικόνα. Επιπλέον, εξαιτίας της ευαισθησίας της παραγώγου σε μικρές αλλαγές φωτεινότητας, βλέπουμε στο αριστερό μέρος των τελευταίων εικόνων κάποια απομονωμένα *pixel* νησίδες, που αναγκαστικά σχηματίζουν μόνα τους *superpixels*. Επίσης, μπορούμε να παρατηρήσουμε ότι εν γένει η εφαρμογή του αλγορίθμου στην παράγωγο της εικόνας δημιουργεί *superpixel*, που σέβονται περισσότερο τα περιγράμματα των αντικειμένων. Από την άλλη όμως, και στις δύο μεθόδους παρατηρούμε, ότι κάποια κεντροειδή περιοχών που δεν μοιάζουν μεταξύ τους, έρχονται πολύ κοντά χωρικά, εξαιτίας του μη κυρτού περιγράμματος κάποιων *superpixel*, που μπορούν να οδηγήσουν τα κέντρα τους ακόμη και εκτός της περιοχής τους. Το αντίκτυπο αυτού θα φανεί στη συνέχεια μέσα από την ποιότητα της κατάτμησης.

Στο πρόβλημά μας θα χρησιμοποιήσουμε τα *SLIC superpixels* ([1]) που είναι το αποτέλεσμα της εκτέλεσης του *k - means* σε ένα 5-διάστατο χωροχρωματικό χώρο, όπου οι 2 πρώτες συντεταγμένες είναι η  $[x, y]$  θέση κάθε *pixel*, ενώ οι 3 τελευταίες χρωματικές διαστάσεις είναι το  $[l, a, b]$  διάνυσμα χρώματος κάθε *pixel*, στο χώρο χρώματος *CIELAB* που θεωρείται ως αντιληπτικά ομοιόμορφος για μικρές χρωματικές αποστάσεις (σαν ευκλείδειος δηλαδή). Κατ' αυτόν τον τρόπο εκμεταλλευόμαστε τοπική πληροφορία που έχει σχέση με την εγγύτητα και με την ομοιότητα. Ο αλγόριθμος παίρνει σαν είσοδο το μέγεθος των *superpixel* (το οποίο είναι προσεγγιστικά τηρούμενο για κάθε ένα) και μια παράμετρο που ρυθμίζει το πόσο boundary adherent θα είναι η έξοδος. Αν αυτή η παράμετρος γίνει πολύ μεγάλη τα περιγράμματα γίνονται πολύ ευαίσθητα στις μικρές αλλαγές τιμών των *pixel* και, όπως και αυτά του *watershed*, γίνονται πολύ μη κυρτά χαλώντας την ποιότητα της κατάτμησης. Από την άλλη αν γίνει πολύ μικρή τότε εμφανίζονται φαινόμενα "bleeding" στην εικόνα, δηλαδή τα *superpixel* τεμνονται από ένα περίγραμμα αντικειμένου όπως φαίνεται στο παρακάτω σχήμα 2.5. Η μέθοδος αυτή ανήκει στις προτεινόμενες (state-of-the-art) μεθόδους κατασκευής *superpixel* σύμφωνα με το [40] που συγκρίνει όλες τις γνωστές μεθόδους ως προς το οπτικό αποτέλεσμα, την χρονική διάρκεια, το boundary adherence και πολλές άλλες μετρικές.

Στο πρώτο σχήμα μπορούμε να δούμε πως όσο απαιτούμε από τον αλγόριθμο να σέβεται τα περιγράμματα των αντικειμένων, τόσο πιο πολύπλοκα γίνονται τα περιγράμματα των *superpixel*, ενώ από την άλλη, μικρή προσαρμογή υποδηλώνει φαινόμενα "leaking". Αυτά, επίσης, θα επηρεάσουν αρνητικά την τελική κατάτμηση, η οποία δεν θα μπορεί να βελτιωθεί περαιτέρω αν τα δομικά αντικείμενα δεν συνθέτουν πραγματικά (σημασιολογικά) αντικείμενα στην εικόνα. Τέτοια φαινόμενα παρατηρούμε και στο δεύτερο σχήμα, στο οποίο αλλάζουμε μόνο το μέγεθος των *superpixel*, για την περίπτωση όλο και μεγαλύτερων *superpixel*. Προφανώς, όπως θα δούμε και παρακάτω, υπάρχει σε αυτό το σημείο ένα trade-off μεταξύ χρονικής πολυπλοκότητας (μεγάλα, αλλά λίγα *superpixel*) και απόδοσης (μικρά *superpixel*, που δεν παρουσιάζουν "leaking") τουλάχιστον μέχρι κάποιο βαθμό.



(α') Boundary Adherence

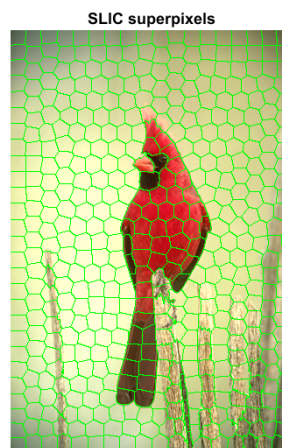
(β') Superpixel Sizes

Σχήμα 2.5: SLIC parameters

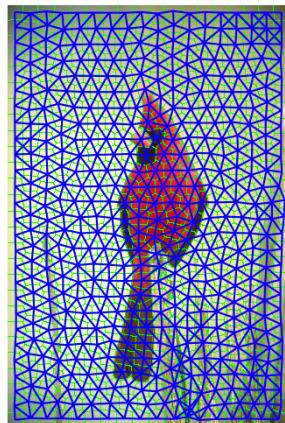
Η σύνδεση των *superpixel* μεταξύ τους για την κατασκευή του γράφου είναι το επόμενο στάδιο της ιεραρχικής δομής με την οποία αντιμετωπίζουμε την ολοκλήρωση της τοπικής πληροφορίας. Θα χρησιμοποιήσουμε ένα Region Adjacency Graph (*RAG*) για την σύνδεση των *superpixel*. Ο λόγος είναι ότι ο γράφος που δημιουργείται είναι αρκετά αραιός, πράγμα που θα επιταχύνει τον αλγόριθμο κατάτμησης στην συνέχεια (με *Lanczos* μεθόδους στον υπολογισμό των ιδιοτιμών) ενώ είναι και διαισθητικά σωστή επιλογή να μην συνδέουμε *superpixel* που βρίσκονται πολύ μακριά, αφού όχι μόνο η χρωματική αντίθεση αλλά και η χωρική απόσταση είναι μέτρα ανομοιότητας των κόμβων. Επίσης, δίνει πάντα συνεκτικό γράφο σε αντίθεση με τον *K-nn graph*, με πιά γενική μορφή, αφού κόμβοι με πιά περίεργα περιγράμματα έχουν εν γένει περισσότερους γείτονες. Βέβαια το τελευταίο παρατηρήσαμε προηγουμένως πως συμβαίνει και στον μη-κατευθυνόμενο *K-nn* (symmetric or mutual). Μπορούμε να σκεφτούμε τον *RAG-SLIC* γράφο σαν μια άλλη εκδοχή του Voronoi Tassellation το οποίο εκμεταλλεύεται ο *K-nn graph*. Σε αυτή την εκδοχή το tassellation γίνεται από την δημιουργία των superpixel, με βάση την απόσταση κάποιων προτοποητημένων (όχι όμως σταθερών) κέντρων, διαδικασία που καταλήγει λόγω *k-means* στην δημιουργία περιοχών που χωρίζονται από την μεσοκάθετο, των κέντρων όμως, όχι των δεδομένων. Η σύνδεση τέλος, δεν γίνεται από τα κοντινότερα κέντρα, αλλά από τα εφαπτόμενα περιγράμματα. Σίγουρα, θα ήταν ένα πολύ χρήσιμο αντικείμενο έρευνας η σύγκριση διάφορων γράφων σε αλγορίθμους κατάτμησης, γιατί στην πράξη υπάρχει μεγάλη επίδραση του σταδίου αυτού στο τελικό αποτέλεσμα της κατάτμησης. Οπτικά το αποτέλεσμα της κατάτμησης φαίνεται στα παρακάτω σχήματα. Παρατηρούμε την συμπαγή δομή των SLIC superpixel, και την κανονική δομή των γράφων *RAG* που σχηματίζονται που οδηγούν σε συνεπή και γρήγορη κατάτμηση των εικόνων.



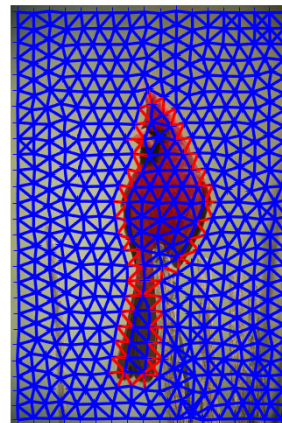
(α') Original Image



(β') SLIC superpixels



(γ') RAG-SLIC Graph

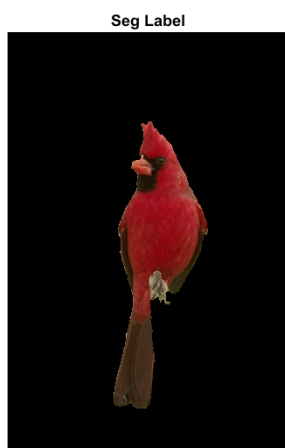


(δ') Graph Output

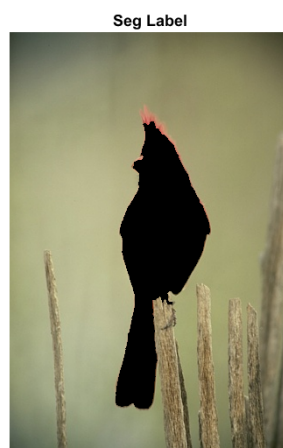
Σχήμα 2.6: bird image. RAG-SLIC graph and Normalized Cut segmentation



(α') Coloured Clusters



(β') Cluster 1

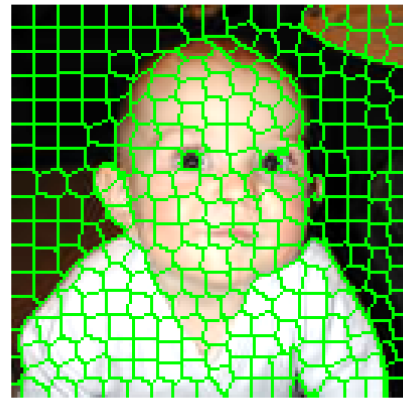


(γ') Cluster 2

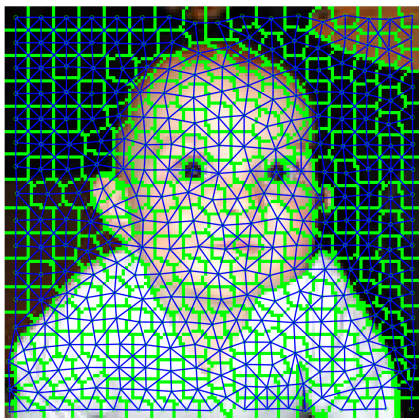
Σχήμα 2.7: bird image. Clusters extended to pixels



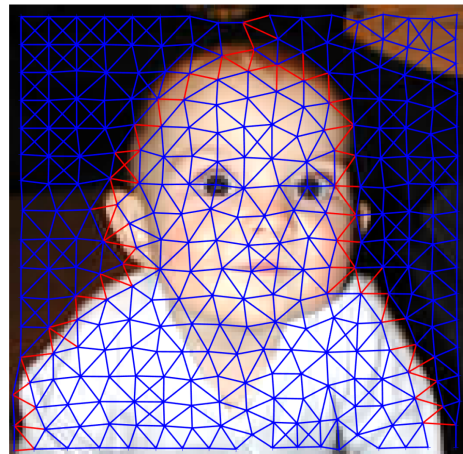
(α') Original Image



(β') SLIC Superpixels



(γ') RAG-SLIC graph



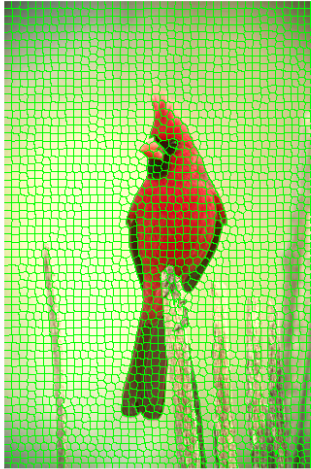
(δ') Normalized Cut Output



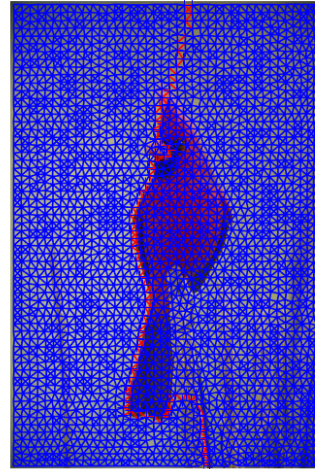
(ε') Output Extended to pixels

Σχήμα 2.8: baby segmentation

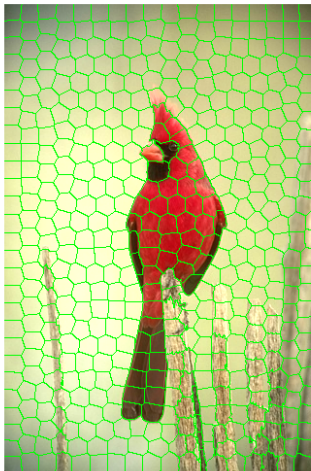




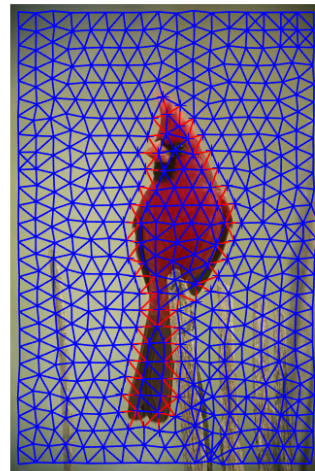
(α') SLIC Size=8



(β') NCut



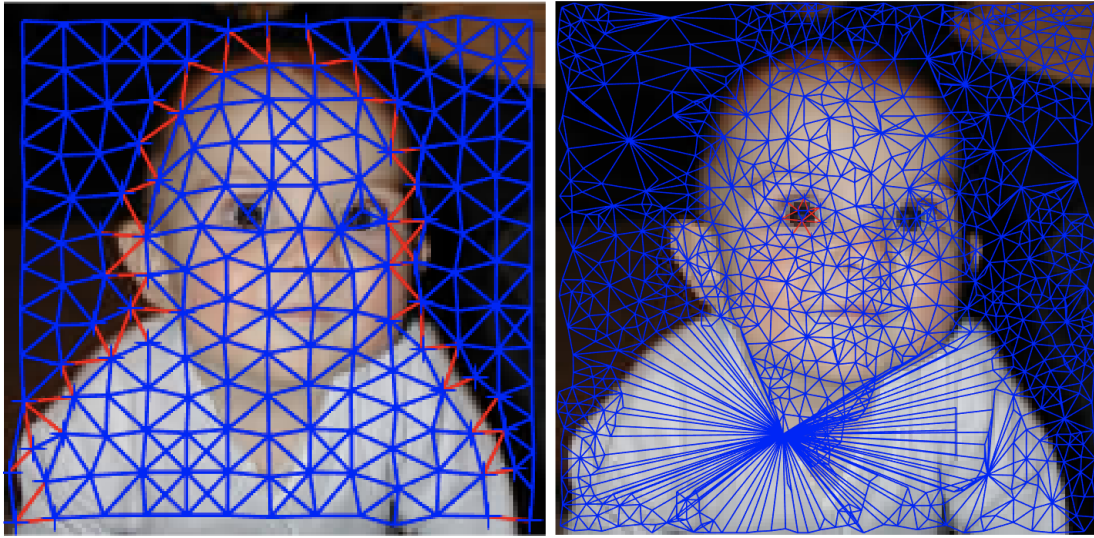
(γ') SLIC Size=18



(δ') NCut

Σχήμα 2.9: Size Parameter. Smaller superpixels is not always a better choice.

Περσσότερες εικόνες στο κεφάλαιο 3, μετά την περιγραφή του αλγορίθμου κατάτμησης και των χαρακτηριστικών που λαμβάνονται στην κατασκευή του γράφου. Μια εποπτική σύγκριση της κατάτμησης με *Watershed superpixels* και με *SLIC superpixels* ως προς την ποιότητα και το χρόνο φαίνεται στο παρακάτω σχήμα 2.10:



(α') SLIC Superpixels

(β') Watershed Superpixels

Σχήμα 2.10: Σύγκριση ποιότητας και χρόνου μεταξύ των δύο μεθόδων κατασκευής superpixel. Η πρώτη μέθοδος εκτός από το ότι παρουσιάζει εμφανώς καλύτερα αποτελέσματα, είναι και πολύ πιο γρήγορη. Ο σχηματισμός του γράφου πήρε 0.3 δευτερόλεπτα και για τις δύο, ενώ στο πρώτο η τομή πήρε 0.005 δευτερόλεπτα και στο δεύτερο 0.05 δευτερόλεπτα.

Στο παραπάνω σχήμα 2.10 χρησιμοποιήθηκε για την τελική κατάτμηση ο αλγόριθμος Normalized Cut που θα αναλυθεί στην συνέχεια. Είναι εμφανές ότι τα SLIC superpixels εκτός από το γεγονός ότι επιταχύνουν την τελική κατάτμηση κατά 10 φορές, επειδή είναι πιο αραιά τοποθετημένα στο χώρο και δημιουργούν μια σχετικά κανονική δομή στον RAG γράφο, η ομαλή τοποθέτησή τους στο χώρο επιτρέπει σε πολύ καλύτερα αποτελέσματα κατάτμησης από ότι τα Watershed superpixels, που όπως φαίνεται στην εικόνα κατευθύνουν την κατάτμηση να διαχωρίσει μικρές κλάσεις λόγω απότομων αραιωμάτων στον γράφο της εικόνας.

## 2.2 Εισαγωγή στη Φασματική Θεωρία Γράφων

### 2.2.1 Τελεστές επί των κόμβων

Από την στιγμή που συσχετίσαμε τον γράφο με ένα πίνακα (γεινιάσης), γίνεται φανερό ότι μπορούμε να τον σκεφτόμαστε σαν ένα γραμμικό τελεστή επί των κορυφών. Ορίζοντας ένα διάνυσμα  $f = (f_1, \dots, f_n)$  με τιμές πάνω στους κόμβους του γράφου και θεωρώντας μη κατευθυνόμενους γράφους χωρίς ακμές ανακύκλωσης, η επίδραση του τελεστή πάνω στο διάνυσμα είναι:

$$(Af)_i = \sum_{j \sim i} f_j$$

Δηλαδή η τιμή του  $Af$  στην κορυφή  $u_i$  είναι το άθροισμα των τιμών των γειτόνων του (και του εαυτού του αν υπάρχει ανακύκλωση). Επειδή κατα κάποιο τρόπο ο τελεστής

”διαχέει” την πληροφορία της τιμής της κορυφής  $u_i$  στους γείτονές του, ονομάζεται τελεστής διάχυσης. Γνωρίζουμε από την γραμμική άλγεβρα ότι τα ιδιοδιανύσματα ενός πίνακα είναι εκείνα τα διανύσματα στα οποία ο τελεστής επιδρά πολλαπλασιαστικά. Η πολλαπλασιαστική σταθερά είναι η ιδιοτιμή του πίνακα. Τα ιδιοδιανύσματα και οι ιδιοτιμές του πίνακα γειτνίασης έχουν την εξής γεωμετρική ερμηνεία:

$$\lambda f_i = \sum_{j \sim i} f_j, \quad \forall i = 1, \dots, m$$

δηλαδή για κάθε κόμβο το άθροισμα των τιμών των γειτόνων είναι σταθερό πολλαπλάσιο της υπάρχουσας τιμής του κόμβου. Παραδείγματος χάρη αν πάρουμε το σταθερό διάνυσμα  $f = (1, 1, \dots, 1)$  η επίδραση του τελεστή είναι η εύρεση του βαθμού των κορυφών, ενώ το διάνυσμα αυτό είναι ιδιοδιάνυσμα αν και μόνο αν ο γράφος είναι  $\kappa$ -κανονικός, δηλαδή όλες οι κορυφές του έχουν τον ίδιο βαθμό  $\kappa$ . Το  $\kappa$  είναι ιδιοτιμή του πίνακα για το ιδιοδιάνυσμα αυτό.

Στην γενικότερη περίπτωση γράφου με βάρη ο αντίστοιχος πίνακας βαρών επιδρά στο διάνυσμα των κορυφών ως:  $(Wf)_i = \sum_{j \sim i} w_{ij} f_j + w_{ii} f_i$ . Ο τελεστής αυτός λέγεται συνήθως τελεστής μετάδοσης θερμότητας (heat diffusion operator). Φάσμα (*Spectrum*) ενός πίνακα ονομάζεται το σύνολο των ιδιοτιμών και ιδιοδιανυσμάτων του. Θα ορίζουμε ως φάσμα ενός γράφου, το φάσμα του πίνακα γειτνίασης (ή βαρών) που τον περιγράφει. Η φασματική θεωρία γράφων είναι ο τομέας που μελετά ιδιότητες των γράφων συσχετίζοντάς τις με τις ιδιοτιμές και τα ιδιοδιανύσματά τους. Θα δούμε πως το φάσμα του γράφου δίνει πληροφορία για πολλές καθολικές ιδιότητες του όπως την συνεκτικότητα, την διάμετρο, τη διμερότητα, τον ισοπεριμετρικό αριθμό, τις συμμετρίες κλπ. Ακριβώς επειδή είναι καθολική ποσότητα μπορεί να αποτελέσει και ένα μέτρο σύγκρισης μεταξύ γράφων πράγμα πολύ χρήσιμο σε εφαρμογές ισομορφισμού γράφων ([32]) και άλλων *combinatorial* προβλημάτων που συχνά στην πράξη είναι δύσκολο να λυθούν (*NP – complete*). Κάποια παραδείγματα γράφων και των φασμάτων τους φαίνονται στην συνέχεια.

Για την μελέτη των ιδιοτιμών του πίνακα βαρών (υποθέτουμε θετικά βάρη), μπορούμε να χρησιμοποιήσουμε την θεωρία *Perron – Frobenius*. Για μια εκτεταμένη αναφορά ο αναγνώστης παραπέμπεται στο [27]. Κάποια πορίσματα της θεωρίας αυτής είναι τα εξής:

1. Η μεγαλύτερη ιδιοτιμή του πίνακα γειτνίασης είναι τουλάχιστον ίση με τον μέσο βαθμό του γράφου και το πολύ ίση με τον μέγιστο βαθμό του γράφου. (Αυτό κατ’ εξαίρεση προκύπτει από τον *variational* ορισμό των ιδιοτιμών που θα δούμε αργότερα)
2. Αν ο γράφος είναι συνεκτικός η μεγαλύτερη ιδιοτιμή του έχει αλγεβρική πολλαπλότητα 1 και το αντίστοιχο ιδιοδιάνυσμα έχει όλα του τα στοιχεία θετικά.
3. Η μεγαλύτερη και η μικρότερη ιδιοτιμή του γράφου είναι αντίθετες αν και μόνο αν ο γράφος είναι διμερής.
4. Με την αύξηση του βάρους μιας ακμής, το *spectral radius*  $\rho_A = \max_i(|\lambda_i|)$  αυξάνεται.



Έχοντας περιορίσει την συζήτηση μας για μη-κατευθυνόμενους γράφους και άρα για συμμετρικούς πίνακες βαρών, είμαστε σίγουροι ότι υπάρχουν  $n$  ιδιοτιμές πραγματικές. Μάλιστα, ισχύει το εξής θεώρημα από γραμμική άλγεβρα.

### Θεώρημα 2.1.

1. Αν ο πίνακας  $A$  είναι συμμετρικός (γεν. Ερμιτιανός) τότε όλες οι ιδιοτιμές του είναι πραγματικές.
2. Αν ο πίνακας  $A$  είναι συμμετρικός (γεν. Ερμιτιανός) τότε δύο ιδιοδιανύσματα του που αντιστοιχούν σε διαφορετικές ιδιοτιμές είναι κάθετα.
3. Αν ο πίνακας  $A$  είναι συμμετρικός και  $\lambda_1, \dots, \lambda_n$  είναι το σύνολο των ιδιοτιμών του (όχι απαραίτητα διαφορετικών) τότε υπάρχουν διανύσματα ανα δύο ορθογώνια  $x_1, \dots, x_n$ ,  $x_i \in \mathbb{R}^n$  τέτοια ώστε τα  $x_i$  να είναι ιδιοδιανύσματα των ιδιοτιμών  $\lambda_i$ .

Μια καλογραμμένη, πλήρης απόδειξη μπορεί να βρεθεί στο [42].

Θα σταθούμε λίγο ακόμη στους τελεστές επί των κορυφών για να καταδείξουμε άλλον ένα πολύ σημαντικό τελεστή διάχυσης, που ονομάζεται *walk operator*. Ορίζεται ως :  $P = WD^{-1}$  και επιδρά σε κάθε κόμβο :

$$(Pf)_i = \sum_{j \sim i} \frac{f_j}{d_j}$$

Λέγεται *walk matrix* γιατί κωδικοποιεί τις δυναμικές ενός τυχαίου περιπατητή που μεταβαίνει σε κάθε επόμενη κορυφή ανάλογα με το βάρος της ακμής σύνδεσης. Η εφαρμογή του σε ένα διάνυσμα  $f$  που περιγράφει τις πιθανότητες μετάβασης από ένα κόμβο σε έναν άλλο, δίνει την πιθανότητα μετάβασης δύο βημάτων, δηλαδή  $(Pf)_i = \mathbb{P}(\mathcal{X}_2 = u_i)$ , όπου  $\mathcal{X}_t$  είναι η τυχαία μεταβλητή που δείχνει την κατάσταση του τυχαίου περιπατητή σε χρόνο  $t$ . Επειδή όλες οι ιδιοτιμές του πίνακα ανήκουν στο διάστημα  $[0,1]$  (στοχαστικός πίνακας) και πάντα το διάνυσμα  $f = s \cdot d = s \cdot (d_1, \dots, d_n)$  είναι ιδιοδιάνυσμα του γράφου  $\forall s \in \mathbb{R}$  με ιδιοτιμή  $\lambda_n = 1$ , η επαναληπτική εφαρμογή αυτού του τελεστή θα συγκλίνει τελικά στην αναλλοίωτη κατανομή  $\pi = \frac{d}{\sum_n d_i}$ , του γράφου. Θα δούμε στην συνέχεια τη σχέση που έχουν οι τυχαίοι περίπατοι με τα προβλήματα της φασματικής γραφοθεωρητικής κατάτμησης.

### 2.2.2 Graph Laplacians

Ένας ακόμη πίνακας που σχετίζεται με τον γράφο και που το φάσμα του κρύβει πολύ μεγάλη πληροφορία για την δομή του είναι ο Graph Laplacian πίνακας που ορίζεται ως:

$$\mathbf{L} = \mathbf{D} - \mathbf{W}$$

Για κίνητρο θα αναφέρουμε μόνο κάποιες εφαρμογές του που σχετίζονται με τυχαίους περιπατητές ([15]), διάδοση θερμότητας, *graph embedding*, συνεκτικότητα, αραιές τομές, ενώ

χρησιμοποιείται συχνά στο φυσικό ανάλογο του γράφου με κυκλώματα ([12]) ή ελατήρια. Ακόμη, το φάσμα του δίνει πλούσια πληροφορία για πολλές καθολικές ιδιότητες του γράφου. Παραδείγματος χάρη δίνει πληροφορία για το πλήθος των συνεκτικών δέντρων του γράφου:

**Θεώρημα 2.2.** (*Kirchhoff's matrix tree theorem*) Δοσμένου ενός συνεκτικού αβαρούς γράφου, ο αριθμός των συνεκτικών δέντρων του είναι:  $t(G) = \frac{1}{n} \lambda_2 \lambda_3 \cdots \lambda_n$ , όπου  $\lambda_2, \dots, \lambda_n$  είναι οι μη μηδενικές ιδιοτιμές του Graph Laplacian.

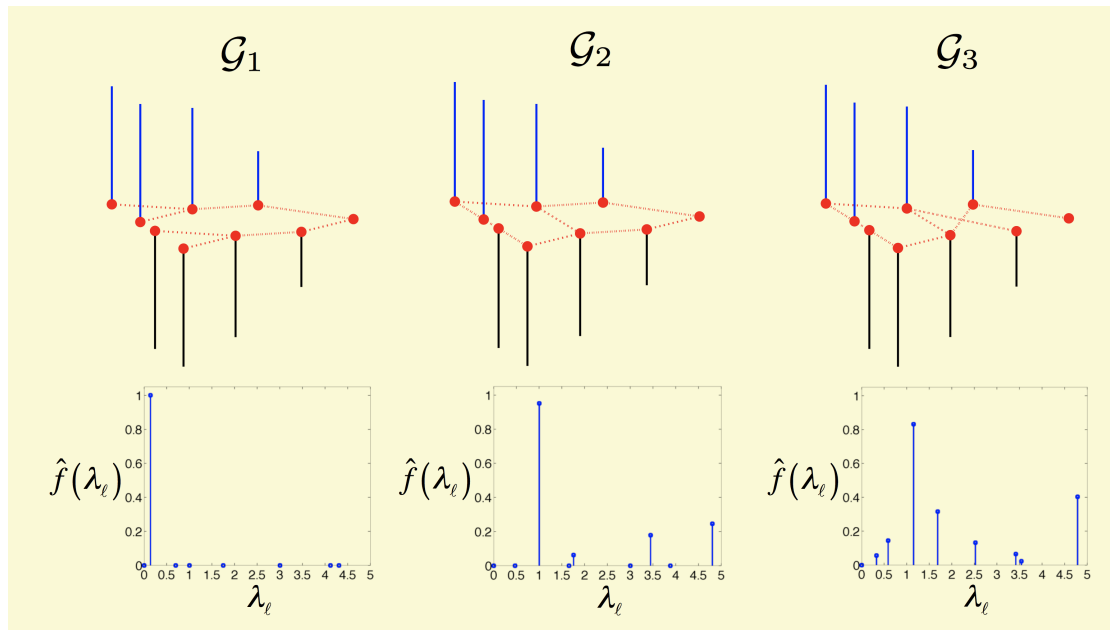
Και αυτός ο πίνακας είναι συμμετρικός και επομένως έχει πραγματικές ιδιοτιμές και ένα πλήρες σετ από κάθετα ιδιοδιανύσματα. Στην ειδική περίπτωση που ο γράφος είναι  $k$ -κανονικός ισχύει  $L = kI - A$  και επομένως  $\lambda_i = d - \mu_i$ , όπου  $\lambda_i, \mu_i$  οι ιδιοτιμές των  $L, A$  αντίστοιχα. Επίσης, οι δύο πίνακες έχουν κοινά ιδιοδιανύσματα. Στην γενικότερη περίπτωση όμως το φάσμα τους διαφέρει.

Στο εξής θα θεωρήσουμε μη-κατευθυνόμενους γράφους με θετικά βάρη στις ακμές. Οι ακμές ανακύκλωσης δεν αλλάζουν τον graph laplacian αφού προστίθενται εξίσου στον πίνακα βαρών και τον πίνακα βαθμών. Μια ακόμη ιδιότητα του graph laplacian που δεν έχει ο πίνακας γειτνίασης/βαρών είναι ότι ο  $L$  είναι θετικά ημι-ορισμένος. Ο τρόπος να δει κανείς ότι ο πίνακας βαρών δεν έχει αυτή την ιδιότητα είναι να σκεφτεί ότι σε γράφους χωρίς ανακύκλωση το  $\text{trace}(W) = 0$ . Όμως επειδή το  $\text{trace}$  είναι ίσο με το άθροισμα των ιδιοτιμών, αν ο γράφος δεν είναι κενός, θα έχει και θετικές και αρνητικές ιδιοτιμές. Συνεπώς, δεν είναι θετικά (ημι-)ορισμένος. Αντίθετα, ο graph laplacian έχει διαγώνια υπεροχή και θετικά στοιχεία στην διαγώνιο άρα από το κριτήριο *Sylvester* είναι θετικά ημι-ορισμένος. Επειδή μάλιστα οι γραμμές του αθροίζουν πάντα σε 0, το σταθερό διάνυσμα  $\mathbf{1}_n$  είναι πάντα ιδιοδιάνυσμα του πίνακα, με ιδιοτιμή 0. Άρα, ποτέ ο *graph laplacian* δεν μπορεί να είναι θετικά ορισμένος. Αυτό μπορεί να το δείξει κανείς και σχηματίζοντας την τετραγωνική μορφή του πίνακα  $L$ .

$$\begin{aligned} x^T L x &= x^T D x - x^T W x = \sum_{i=1}^m d_i x_i^2 - \sum_{i,j=1}^m w_{ij} x_i x_j \\ &= \frac{1}{2} \left( \sum_{i=1}^m d_i x_i^2 - 2 \sum_{i,j=1}^m w_{ij} x_i x_j + \sum_{j=1}^m d_j x_j^2 \right) \quad (2.5) \\ &\iff x^T L x = \frac{1}{2} \sum_{i,j=1}^m w_{ij} (x_i - x_j)^2 \geq 0 \end{aligned}$$

Η τετραγωνική μορφή του πίνακα μας δίνει και μια παραπάνω χρήσιμη πληροφορία για την συνάρτηση  $x$  επί των κορυφών. Μας λέει πόσο ομαλή είναι η συνάρτηση αυτή πάνω στον γράφο. Πράγματι, εκτός του ότι η σταθερή συνάρτηση δίνει την μηδενική μορφή στην τετραγωνική μορφή, άλλες συναρτήσεις που δεν αλλάζουν απότομα μεταξύ των ακμών δίνουν μικρή τιμή στην τετραγωνική μορφή. Γιαυτό και παίζει κεντρικό ρόλο στην επιστήμη του *discrete signal processing on graphs* ([38],[35]), που ασχολείται με την γενίκευση κλασικών προβλημάτων και αλγορίθμων σημάτων, όπως ο μετασχηματισμός Fourier, η γραμμική

πρόβλεψη, η συμπίεση σημάτων, η διάχυση κλπ. πάνω σε γράφους. Στο παρακάτω σχήμα 2.11 παρατηρεί κανείς ότι η ομαλότητα ενός σήματος πάνω σε μια διακριτή δομή όπως ο γράφος, εξαρτάται εκτός από τις τιμές του σήματος και από τις συνδέσεις των κόμβων του γράφου. Σε μια συνεχή δομή δεν θα είχαμε κάτι αντίστοιχο. Απεικονίζουμε λοιπόν το ίδιο σήμα  $f(u)$  επί των κορυφών πάνω σε 3 διαφορετικούς γράφους, με ίδιες κορυφές αλλά διαφορετικές ακμές. Με κόκκινο βλέπουμε τις θετικές τιμές και με μαύρο τις αρνητικές. Κατόπιν, απεικονίζουμε την συνάρτηση στο χώρο του φάσματος κάθε γράφου. Όπως στο συνεχή χώρο είχαμε τα μιγαδικά εκθετικά σαν βάση του πυρήνα του μετασχηματισμού *Fourier*, τώρα έχουμε σαν βάση τα ιδιοδιανύσματα του *graph Laplacian*. Δεδομένου ότι το πεδίο ορισμού της αρχικής συνάρτησης είναι πάνω στο χώρο κορυφών του γράφου, η μετασχηματισμένη θα έχει ισάριθμο πεδίο ορισμού αλλά στο χώρο των ιδιοτιμών του *graph Laplacian*. Τελικά, ο μετασχηματισμός είναι απλά το εσωτερικό γινόμενο της αρχικής με κάθε ιδιοδιάνυσμα του *graph Laplacian*. Παρατηρούμε πόσο πιά πλούσιο φασματικό περιεχόμενο έχουν οι συναρτήσεις που ορίζονται πάνω στους δύο τελευταίους γράφους, οι οποίοι έχουν πολλές συνδέσεις μεταξύ κόμβων με αρνητική και θετική τιμή σήματος. Επιβεβαιώνουμε το ότι η εποπτική ομαλότητα των σημάτων ποσοτικοποιείται με την τετραγωνική μορφή του *graph Laplacian*, υπολογίζοντας ότι:  $f^\top L_1 f = 0.13$ ,  $f^\top L_2 f = 1.31$ ,  $f^\top L_3 f = 1.81$ .



Σχήμα 2.11: Απεικονίζουμε το ίδιο σήμα σε 3 διαφορετικούς γράφους, για να καταδείξουμε ότι η ομαλότητα του εξαρτάται και από την υποδόσκουσα δομή. Σχήμα από [38]

Το γεγονός ότι ο *graph laplacian* είναι θετικά ημι-ορισμένος σημαίνει ότι μπορεί να παραγοντοποιηθεί ως το γινόμενο ενός πίνακα και του αναστρόφου του:  $L = BB^T$ . Στην περίπτωση μας ο πίνακας αυτός έχει μεγάλη σημασία και απλή ερμηνεία. Ας ορίσουμε τον πίνακα πρόσπτωσης ενός κατευθυνόμενου γράφου, με θετικά βάρη στις ακμές ως τον πίνακα

με στοιχεία:

$$b_{ij} = \begin{cases} +\sqrt{w_{ij}}, & i \neq j \quad s(e_j) = u_i \\ -\sqrt{w_{ij}}, & i \neq j \quad t(e_j) = u_i \\ 0 & \text{otherwise.} \end{cases}$$

Αν ορίσουμε και μια συνάρτηση κατευθυντικότητας των ακμών  $\sigma : E \rightarrow V \times V$  που αναθέτει σε κάθε ακμή μια πηγή και ένα προορισμό, τότε ο graph laplacian παραγοντοποιείται στο γινόμενο:  $L = B_\sigma B_\sigma^T$ , όπου ο  $B_\sigma$  είναι ο πίνακας πρόσπτωσης οποιουδήποτε κατευθυνόμενου γράφου  $\mathcal{G}_\sigma$  που προκύπτει από τον αντίστοιχο μη-κατευθυνόμενο  $\mathcal{G}$  μετά την εφαρμογή της συνάρτησης  $\sigma$ . Όπως έχουμε δει όμως ο graph laplacian γράφεται και ως  $D - W$  άρα δεν εξαρτάται από τον τύπο της συνάρτησης κατευθυντικότητας. Ο λόγος που υπάρχει είναι για συνέπεια των ορισμών με τους κατευθυνόμενους γράφους. Θα δούμε στη συνέχεια ένα θεώρημα που αφορά τον πίνακα πρόσπτωσης και βρίσκεται στην καρδιά του spectral graph clustering.

Ας δούμε πρώτα το φάσμα του graph laplacian στη γενικότερη μοφή που ο γράφος δεν είναι κανονικός. Η επίδραση του τελεστή graph laplacian πάνω σε ένα διάνυσμα  $f$  ορισμένο στους κόμβους του γράφου είναι:

$$(Lf)_i = \sum_{j \sim i} w_{ij}(f_i - f_j)$$

Η επίδραση αυτή είναι αντίστοιχη της *Laplacian* σε συνεχείς συναρτήσεις ([9],[13]). Αυτό είναι ένα κεντρικό σημείο στον κλάδο του discrete calculus, που ασχολείται με ανάλυση συναρτήσεων πάνω σε διακριτές δομές, που δεν έχουν απαραίτητα μια υποβόσκουσα συνεχή δομή. Για να το κάνουμε αυτό πιό ξεκάθαρο, αν ο  $B$  είναι ο πίνακας πρόσπτωσης του  $\mathcal{G}_\sigma$ , τότε ο  $B^T$  ορίζει ένα είδος διακριτής παραγώγου κατά μήκος διανυσμάτων εφαπτόμενων πάνω σε ένα manifold (discrete covariant derivative)  $\nabla : \mathbb{R}^V \times \mathcal{X}(\mathcal{G}) \rightarrow \mathbb{R}^V$  πάνω στις 0-φόρμες, που είναι το σύνολο των συναρτήσεων του  $\mathbb{R}^V$ . Για κάθε κόμβο  $u_i \in V$  συγκεντρώνουμε το σύνολο των ακμών που έχουν ένα άκρο στο  $u_i$

$$T_{u_i}G = \{(u_i, u_j) | w_{ij} \neq 0\} \cup \{(u_j, u_i) | w_{ji} \neq 0\}$$

σε ένα σεντ που μοιάζει με το διακριτό εφαπτόμενο χώρο του  $u_i$ . Η ένωση των χώρων αυτών για κάθε κορυφή είναι το λεγόμενο tangent bundle TG. Επομένως, ένα διακριτό διανυσματικό πεδίο είναι μια συνάρτηση  $X : V \rightarrow TG$  που αναθέτει σε κάθε κόμβο  $u_i \in V$  κάποια ακμή  $X(u_i) = e_k \in T_{u_i}G$ , και ονομάζουμε όλον αυτό το σύνολο των διακριτών διανυσματικών πεδίων ως  $\mathcal{X}(G)$ . Κατόπιν, για κάθε συνάρτηση  $f \in \mathbb{R}^V$  και για κάθε διανυσματικό πεδίο  $X \in \mathcal{X}(G)$ , ορίζουμε την συνάρτηση  $\nabla_X f$ , ένα διακριτό ανάλογο της covariant παραγώγου της συνάρτησης  $f$  ως προς το διανυσματικό πεδίο  $X$ ,

$$(\nabla_X f)(u_i) = B^T(f)(X(u_i))$$

εάν λοιπόν  $e_k = (u_i, u_j)$ , τότε

$$(\nabla_X f)(u_i) = \sqrt{w_{ij}}(f_i - f_j)$$

ενώ αν  $e_k = (u_j, u_i)$ , τότε

$$(\nabla_X f)(u_i) = \sqrt{w_{ij}}(f_j - f_i)$$

τότε ο graph laplacian  $L$  δίνεται από τον τύπο  $L = BB^T$  και για κάθε κόμβο έχουμε:

$$(Lx)_i = \sum_{j \sim i} w_{ij}(x_i - x_j)$$

Επομένως, ο graph laplacian εμφανίζεται να είναι ένα διακριτό ανάλογο του *Bohner Laplacian*.

Ως προς τις ιδιοτιμές του graph laplacian θα παραθέσουμε το παρακάτω θεώρημα ([13]):

**Θεώρημα 2.3.** Δοθέντος ενός μη κατευθυνόμενου γράφου  $n$  κόμβων και  $c$  συνεκτικών συνιστωσών και ενός πίνακα πρόσπτωσης  $B$ , τότε:

1.  $c = \dim(\text{Ker}(B^T))$  και  $\text{rank}(B) = n - c$
2. Ο πυρήνας του  $B^T$  έχει μια βάση που αποτελείται από δείκτριες συναρτήσεις των συνεκτικών συνιστωσών του γράφου.
3. Ο πληθίριθμος των μηδενικών ιδιοτιμών του graph laplacian είναι ίσος με τον αριθμό των συνεκτικών συνιστωσών του γράφου.

Αντίστροφα, το θεώρημα αυτό μας λέει ότι όταν ο γράφος είναι συνεκτικός, ισχύει  $\lambda_2 > 0$ . Μάλιστα, η πρώτη μη-μηδενική ιδιοτιμή του graph laplacian ονομάζεται αλγεβρική συνεκτικότητα του γράφου ή αριθμός *Fiedler*. Ο λόγος είναι ότι όσο πιο μακριά από το 0 βρίσκεται, τόσο πιο πυκνά συνδεδεμένος είναι ο γράφος. Αυτό το είδαμε στο προηγούμενο θεώρημα για τον πίνακα γειτνίασης/βαρών και την αύξηση της του spectral radius με την προσθήκη/ενδυνάμωση των ακμών. Ανάποδα, κοιτάζοντας συνεκτικούς γράφους με μεγάλες δεύτερες ιδιοτιμές (δες σχήμα) παρατηρούμε ότι η αύξηση της συνεκτικότητας δεν σχετίζεται απαραίτητα με την προσθήκη μεγάλου αριθμού ακμών, αλλά με την "έξυπνη" σύνδεσή τους, ώστε η αφαίρεση οποιονδήποτε από αυτές να μην χωρίζει το γράφο σε πολλές συνιστώσες. Μάλιστα, υπάρχουν γράφοι με πληθικότητα ακμών  $|E| = \mathcal{O}(|V|)$ , που έχουν πάρα πολύ ισχυρή συνεκτικότητα. Αυτοί οι γράφοι ονομάζονται *expanders* και χρησιμοποιούνται πολύ σε εφαρμογές τηλεπικοινωνιών ή δικτύων που ο στόχος είναι η κατασκευή ενός *robust* συστήματος, με μεγάλη ανεκτικότητα στην καταστροφή κάποιων συνδέσεων. Θα θεμελιώσουμε καλύτερα τα παραπάνω στο επόμενο κεφάλαιο μέσω του ισοπεριμετρικού προβλήματος και της σταθεράς *Cheeger*.

Το επόμενο που θα χρειαστεί είναι να ορίσουμε μια σταθμισμένη εκδοχή του graph laplacian που θα χρησιμοποιήσουμε στα επόμενα κεφάλαια. Στην βιβλιογραφία υπάρχουν δύο εκδοχές normalized graph laplacian με τους εξής τύπους:

1.  $L_{sym} = D^{-1/2}LD^{-1/2}$

$$2. L_{rw} = D^{-1}L$$

Ο  $L_{sym}$  είναι συμμετρικός πίνακας, ενώ ο  $L_{rw}$  σχετίζεται πολύ με τυχαίους περιπατητές στο γράφο και με rapidly mixing Markov chains. Τα επόμενα θεωρήματα είναι πολύ χρήσιμα για την συνέχεια οπότε τα παραθέτουμε χωρίς απόδειξη. Ίσως πρώτα όμως πρέπει να φτιάξουμε το πλαίσιο στο οποίο δουλεύουμε και αποδεικνύουμε αυτά τα θεωρήματα. Το επόμενο θεώρημα που εξετάζει τις ιδιοτιμές από την σκοπιά της θεωρίας βελτιστοποίησης θα μας βοηθήσει.

**Θεώρημα 2.4. (Courant-Fischer)** Αν ο  $A$  είναι συμμετρικός πίνακας  $n \times n$ , με ιδιοτιμές  $\lambda_1 \leq \dots \leq \lambda_n$  και ιδιοδιανύσματα  $\{u_1, \dots, u_n\}$ , τότε:

$$\lambda_k = \min_{\substack{S \subseteq \mathbb{R}^n \\ \dim(S)=k}} \max_{\substack{x \in S \\ x \neq 0}} \frac{x^T A x}{x^T x} = \max_{\substack{S \subseteq \mathbb{R}^n \\ \dim(S)=n-k+1}} \min_{\substack{x \in S \\ x \neq 0}} \frac{x^T A x}{x^T x}$$

Και η λύση που πετυχαίνει το ελάχιστο είναι το ιδιοδιάνυσμα της ιδιοτιμής αυτής ( $u_k$ ). Για παράδειγμα αυτό το θεώρημα μας λέει πως :

$$\lambda_1 = \min_{x \in \mathbb{R}^n} \frac{x^T A x}{x^T x}, \quad \lambda_n = \max_{x \in \mathbb{R}^n} \frac{x^T A x}{x^T x}, \quad \lambda_{k+1} = \min_{x \in \{u_1, \dots, u_k\}^\perp} \frac{x^T A x}{x^T x}$$

Ο λόγος  $R_A(x) = \frac{x^T A x}{x^T x}$  ονομάζεται στην βιβλιογραφία πηλίκο Rayleigh και χρησιμοποιείται σε πολλά προβλήματα βελτιστοποίησης ειδικά σε προβλήματα ιδιοτιμών, π.χ. στη PCA ή στο Normalized Cuts που θα αναλύσουμε στη συνέχεια.

**Θεώρημα 2.5. (Poincare separation theorem)** Αν ο  $A$  είναι συμμετρικός πίνακας  $n \times n$ , με ιδιοτιμές  $\lambda_1 \leq \dots \leq \lambda_n$  και ο  $P$  είναι  $n \times m$  πίνακας με την ιδιότητα  $P^T P = I$  ( $m \leq n$ ), τότε ο πίνακας  $B = P^T A P$  έχει τις εξής ιδιότητες:

1. Αν  $\mu_1 \leq \dots \leq \mu_m$  είναι οι ιδιοτιμές του  $B$  τότε  $\lambda_i \leq \mu_i \leq \lambda_{n-m+i}, \forall i \in \{1, \dots, m\}$
2. Αν  $\lambda_i = \mu_i$ , τότε υπάρχει ένα ιδιοδιάνυσμα  $u$  του  $B$  με ιδιοτιμή  $\mu_i$  τέτοιο ώστε το  $Ru$  να είναι ιδιοδιάνυσμα του  $A$  με ιδιοτιμή  $\lambda_i$ .

Το τελευταίο έχει σαν αποτέλεσμα ότι:

$$\lambda_1 + \dots + \lambda_m \leq \text{tr}(P^T A P) \leq \lambda_{n-m+1} + \dots + \lambda_n$$

Μια πλήρης απόδειξη των θεωρημάτων αυτών μπορεί να βρεθεί στο [13].

**Θεώρημα 2.6.** Αν  $G = (V, W)$  είναι ένας γράφος χωρίς απομονωμένες κορυφές, τότε οι γράφοι Laplacian που σχετίζονται μάλιστα έχουν τις εξής ιδιότητες ([9]):

1. Ο συμμετρικός Laplacian είναι συμμετρικός και θετικά ημιορισμένος με τετραγωνική μορφή:

$$x^T L_{sym} x = \frac{1}{2} \sum_{i,j=1}^m w_{ij} \left( \frac{x_i}{\sqrt{d_i}} - \frac{x_j}{\sqrt{d_j}} \right)^2, \forall x \in \mathbb{R}^m$$

2. Οι πίνακες  $L, L_{rw}$  είναι όμοιοι και επομένως έχουν τις ίδιες ιδιοτιμές και κάθε ιδιοδιάνυσμα του πρώτου αν πολλαπλασιαστεί από αριστερά με το  $D^{1/2}$  αποτελεί ιδιοδιάνυσμα του δεύτερου για την ίδια ιδιοτιμή. Αυτό μάλιστα σημαίνει ότι παρότι ο  $L_{rw}$  δεν είναι συμμετρικός έχει πραγματικές ιδιοτιμές και πλήρη βάση ιδιοδιανυσμάτων. Αν και όχι πάντα ορθοκανονική.
3. Ένα διάνυσμα  $u \neq 0$  είναι λύση του γενικευμένου προβλήματος ιδιοτιμών  $Lu = \lambda Du$  αν και μόνο αν το  $D^{1/2}u$  είναι ιδιοδιάνυσμα του  $L_{sym}$  για την ιδιοτιμή  $\lambda$  και αν και μόνο αν το ίδιο διάνυσμα είναι και ιδιοδιάνυσμα του  $L_{rw}$  για την ίδια ιδιοτιμή.
4. Οι πίνακες  $L, L_{rw}$  έχουν τον ίδιο πυρήνα. Κάθε στοιχείο του πυρήνα τους είναι στοιχείο του πυρήνα του  $L_{sym}$  αν πρώτα πολλαπλασιαστεί από αριστερά με  $D^{1/2}$ . Άρα, το  $\mathbf{1}$  ανήκει στο πυρήνα του  $L_{rw}$  και το  $D^{1/2}\mathbf{1}$  στον πυρήνα του  $L_{sym}$ .
5. Κάθε ιδιοτιμή του συμμετρικού Λαπλάσιαν ανήκει στο διάστημα  $[0, 2]$ . Επιπλέον, αν η μεγαλύτερη ιδιοτιμή είναι ίση με 2, τότε ο γράφος έχει ένα μη-τετριμμένο, συνεκτικό διμερές κομμάτι.
6. Αν έχουμε τουλάχιστον 2 κορυφές και ο γράφος δεν είναι πλήρης, τότε η δεύτερη ιδιοτιμή του συμμετρικού Λαπλάσιαν είναι μικρότερη ή ίση με 1. Από την άλλη, ο γράφος θα είναι πλήρης αν και μόνο αν  $\lambda_2 = n/(n-1)$ .
7. Για κάθε γράφο χωρίς απομονωμένες κορυφές και με περισσότερες από 2 κορυφές ισχύει για τις ιδιοτιμές της συμμετρικής Λαπλασιαν ότι:  $\lambda_n \geq n/(n-1)$ .
8. Το φάσμα ενός γράφου είναι η ένωση των φασμάτων των συνεκτικών συνιστωσών του.





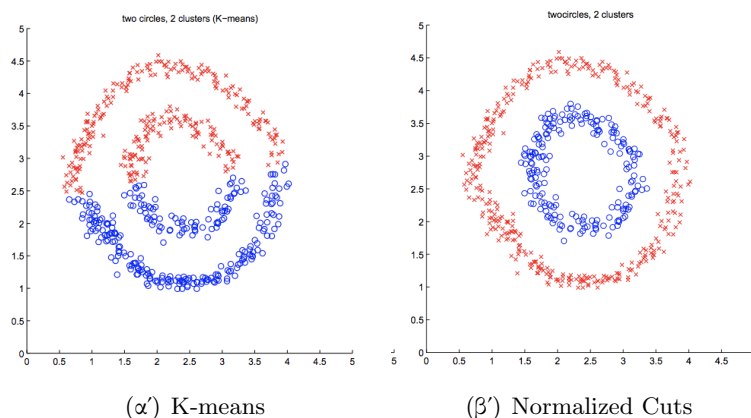
## Κεφάλαιο 3

# Φασματική κατάτμηση γράφων με την ελαχιστοποίηση σταθμισμένων τομών

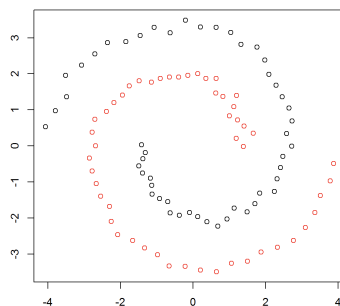
### 3.1 Εισαγωγή

Στο προηγούμενο κεφάλαιο, αναφέραμε πληθώρα χρήσεων της φασματικής θεωρίας γράφων για εύρεση διαφόρων (κυρίως καθολικών) ιδιοτήτων ενός γράφου, καθώς και διάφορα προβλήματα που στην βιβλιογραφία επιλύονται με εύρεση του φάσματος ενός γράφου, όπως ο ισομορφισμός δύο γράφων κλπ. Σε αυτό το κεφάλαιο θα εστιάσουμε σε μεθόδους γραφοθεωρητικής κατάτμησης με εφαρμογή σε σημασιολογική κατάτμηση εικόνων, σε εντοπισμό κοινοτήτων σε κοινωνικά δίκτυα ή σε point cloud δεδομένα. Ο στόχος της συσταδοποίησης είναι κυρίως το exploratory data analysis, δηλαδή η εποπτική επεξεργασία των δεδομένων, η απεικόνιση τους σε χώρους που κάποια χαρακτηριστικά είναι πιο εμφανή, η ανακατασκευή των κλάσεων της υπόθεσης (π.χ. από stochastic block models ή low-dimensional manifolds) καθώς και η συμπίεση δεδομένων με ελάττωση της διάστασης των χαρακτηριστικών. Βιβλιογραφικά, υπάρχουν πολλές μέθοδοι για κατάτμηση δεδομένων ([20]), με πιο γνωστή τη μέθοδο του k-means. Συνοπτικά, αναφέρουμε ότι σε αυτή τη μέθοδο θέτουμε μια αρχική κατάτμηση των δεδομένων τοποθετώντας τυχαία ή με κάποιο πιο έξυπνο τρόπο τα κεντροειδή των κλάσεων. Στη συνέχεια, αναθέτουμε τα σημεία στα κοντινότερα (με κάποια έννοια απόστασης) κεντροειδή. Κατόπιν, ανανεώνουμε τη θέση των κεντροειδών, ώστε να βρίσκονται στο κέντρο βάρους των σημείων της κλάσης τους και επαναλαμβάνουμε αρκετές φορές αυτή τη διαδικασία ανάθεσης σημείων σε κλάσεις και ανανέωσης των κεντροειδών. Ο στόχος αυτής της επαναληπτικής διαδικασίας είναι να μειώσει την απόσταση των κεντροειδών από τα (άγνωστα) σημεία κάθε κλάσης. Αν και υπάρχουν κάποιες θεωρητικές εγγυήσεις για την επίδοση του k-means, επειδή η ελαχιστοποίηση της μετρικής που χρησιμοποιεί είναι NP-hard, συχνά μπορεί να κολλήσει σε τοπικά ελάχιστα της αντικειμενικής συνάρτησης, κάτι που αποβαίνει μοιραίο για την επίδοση του αλγορίθμου.

Ο λόγος που θα θέλαμε να χρησιμοποιήσουμε γραφοθεωρητικές μεθόδους, είναι το ότι ο  $k$ -means δρα σε διανύσματα χαρακτηριστικών και όχι σε γράφους, οπότε αν μας δίνεται εξαρχής ένας γράφος προς κατάτμηση π.χ. ένα κοινωνικό δίκτυο ή ένα δίκτυο πρωτεϊνών, δεν υπάρχει άμεσος τρόπος να εφαρμόσουμε τον  $k$ -means διατηρώντας ταυτόχρονα τη σημασιολογία του γράφου. Επιπλέον, όπως φαίνεται και στο παρακάτω σχήμα ο κλασικός  $k$ -means δεν έχει την δυνατότητα να αναγνωρίσει μη-κυρτά clusters. Αντίθετα, οι μέθοδοι φασματικής γραφοθεωρητικής κατάτμησης δεν κάνουν υποθέσεις για την δομή των άγνωστων clusters και γι' αυτό μπορούν να κατατμήσουν σχήματα, όπως οι διαπλεκόμενες σπείρες ή οι ομόκεντροι κύκλοι που φαίνονται στο παρακάτω σχήμα (3.1,3.2). Τελικά βέβαια, μπορούμε να υπερβούμε το περιορισμό της ευκλείδειας απόστασης που καταλήγει σε διαχωρισμό μόνο κωρτών *cluster* χρησιμοποιώντας kernel  $k$ -means. Υπάρχουν πυρήνες για τους οποίους ο αλγόριθμος του  $k$ -means ισοδυναμεί με τις φασματικές μεθόδους κατάτμησης. Ο αναγνώστης παραπέμπεται στα ([34],[2]). Για τους ίδιους λόγους αποτυγχάνουν και άλλες μέθοδοι που χρησιμοποιούν επαναληπτικές διαδικασίες όπως ο  $EM$  αν υποθέτουν κάποια μορφή για τις κατανομές.



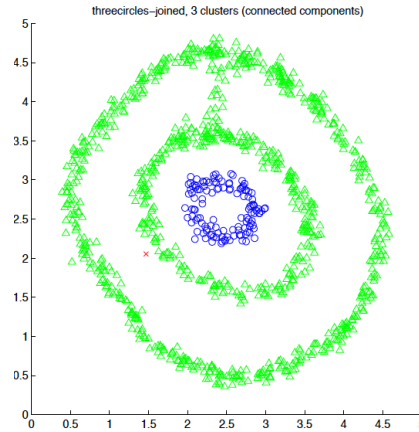
Σχήμα 3.1: Κατάτμηση ομόκεντρων κύκλων



Σχήμα 3.2: Κατάτμηση διαπλεκόμενων σπειρών με Normalized Cuts

Επιπλέον, αφού το πρόβλημα είναι η κατάτμηση ενός γράφου σε  $K$  κομμάτια, φροντίζοντας να συσταδοποιήσουμε διαφορετικά, σημεία που είναι μακριά και άλλα έχουν μικρή

ομοιότητα, θα μπορούσε κανείς να υποθέσει ότι αν ο γράφος είναι σωστά κατασκευασμένος, με μια κατωφλιοποίηση στις ακμές του γράφου στο σημείο που θα μας έδιναν  $K$  συνιστώσες θα πέραμε αρκετά καλές κλάσεις. Η μέθοδος αυτή αποτυγχάνει για το λόγο ότι βασίζεται πολύ στην καλή κατασκευή του γράφου, πρόβλημα που όμως είναι επίσης μη επιβλεπόμενο. Στο παρακάτω σχήμα φαίνεται ένα παράδειγμα με 3 κλάσεις που ο αλγόριθμος διαχωρίζει μια μοναδική κορυφή σε μια κλάση.



Σχήμα 3.3: Αλγόριθμος κατωφλιοποίησης γράφου

Παρότι το ζητούμενο είναι η γραφοθεωρητική κατάτμηση, δεν είναι ακόμη ξεκάθαρο πως υπεισέρχεται η φασματική θεωρία γράφων στην επίλυση του προβλήματος. Ο λόγος είναι η χρήση της μετρικής μας, της οποίας η βελτιστοποίηση θα καταλήξει σε ένα πρόβλημα ιδιοδιανυσμάτων. Σε αυτό το σημείο κάποιος θα μπορούσε να αναρωτηθεί γιατί δεν βρίσκουμε στον γράφο την ελάχιστη τομή (έστω ότι ψάχνουμε 2 κλάσεις), πρόβλημα που έχει επιλυθεί με την μοντελοποίηση του σαν γραμμικό πρόγραμμα και χρησιμοποιώντας την δυϊκότητα ελάχιστης τομής-μέγιστης ροής. Το τελευταίο πρόβλημα λύνεται με τον αλγόριθμο Ford-Fulkerson ή με άλλες γενικότερες παραλλαγές του. Το πρόβλημα με την ελάχιστη τομή είναι ότι συνήθως απομονώνει πολύ πολύ μικρά κομμάτια του γράφου, ακόμη και μια μεμονωμένη κορυφή, και συνεπώς το αποτέλεσμα δεν έχει κανένα σημασιολογικό νόημα (πιό πολύ κόβει outliers: ένα pixel σε μια εικόνα ή ένα κόμβο σε ένα κοινωνικό δίκτυο). Η μετρική που θα χρησιμοποιήσουμε ονομάζεται Normalized Cut [37] και εκτός από την ελαχιστοποίηση της τομής ταυτόχρονα μεγιστοποιεί τον όγκο κάθε κλάσης, μετρούμενο ως προς το συνολικό βάρος των ακμών της. Φορμαλιστικά για  $K$  κλάσεις  $(A_1, \dots, A_K)$ ,

$$Ncut(A_1, \dots, A_K) = \sum_{i=1}^K \frac{links(A_i, \bar{A}_i)}{vol(A_i)} = \sum_{i=1}^K \frac{cut(A_i, \bar{A}_i)}{vol(A_i)}, \quad (3.1)$$

όπου

$$links(A, B) = \sum_{u_i \in A, u_j \in B} w_{ij},$$

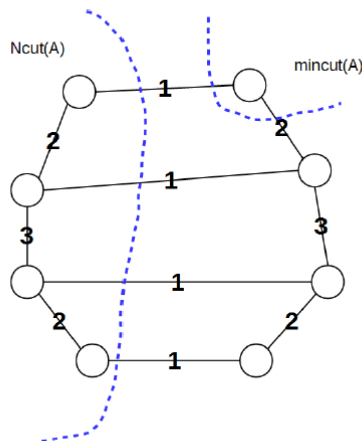
$$cut(A) = links(A, \bar{A}),$$

$$\text{assoc}(A) = \text{links}(A, A),$$

$$\text{vol}(A) = \sum_{u_i \in A} d(u_i) = \text{cut}(A) + \text{assoc}(A)$$

Στο παρακάτω σχήμα βλέπουμε πως επιδρά σε ένα απλό γράφο.

$$\min \text{Ncut}(A).$$



$$\min \text{Ncut}(A) = \frac{4}{3+6+6+3} + \frac{4}{3+6+6+3} = \frac{4}{9}.$$

Σχήμα 3.4: Min Cut versus Min Normalized Cut

## 3.2 Φασματική κατάτμηση γράφου σε 2 κλάσεις

Στην ενότητα αυτή θα μοντελοποιήσουμε το πρόβλημα της σταθμισμένης τομής σαν ένα πρόβλημα συνδυαστικής βελτιστοποίησης, και θα το επιλύσουμε χαλαρώνοντάς το σε ένα επιλύσιμο πρόβλημα ιδιοτιμών. Θα επικεντρωθούμε στην κατάτμηση σε 2 κλάσεις γιατί παρουσιάζει ιδιαίτερο ενδιαφέρον και μεγάλη βιβλιογραφία σε πολλούς επιστημονικούς χώρους όπως οι θεωρητικοί αλγόριθμοι, η μηχανική μάθηση, η όραση υπολογιστών, η θεωρία βελτιστοποίησης και κατόπιν θα επεκτείνουμε την ανάλυση σε πολλές κλάσεις.

### 3.2.1 Μοντελοποίηση προβλήματος

Έστω ότι μας δίνεται ένας γράφος  $\mathcal{G}(\mathcal{V}, \mathcal{E})$  με κορυφές  $\{u_1, \dots, u_n\}$  στις οποίες ορίζουμε μια πραγματική συνάρτηση  $x : \mathcal{V} \rightarrow \mathbb{R}$ . Στο κεφάλαιο αυτό θα θεωρούμε κυρίως συνδεδεμένους γράφους. Αν είναι μη-συνδεδεμένοι, τότε σίγουρα δεν θα υπάρχουν απομονωμένοι κόμβοι, έτσι ώστε ο βαθμός κάθε ακμής να είναι θετικός. Έστω επίσης ότι σε κάθε ακμή  $\{u_i, u_j\}$  ανατίθεται ένα βάρος  $w_{ij} \geq 0$ . Στόχος είναι ο διαχωρισμός των κορυφών σε δύο κλάσεις  $(A, \bar{A})$ , επομένως έχει νόημα η συνάρτηση επί των κορυφών να έχει την μορφή:

$$x_i = \begin{cases} a, & u_i \in A \\ b, & u_i \notin A \end{cases}$$

όπου προφανώς  $a \neq b$ . Αν ορίσουμε  $X = (x_1, \dots, x_n)$  το διάνυσμα των τιμών της συνάρτησης πάνω στις κορυφές τότε η σωστή ερμηνεία του διανύσματος αυτού είναι ότι αποτελεί ένα σημείο  $\mathcal{P}(X)$  του πραγματικού προβολικού χώρου  $\mathbb{RP}^{n-1}$  με ομογενείς συντεταγμένες  $(x_1 : \dots : x_n)$ . Ο λόγος που χρειαζόμαστε την σωστή ερμηνεία του διανύσματος είναι για να ορίσουμε ποιά απόσταση μεταξύ διανυσμάτων είναι η κατάλληλη σε αυτό το χώρο, όπως θα δούμε στην συνέχεια. Από την τετραγωνική μορφή της *Laplacian* που είδαμε στο προηγούμενο κεφάλαιο έχουμε:

$$X^T \mathbf{L} X = (a - b)^2 \text{cut}(A, \bar{A})$$

ενώ εύκολα αποδεικνύεται ότι:

$$X^T \mathbf{D} X = \text{vol}(A)a^2 + \text{vol}(\bar{A})b^2$$

Όμως,

$$N\text{cut}(A, \bar{A}) = \frac{\text{cut}(A, \bar{A})}{\text{vol}(A)} + \frac{\text{cut}(\bar{A}, A)}{\text{vol}(\bar{A})} = \frac{\text{vol}(G)}{\text{vol}(A)\text{vol}(\bar{A})} \text{cut}(A, \bar{A}),$$

αν επιλέξουμε κατάλληλα τα  $a, b$  μπορούμε να εκφράσουμε την σταθμισμένη τομή σαν ένα γενικευμένο *Rayleigh* κριτήριο.

$$N\text{cut}(A, \bar{A}) = \frac{X^T \mathbf{L} X}{X^T \mathbf{D} X} \iff \frac{(a - b)^2}{\text{vol}(A)a^2 + \text{vol}(\bar{A})b^2} = \frac{\text{vol}(G)}{\text{vol}(A)\text{vol}(\bar{A})}$$

Το τελευταίο είναι ισοδύναμο μετά από λίγη άλγεβρα με:  $\text{vol}(A)a + \text{vol}(\bar{A})b = 0$  ή ισοδύναμο  $X^T \mathbf{D} \mathbf{1}_N = 0$ . Μάλιστα, δεν χρειάζεται καν να επιβάλλουμε τον περιορισμό  $a \neq b$ . Αρκεί να απαιτήσουμε κανένα cluster να μην είναι κενό και το διάνυσμα να μην είναι το μηδενικό. Τότε,  $\text{vol}(A), \text{vol}(\bar{A}) \neq 0$  και από την τελευταία εξίσωση που συνδέει τα  $a, b$  γραμμικά προκύπτει  $a, b \neq 0$ . Τότε όμως αναγκαστικά,  $a \neq b$  αλλιώς  $\text{vol}(G) = 0$ , πράγμα άτοπο. Το ότι ο περιορισμός εξάγεται φυσικά από τον τύπο του προβλήματος είναι πολύ σημαντικό για να μπορούμε ταυτόχρονα να είμαστε σίγουροι ότι έχουμε κάνει σωστή μοντελοποίηση του προβλήματος και ότι μπορούμε να λύσουμε αποδοτικά την χαλαρωμένη εκδοχή που θα δούμε στην συνέχεια. Στην συνέχεια, λόγω ισοδυναμίας των παραπάνω, θα απαιτούμε απλά τα στοιχεία του διανύσματος να είναι μη μηδενικά.

Έστω λοιπόν το σετ των διανυσμάτων μας:

$$\mathcal{X} = \{(x_1, \dots, x_n) | x_i \in \{a, b\}, a, b \in \mathbb{R}, a, b \neq 0\}$$

καθώς και το σετ των λύσεών μας:

$$\mathcal{K} = \{X \in \mathcal{X} | X^T \mathbf{D} \mathbf{1}_N = 0\}.$$

Ψάχνουμε λύσεις στο  $\mathbb{RP}^{n-1}$  τέτοιες ώστε να επιλύουν το εξής πρόβλημα:

$$\begin{aligned} \min_X \quad & \frac{X^T \mathbf{L} X}{X^T \mathbf{D} X} \\ \text{s.t.} \quad & X^T \mathbf{D} \mathbf{1}_N = 0, \quad X \in \mathcal{X} \end{aligned} \tag{3.2}$$

Επειδή και η αντικειμενική συνάρτηση και οι περιορισμοί είναι *scale – invariant* μπορούμε να σχηματίσουμε το ισοδύναμο πρόβλημα:

$$\begin{aligned} \min_X \quad & X^T \mathbf{L} X \\ \text{s.t.} \quad & X^T \mathbf{D} X = 1, \quad X^T \mathbf{D} \mathbf{1}_N = 0, \quad X \in \mathcal{X} \end{aligned} \quad (3.3)$$

Η ισοδυναμία αυτή έγγειται όχι στο ίδιο σύνολο λύσεων αλλά στο γεγονός ότι αν βρούμε μια λύση στο δεύτερο πρόβλημα τότε κάθε πολλαπλάσιό της είναι λύση του αρχικού. Από την άλλη, για να είμαστε σίγουροι ότι δεν θα χάσουμε καμία λύση πρέπει να δείξουμε και το ανάποδο, δηλ ότι κάθε λύση του πρώτου προβλήματος, έστω  $X^*$  αντιστοιχεί σε μια λύση του δεύτερου προβλήματος. Εδώ αυτή η λύση είναι η  $X^*/(X^{*\top} \mathbf{D} X^*)^{1/2}$ . Μπορούμε να θεωρήσουμε ότι έχουμε το ίδιο σετ λύσεων μεταξύ των δύο προβλημάτων, όπως αυτά εκφράζονται από την ισοδυναμία των σημείων στο προβολικό χώρο. Δυστυχώς, το τελευταίο πρόβλημα έχει αποδειχθεί από τον C.Papadimitriou, ότι ακόμη και για  $K=2$  και ακόμη και για επίπεδους γράφους είναι *NP – complete*. Όπως συνηθίζεται στην συνδυαστική βελτιστοποίηση θα λύσουμε το πρόβλημα σε ένα μεγαλύτερο συνεχή χώρο επακριβώς και κατόπιν θα στρογγυλοποιήσουμε την λύση μας σε μια προσεγγιστική λύση. Για αυτή τη λύση υπάρχουν και θεωρητικές εγγυήσεις για το πόσο καλή προσέγγιση αποτελεί. Η διαισθητικά σωστή χαλάρωση είναι να αφήσουμε το διάνυσμα να παίρνει τιμές σε όλο το  $\mathbb{R}^n$ . Ωστόσο υπάρχουν και άλλοι τρόποι να κάνει κανείς χαλάρωση στο πρόβλημα όπως να αλλάξει την μετρική της απόστασης. Θα αναφέρουμε κάτι τέτοιο στην συνέχεια. Για να εκμεταλλευτούμε το θεώρημα του προηγούμενου κεφαλαίου, πρώτα θέτουμε  $Y = \mathbf{D}^{1/2} X \iff X = \mathbf{D}^{-1/2} Y$ , οπότε το πρόβλημα στο συνεχή χώρο γίνεται:

$$\begin{aligned} \min_Y \quad & Y^\top \mathbf{L}_{sym} Y \\ \text{s.t.} \quad & Y^\top Y = 1, \quad Y^\top \mathbf{D}^{1/2} \mathbf{1}_N = 0 \end{aligned} \quad (3.4)$$

όπου  $\mathbf{L}_{sym} = \mathbf{D}^{-1/2} \mathbf{L} \mathbf{D}^{-1/2}$ . Τα ολικά ελάχιστα της συνάρτησής μας βρίσκονται στην θέση οποιουδήποτε ιδιοδιανύσματος του συμμετρικού Λαπλάσιαν που αντιστοιχεί στην δεύτερη ιδιοτιμή. Από τον δεύτερο περιορισμό αυτό το ιδιοδιάνυσμα πρέπει να έχει μέτρο μονάδα. Για να ικανοποιήσουμε τον τελευταίο περιορισμό παρατηρούμε τα εξής: Κοιτάζοντας τις ιδιοτιμές του πίνακα Λαπλάσιαν παρατηρήσαμε ότι πάντα έχει ιδιοτιμή το 0 και ιδιοδιάνυσμα το σταθερό διάνυσμα (αν ο γράφος είναι συνδεδεμένος). Αν όχι μπορούμε να δουλέψουμε ξεχωριστά σε κάθε συνεκτική συνιστώσα γιατί ο σχηματισμένος πίνακας είναι block-διαγώνιος και συνεπώς η ένωση των επιμέρους φασμάτων των υποπίνακων, συγκροτεί το φάσμα όλου του πίνακα). Συνεπώς, με αλλαγή μεταβλητής καταλήγουμε ότι το  $\mathbf{D}^{1/2} \mathbf{1}_N$  ανήκει στο πυρήνα του  $\mathbf{L}_{sym}$ . Επειδή ο πίνακας είναι συμμετρικός όμως υπάρχει τρόπος, όπως είδαμε στο προηγούμενο κεφάλαιο να βρούμε ένα πλήρες σετ από κάθετα ιδιοδιανύσματα. Συνεπώς και το δεύτερο ιδιοδιάνυσμα θα είναι κάθετο στο  $\mathbf{D}^{1/2} \mathbf{1}_N$ .

Επόμενως, η λύση στο αρχικό χαλαρωμένο πρόβλημα είναι η  $Z = \mathbf{D}^{-1/2} v_2$ , όπου  $v_i$  είναι τα ιδιοδιανύσματα του συμμετρικού Λαπλάσιαν. Μπορεί να αποδείξει κανείς ότι το  $Z$  είναι ιδιοδιάνυσμα του  $\mathbf{L}_{rw}$  που ορίσαμε στο προηγούμενο κεφάλαιο. Ο  $\mathbf{L}_{rw}$  έχει ιδιοδιανύσματα

που έχουν μεγαλύτερη φυσική σημασία από αυτά του  $\mathbf{L}_{sym}$ , γιατί σε συνδεδεμένο γράφο το διάνυσμα  $\mathbf{1}_N$  είναι ιδιοδιάνυσμα του πίνακα και επομένως σε γράφο με πολλές συνεκτικές συνιστώσες η δείκτρια συνάρτηση κάθε συνεκτικής συνιστώσας είναι επίσης ιδιοδιάνυσμα του πίνακα. Επιπλέον, όταν κάνουμε μικρή διαταραχή προσθέτοντας λίγες ακμές στον γράφο, επειδή η συνάρτηση των ιδιοδιανυσμάτων επί των ακμών είναι συνεχής περιμένουμε ομαλή συμπεριφορά, δηλαδή σε γράφο με καλά ομαδοποιημένες κλάσεις περιμένουμε πως τα ιδιοδιανύσματα θα είναι κοντά στις δείκτριες συναρτήσεις κάθε κλάσης. Αυτό ισχύει προφανώς και για παραπάνω από 2 κλάσεις. Από την άλλη όμως είναι πιο αριθμητικά σταθεροί οι αλγόριθμοι που υπολογίζουν ιδιοδιανύσματα σε συμμετρικούς πίνακες των οποίων είναι εξασφαλισμένο ότι οι ιδιοτιμές είναι πραγματικές ενώ ο  $\mathbf{L}_{rw}$  δεν είναι συμμετρικός, οπότε αυτή η "συνέχεια" των ιδιοδιανυσμάτων χάνεται σαν πληροφορία στην αριθμητική αστάθεια. Υπάρχουν και άλλοι λόγοι που θα ήθελε κανείς να χρησιμοποιήσει τον  $\mathbf{L}_{rw}$ , όπως για την εποπτική επεξεργασία του διαγράμματος των ιδιοτιμών που όπως θα δούμε στην συνέχεια έχει μεγάλη σημασία. Γι'αυτό και προτείνεται η χρήση του από το [25]. Ωστόσο, αυτός δεν είναι αρκετός λόγος για να χρησιμοποιηθεί στον τελικό αλγόριθμο κατάτμησης αφού σαν προστάδιο για την εποπτεία μπορούμε να τον χρησιμοποιήσουμε ανεξάρτητα.

### 3.2.2 Σύγκριση με άλλους αλγορίθμους φασματικής κατάτμησης

Σε αυτό το σημείο θα θέλαμε να σταθούμε για να αναρωτηθούμε γιατί πρέπει κανείς να σταθμίσει τον πίνακα Λαπλάσιαν για να λύσει ένα πρόβλημα κατάτμησης. Η τετραγωνική μορφή που ελαχιστοποιήσαμε μοιάζει ελαφρά διαφορετική από την τετραγωνική μορφή του πραγματικού Λαπλάσιαν. Επειδή  $X^T \mathbf{L} X = 1/2 \sum_{i,j} w_{ij} (x_i - x_j)^2 = cut(A, \bar{A})$ , η ελαχιστοποίηση της τετραγωνικής μορφής της συνάρτησης μας δίνει την ελάχιστη τομή του γράφου. Φυσικά, πρέπει να κοιτάζουμε μη-σταθερά διανύσματα που μηδενίζουν αυτή την τετραγωνική μορφή. Το φυσικό νόημα αυτού είναι ότι δεν θέλουμε καμία από τις δύο κλάσεις να είναι κενή. Είδαμε όμως ότι η ελάχιστη τομή δεν δίνει καλά αποτελέσματα για εφαρμογές που πολύ μικρά cluster δεν έχουν φυσικό νόημα. Θα θέλαμε συνεπώς η αντικειμενική συνάρτησή μας να "τιμωρεί" πολύ μικρά clusters. Η στάθμιση όμως στο *Normalized Cut* είναι πιά εξεζητημένη από αυτή που θα μπορούσαμε να πετύχουμε απλά κοιτώντας διανύσματα που έχουν π.χ. σταθερό μέτρο. Στην πραγματικότητα το παραπάνω πρόβλημα που περιγράψαμε λύνει μια μετρική γνωστή ως *Ratio Cut*, η οποία ορίζεται ως:

$$Rcut(A_1, \dots, A_K) = \sum_{i=1}^K \frac{links(A_i, \bar{A}_i)}{size(A_i)} = \sum_{i=1}^K \frac{cut(A_i, \bar{A}_i)}{size(A_i)},$$

όπου με την έννοια  $size(A)$  ορίζουμε το πλήθος των στοιχείων μιας κλάσης. Η λύση αυτού του προβλήματος βελτιστοποίησης με την ίδια διαδικασία με παραπάνω είναι το δεύτερο (για 2 κλάσεις) ιδιοδιάνυσμα του *unnormalized* Λαπλάσιαν. Υπάρχουν πολύ καλοί λόγοι και εμπειρικοί και πρακτικοί για να αποφύγει κανείς αυτή την μετρική. Στο [26] αποδεικνύεται ότι η κατάτμηση που παράγει το Ratio - Cut δεν είναι "συνεπής", με την έννοια ότι η

κατάτμηση που γίνεται σε πεπερασμένο αριθμό στοιχείων δεν δίνει μια χρήσιμη κατάτμηση όσο το μέγεθος των δεδομένων τείνει στο άπειρο. Από την άλλη, αυτό που επηζητάμε όταν κάνουμε κατάτμηση είναι η δημιουργία κλάσεων με μεγάλη εσωτερική ομοιογένεια και μεγάλη εξωτερική ανομοιογένεια. Κοιτάζοντας μόνο ελάχιστες τομές εξασφαλίζουμε μόνο το δεύτερο. Ωστόσο, όταν πρόκειται για την σταθμισμένη τομή παρατηρούμε το εξής εντυπωσιακό. Ορίζοντας κανείς με αντίστοιχο τρόπο το *Normalized Association*, ως:

$$Nassoc(A_1, \dots, A_K) = \sum_{i=1}^K \frac{assoc(A_i)}{vol(A_i)},$$

κάποιος θα ήθελε ταυτόχρονα να ελαχιστοποιήσει την τομή και να μεγιστοποιήσει το *association*. Αυτές οι δύο ανταγωνιστικές μετρικές φαίνεται να ικανοποιούνται μαζί όμως στην περίπτωση μας. Χρησιμοποιώντας κανείς την ταυτότητα:  $assoc(A) + cut(A) = vol(A)$ , μπορεί να αποδείξει ότι:  $Ncut(A_1, \dots, A_K) = K - Nassoc(A_1, \dots, A_K)$ . Αντίστοιχο ορισμό για το *Ratio Cut* δεν έχουν αυτή την ιδιότητα. Γιαυτό το λόγο καταλήγουμε με μια μετρική στην οποία το μέγεθος ενός cluster μετριέται με βάση το βάρος των συνολικών ακμών και όχι με το πλήθος τους. Στην βιβλιογραφία ([4]) το αντίστοιχο αποτέλεσμα έχει χρησιμοποιηθεί σε αλγόριθμους κατάτμησης με τυχαίους περιπατητές. Το ίδιο κάνει και ο αλγόριθμος του *Page – Rank* [30] για την ταξινόμηση της σημαντικότητας των ιστοσελίδων. Μετράει την σημασία ενός κόμβου (μιας ιστοσελίδας) από την σημασία των κόμβων στις οποίες δείχνει. Η αναδρομικότητα του ορισμού αυτού γεννάει την ιδέα για κατασκευή μιας επαναληπτικής διαδικασίας στην οποία ζητάμε τα στάσιμα σημεία. Μοντελοποιώντας μάλιστα αυτό το πρόβλημα σαν τυχαίο περίπατο σε μια μαρκοβιανή αλυσίδα, καταλήγουμε να ζητάμε την αναλλοίωτη κατανομή της αλυσίδας αυτής. Μπορούμε να αποδείξουμε ότι αυτή είναι το μέγιστο ιδιοδιάνυσμα σε ένα πίνακα γειτνίασης μετασχηματισμένο ώστε να γίνει στοχαστικός.

### 3.2.3 Επαναφορά στον διακριτό χώρο

Παρότι λύσαμε το πρόβλημα στο συνεχή χώρο, για να ολοκληρώσουμε έναν προσεγγιστικό αλγόριθμο χρειαζόμαστε και το βήμα που κάνει την διακριτοποίηση της λύσης μας. Αυτό το βήμα είναι ακόμη πιο σημαντικό από το προηγούμενο και μάλιστα μερικές φορές στην θεωρία βελτιστοποίησης προτιμούν να μην λύνουν το πρόβλημα στο συνεχή χώρο επακριβώς αν αυτό σημαίνει ότι θα πλησιάσει το *rounding scheme* πιο κοντά στην βέλτιστη διακριτή λύση.

Μια πρώτη προσπάθεια για την εύρεση μιας καλής διακριτής λύσης είναι να αναθέσει κανείς κόμβους σε κλάσεις ανάλογα με το πρόσημό που έχουν στο δεύτερο ιδιοδιάνυσμα. Παρατηρώντας κανείς την λύση μας στο συνεχή χώρο, δηλαδή το ιδιοδιάνυσμα  $u_2$  του  $\mathbf{L}_{rw}$ , βλέπουμε ότι αυτό πρέπει να είναι κάθετο στο σταθερό διάνυσμα, που είναι το πρώτο ιδιοδιάνυσμα του ίδιου πίνακα. Αυτό σημαίνει όμως ότι το άθροισμα των συνεταγμένων του είναι ίσο με μηδέν, και αφού δεν είναι το μηδενικό ιδιοδιάνυσμα τότε αναγκαστικά έχει και θετικές και αρνητικές συνιστώσες. Αν υπάρχουν και συνιστώσες με μηδενική τιμή



τότε υπάρχουν έξυπνες ευρετικές που κοιτάνε τα μεγέθη των κλάσεων και μοιράζουν τους κόμβους ανάλογα. Ωστόσο, αυτό παρουσιάζει μικρό ενδιαφέρον στην πράξη, γιατί οι γράφοι, αν δεν έχουν πολλές συμμετρίες δεν δίνουν ποτέ μηδενική συντεταγμένη. Πως καταλήγουμε να μοιράζουμε τους κόμβους με βάση το πρόσημο του ιδιοδιανύσματος όμως; Ξέρουμε σίγουρα ότι οι τιμές  $a, b$  της διακριτής λύσης θα είναι ετερόσημες, γιατί τις περιορίσαμε έτσι ώστε να ισχύει:  $\text{vol}(A)a + \text{vol}(\bar{A})b = 0$ . Όμως, πως γνωρίζουμε ότι αν κάποιος κόμβος είναι αρνητικός στην διακριτή λύση, τότε είναι και στην συνεχή και ανάποδα; Μια διασθητική ερμηνεία που χρησιμοποιεί θεωρήματα από διαφορική γεωμετρία, λέει το εξής. Γνωρίζουμε ότι τα σημεία μας είναι σημεία στον προβολικό χώρο  $n-1$  διαστάσεων. Αυτό που στην ουσία ζητάμε να βρούμε είναι πόσο κοντά είναι η λύση  $Z$  του συνεχούς χώρου, στα σημεία του εφικτού χώρου λύσεων  $\mathcal{X}$ . Ή με άλλα λόγια ποιό σημείο  $\mathbb{P}(X) \in \mathbb{P}(\mathcal{X})$  είναι εγγύτερα στην προσεγγιστική λύση  $\mathbb{P}(Z) \in \mathbb{RP}^{n-1}$ . Επειδή το  $\mathcal{X}$  είναι κλειστό ως προς το *antipodal map* το να ελαχιστοποιήσει κανείς την απόσταση  $d(\mathbb{P}(X), \mathbb{P}(Z))$  στο  $\mathbb{RP}^{n-1}$  είναι ισοδύναμο με το να ελαχιστοποιήσει την ευκλείδια απόσταση  $\|X - Z\|_2$ , όπου τώρα οι εκπρόσωποι των προηγούμενων σημείων βρίσκονται πάνω στην μοναδιαία σφαίρα. Επομένως, το τελευταίο είναι ισοδύναμο με το να μεγιστοποιήσει κανείς το εσωτερικό γινόμενο των διανυσμάτων. Αυτό σημαίνει λοιπόν ότι (έστω χ.β.γ.  $a > 0, b < 0$ ) θα αντιστοιχίσουμε όλες τις θετικές τιμές του  $Z$  με το  $a$  και όλες τις αρνητικές με  $b$ , κατατμώντας έτσι τους κόμβους ανάλογα με το πρόσημό τους.

Μια βελτιωμένη λύση είναι να ταξινομήσει κανείς τις συντεταγμένες του διανύσματος και κατόπιν να υπολογίσει την τιμή του *Normalized Cut* στις  $n-1$  τομές που δημιουργούνται από τους κόμβους με τις μεγάλες τιμές και τους κόμβους με τις μικρές τιμές του διανύσματος. Ο αλγόριθμος αυτός ονομάζεται *sweep algorithm* και περιγράφεται αναλυτικά στο παρακάτω σχήμα: Αυτός ο αλγόριθμος υπολογίζει την τιμή της σταθμισμένης

---

#### Algorithm 1 Sweep Algorithm

---

- 1: **Input:** graph  $\mathcal{G} = (\mathcal{V}, \mathcal{E})$  and vector  $x \in R^{|\mathcal{V}|}$ .
  - 2: Sort the vertices of  $V$  in non-decreasing order of values of entries in  $x$ , that is let  $V = \{u_1, \dots, u_n\}$  where  $x_{u_1} \leq x_{u_2} \leq \dots \leq x_{u_n}$ .
  - 3: Let  $i \in \{1, \dots, n-1\}$  be such that  $Ncut(\{u_1, \dots, u_i\}, \{u_{i+1}, \dots, u_n\})$  is minimal.
  - 4: Output  $A = \{u_1, \dots, u_i\}, \bar{A} = \{u_{i+1}, \dots, u_n\}$ .
- 

τομής πολλές φορές αλλά στην πραγματικότητα δεν παίρνει παραπάνω χρόνο ασυμπτωτικά. Αν τον υλοποιήσουμε έξυπνα η πιο βαριά διαδικασία θα παραμένει η εύρεση ιδιοτιμών. Στην πραγματικότητα, αν έχουμε υπολογίσει μόλις το:  $Ncut(\{u_1, \dots, u_i\}, \{u_{i+1}, \dots, u_n\})$ , το να υπολογίσουμε το:

$Ncut(\{u_1, \dots, u_{i+1}\}, \{u_{i+2}, \dots, u_n\})$  παίρνει χρόνο μόνο  $\mathcal{O}(\text{degree}(u_{i+1}))$  γιατί χρειάζεται μόνο να μεταφέρουμε το τελευταίο κόμβο στην άλλη κλάση και άρα να προσθέσουμε τις κατάλληλες ακμές του στην τομή, να αφαιρέσουμε τις υπόλοιπες από την προηγούμενη τομή και να ανανεώσουμε με τον ίδιο τρόπο και το *volume*. Αυτός ο αλγόριθμος παρότι δεν χρησιμοποιείται στο χώρο του *computer vision* έχει μεγάλο ερευνητικό ενδιαφέρον και

χρησιμοποιείται στο *theory* για ένα παρεμφερές πρόβλημα με το *Normalized Cut* που είναι το *sparsest cut* ή η εύρεση του *conductance* ενός γράφου που ορίζεται ως:

$$\varphi(G) = \min_{S \in V} \frac{\text{cut}(S)}{\min(\text{vol}(S), \text{vol}(\bar{S}))}$$

Εύκολα μπορεί κανείς να δει ότι το τελευταίο πρόβλημα είναι  $2 - \text{approximation}$  από το *Normalized Cut*. ( $\varphi \leq \text{Ncut} \leq 2\varphi$ ) Όμως, για το *conductance* έχει αποδειχθεί ότι:  $\lambda_2/2 \leq \varphi(G) \leq \sqrt{2\lambda_2}$ , όπου  $\lambda_2$  είναι η ιδιοτιμή του συμμετρικού λαπλάσιαν ( $[0,2]$ ). Η ανισότητα αυτή έχει ονομαστεί *Cheeger inequality* και αποτελεί αντικείμενο μεγάλης έρευνας. Μάλιστα η απόδειξη της ανισότητας χρησιμοποιεί σαν κατασκευαστικό αλγόριθμο τον *sweepalgorithm*. Τελικά, μπορεί κανείς να συμπεράνει πως η έξοδος του αλγορίθμου αυτού θα είναι:  $\varphi(S) \leq \sqrt{2\lambda_2} \leq 2\sqrt{\varphi(G)}$ . Δυστυχώς, η τετραγωνική προσέγγιση δείχνει ότι ο αλγόριθμος έχει πολύ κακά *worst - case* σενάρια. Μάλιστα, για να αποδειχθεί ότι ο τετραγωνικός λόγος προσέγγισης δεν είναι κάποιο τεχνητό σημείο που εισέρχεται από την απόδειξη, έχουν βρεθεί γράφοι για τους οποίους ισχύει επακριβώς ο τετραγωνικός λόγος. Αυτοί είναι οι γράφοι *Ramanujan*. Ο συγκεκριμένος λόγος προσέγγισης βέβαια είναι αποτέλεσμα του προσεγγιστικού αλγορίθμου που επιλέξαμε να λύσουμε. Υπάρχουν και άλλοι τρόποι να κάνει κανείς χαλάρωση στο πρόβλημα. Μπορούμε να παρατηρήσουμε ότι: Αντί να χρησιμοποιηθεί η τετραγωνική μορφή  $X^T \mathbf{L} X = 1/2 \sum_{i=1}^n w_{ij}(x_i - x_j)^2$  για να μοντελοποιηθεί το *cut*, μπορεί να χρησιμοποιηθεί στο διακριτό χώρο και η *L1* απόσταση αφού:  $1/2 \sum_{i=1}^n w_{ij}|x_i - x_j|$  είναι επίσης ίσο με την τομή όταν το διάνυσμα παίρνει *binary* τιμές. Μάλιστα, με αυτή τη λογική οι συγγραφείς του [3], προτείνουν την εξής χαλάρωση:

$$LR(G, H) = \min_{\substack{d: V \times V \rightarrow \mathbb{R} \\ d \text{ semi-metric}}} \frac{|E_H|}{|E_G|} \cdot \frac{\sum_{\{u,v\}} G_{u,v} d(u, v)}{\sum_{\{u,v\}} H_{u,v} d(u, v)}$$

όπου ο  $G$  είναι ο γράφος που θέλουμε να κατατμήσουμε και ο  $H$  είναι ο μετασχηματισμένος γράφος που δημιουργείται από τον  $G$  αν κρατήσουμε ίδιες τις κορυφές και αντικαταστήσουμε όλες τις ακμές μεταξύ δύο κορυφών  $\{i, j\}$  με ακμές βάρους  $d_i d_j$ . Αποδεικνύουν οι συγγραφείς ότι:

$$LR(G, H) \leq \varphi(G, H) \leq \mathcal{O}(\log |V|) \cdot LR(G, H)$$

Επίσης αν κανείς απαιτήσει μόνο *negative type semi - metric* (δηλαδή *semi - metric* των οποίων η ρίζα είναι ευκλείδεια *semi - metric*).

### 3.2.4 Υπολογιστική πολυπλοκότητα και αλγόριθμοι

Σε μια εικόνα διαστάσεων  $300 \times 300$  έχουμε 90000 pixels. Δηλαδή (αν μείνουμε με αυτή την διακριτοποίηση της εικόνας) ο πίνακας του γράφου μας θα είχε παραπάνω από 8 δισεκατομμύρια θέσεις. Ο υπολογισμός των ιδιοτιμών (που με τους κλασσικούς αλγόριθμους είναι κυβικού χρόνου στο πλήθος των *pixels*) ενός τέτοιου πίνακα δεν θα ήταν ρεαλιστικά υλοποιήσιμος από κανένα υπολογιστή. Γιαυτό θα πρέπει να καταφύγουμε σε μεθόδους που θα διατηρούν όσο γίνεται περισσότερη πληροφορία από τον γράφο αλλά θα

ελαττώνουν την χρονική πολυπλοκότητα του αλγορίθμου. Μια τέτοια πρώτη μέθοδος είναι η δημιουργία των *superpixel*. Είδαμε ότι η ιεραρχική αυτή αντιμετώπιση όχι μόνο είναι ταχύτερη αλλά συνήθως καταλήγει σε πολύ καλά αποτελέσματα. Παρόλαυτά, δεν αρκεί μόνο αυτό. Η επόμενη βελτίωση λοιπόν είναι να εκμεταλλευτούμε την αραιή δομή του γράφου που κατασκευάσαμε. Στον *RAG* γράφο ενσωματώνουμε μόνο τοπική πληροφορία από τα *superpixel* και επομένως δεν συνδέουμε κόμβους που είναι πολύ μακριά μεταξύ τους. Συνεπώς, ο πίνακας του γράφου θα είναι εξαιρετικά αραιός ( $O(N)$  μη μηδενικές θέσεις). Αυτό δεν το εκμεταλλεύονται κλασικοί αλγόριθμοι που διαγωνιοποιούν τον πίνακα ή κάνουν *QR* διάσπαση. Το πρώτο που πρέπει να σκεφτούμε είναι ότι όταν θέλουμε να κατατμήσουμε τον πίνακα σε 2 κλάσεις δεν χρειάζεται να υπολογίσουμε όλο το φάσμα του αλλά μόνο το δεύτερο ιδιοδιάνυσμα. Γι'αυτό το σκοπό θα τροποποιήσουμε τον αλγόριθμο *power method* για να υπολογίζει το ιδιοδιάνυσμα αυτό.

### Power Iteration για τον υπολογισμό του δεύτερου ιδιοδιανύσματος του graph Laplacian

Ο αλγόριθμος *Power Iteration* χρησιμοποιείται για τον υπολογισμό της μεγαλύτερης ιδιοτιμής ενός πίνακα  $A$ . Αρχικά επιλέγεται ένα διάνυσμα  $b_0$  που μπορεί να είναι ή όχι μια προσέγγιση του ζητούμενου ιδιοδιανύσματος. Κατόπιν, επαναλαμβάνεται αρκετές φορές η αναδρομική σχέση:  $b_{k+1} = \frac{Ab_k}{\|Ab_k\|}$  μέχρι την σύγκλιση. Η σύγκλιση αυτή είναι υπό προϋποθέσεις όμως. Η  $k$ -οστός όρος μπορεί να γραφεί και ως:  $b_k = e^{i\varphi_k} u_1 + r_k$ , όπου  $u_1$  είναι το μεγαλύτερο ιδιοδιάνυσμα του πίνακα  $A$  και  $\|r_k\| \rightarrow 0$ . Το αποτέλεσμα αυτό σημαίνει πως ο αλγόριθμος δεν μπορεί να συγκλίνει εκτός και αν  $e^{i\varphi_k} = 1$ . Αποδεικνύεται ότι αν ο πίνακας έχει ιδιοτιμή με μεγαλύτερο μέτρο από τις υπόλοιπες και η επιλογή του αρχικού διανύσματος  $b_0$  είναι τέτοια ώστε να έχει μη μηδενική τιμή στην κατεύθυνση του ιδιοδιανύσματος αυτής της ιδιοτιμής, η ακολουθία  $b_k$  συγκλίνει όντως στο μεγαλύτερο ιδιοδιάνυσμα και η ακολουθία  $\mu_k = \frac{b_k^* Ab_k}{b_k^* b_k}$  στην μεγαλύτερη ιδιοτιμή. Υπάρχει και μια πιθανοτική εκδοχή του αλγορίθμου αυτού στο [42], που δεν θα μας απασχολήσει εδώ όμως.

Το να υπολογίσει κανείς την δεύτερη ιδιοτιμή με αυτό τον αλγόριθμο οφείλεται στις ιδιότητες που έχει ο *graph Laplacian*, που είναι συμμετρικός και θετικά ημιορισμένος. Μάλιστα η πρώτη ιδιοτιμή του είναι 0 ενώ η δεύτερη για συνεκτικούς γράφους αυστηρά μεγαλύτερη του μηδενός. Γιατί όμως ξαφνικά μιλάμε για τις μικρότερες ιδιοτιμές; Η μέθοδος βασίζεται στο γεγονός ότι ο πίνακας είναι θετικά ημιορισμένος πράγμα που χάλαι αν υπολογίσουμε με την παραπάνω μέθοδο τις μεγαλύτερες ιδιοτιμές του αντίθετού του. Η άλλη ιδιότητα που έχει όμως ο *graph Laplacian* είναι ότι οι ιδιοτιμές του είναι φραγμένες. Ο συμμετρικός *graph Laplacian* έχει όλες τις ιδιοτιμές του στο διάστημα  $[0, 2]$ . Επομένως, εύκολα παρατηρεί κανείς ότι ο πίνακας  $\mathbf{M} = 2\mathbf{I} - \mathbf{L}_{sym}$  είναι επίσης θετικά ημιορισμένος, οι μικρότερες ιδιοτιμές του αντιστοιχούν στις μεγαλύτερες του αρχικού, ενώ και οι δύο πίνακες μοιράζονται τα ίδια ιδιοδιανύσματα. Επιπλέον, θα μπορούσε κανείς να υπολογίσει τις μεγαλύτερες ιδιοτιμές και ιδιοδιανύσματα του ψευδοαντιστρώφου του γράφου Λαπλάσιαν, τεχνική που εξαιτίας του αλγορίθμου *Teng – Spielman* ([39]) για την επίλυση

του συστήματος  $L_{sym}x = y$  δίνει ακόμη γρηγορότερα αποτελέσματα και σχεδόν γραμμική πολυπλοκότητα ανεξάρτητη από την σχέση της δεύτερης ιδιοτιμής με το πλήθος των κόμβων του γράφου. Μένει λοιπόν να δείξουμε ότι μπορούμε χρησιμοποιώντας τον προηγούμενο αλγόριθμο να υπολογίσουμε την δεύτερη ιδιοτιμή ενός πίνακα. Η ιδέα βασίζεται στο γεγονός ότι αν επιλέξουμε να εφαρμόσουμε τον αλγόριθμο σε έναν υπόχωρο κάθετο στο πρώτο ιδιοδιάνυσμα θα ανακαλύψουμε το δεύτερο. Γιαυτό μπορούμε να ξεκινήσουμε επιλέγοντας ένα αρχικό διάνυσμα σαυτό τον υπόχωρο. Έχοντας υπολογίσει λοιπόν το πρώτο ιδιοδιάνυσμα  $v$ , μπορούμε να επιλέξουμε σαν αρχικό διάνυσμα το  $x = x - (x^T v) \cdot v$ , όπου το  $x$  επιλέγεται τυχαία ώστε να είναι αρκετά πιθανό να έχει μη μηδενική συνιστώσα στις κατευθύνσεις των πρώτων δύο ιδιοδιανυσμάτων. Στο [42], περισσότερες λεπτομέρειες για το πως μπορεί να γίνει de-randomization του αλγορίθμου αυτού.

Παρατηρούμε επίσης το πλεονέκτημα που μας δίνει η επαναληπτική μέθοδος εξαιτίας του γεγονότος ότι η πολυπλοκότητα του αλγορίθμου είναι ανάλογη της ακρίβειας που επιλέγουμε να έχουμε στην απάντηση μας για το δεύτερο ιδιοδιάνυσμα. Επειδή χρειαζόμαστε μόνο το πρόσημο του δεύτερου ιδιοδιανύσματος για να κάνουμε την κατάτμηση και ξέρουμε ότι εξαιτίας της μορφής του πίνακα οι ιδιοτιμές δεν είναι εξαιρετικά ευαίσθητες σε μικρές διαταραχές, μπορούμε να σταματήσουμε την εκτέλεση του αλγορίθμου μετά από λίγες επαναλήψεις.

Στα επόμενα κεφάλαια που θα μιλήσουμε για περισσότερες κλάσεις και υπολογισμό περισσότερων ιδιοδιανυσμάτων θα καθίσταται ακόμη περισσότερο αναγκαία η εύρεση μιας γενικευμένης μεθόδου για τον υπολογισμό ιδιοδιανυσμάτων από αραιούς πίνακες. Γιαυτό θα καταφύγουμε σε *Krylov subspace* μεθόδους όπως η *Lanczos* ([46]), που μειώνουν την πολυπλοκότητα εύρεσης ιδιοτιμών σε πίνακες διαστάσεων  $O(N)$  από κυβική σε  $O(N^{3/2})$ .

### 3.2.5 Ένα σχόλιο για την δεύτερη και τις μεγαλύτερες ιδιοτιμές

Η δεύτερη ιδιοτιμή είδαμε ότι είναι ένα μέτρο της συνεκτικότητας του γράφου. Αν αντιμετωπίζαμε την ιδανική περίπτωση που ο γράφος είχε  $K$  συνεκτικές συνιστώσες, τότε ο χωρισμός είναι προφανής και θα μπορούσαμε να τον βρούμε με έναν αλγόριθμο αναζήτησης γράφου όπως ο *DFS*, που είναι γραμμικού χρόνου και δεν απαιτεί εύρεση ιδιοτιμών. Ωστόσο, ποτέ δεν θα αντιμετωπίζουμε στην πράξη αυτή την ιδανική περίπτωση και μεταξύ των κλάσεων θα υπάρχουν λίγες ή περισσότερες θορυβώδεις ακμές. Για αυτό το λόγο χρειαζόμαστε μια *robust* εκδοχή της συνεκτικότητας του γράφου. Παρατηρούμε από την ανισότητα *Cheeger*, ότι όσο μικρότερη είναι η δεύτερη ιδιοτιμή τόσο μικρότερη περιορίζεται να είναι η αραιότερη τομή δημιουργώντας ένα γράφο με καλά διαχωρισμένες κλάσεις. Από την άλλη, μια κλίμα, είδαμε ότι έχει πολύ μεγάλη δεύτερη ιδιοτιμή ( $K_n : n/(n-1)$ ). Ωστόσο, ειδικά για μεγάλους γράφους, αυτή η τιμή είναι πολύ μακριά από την μέγιστη ιδιοτιμή που μπορεί να έχει ο συμμετρικός γράφος Λαπλάσιαν, που είναι 2. Τι συμβαίνει όσο η μεγαλύτερη ιδιοτιμή πλησιάζει το 2; Γνωρίζουμε από πριν ότι όταν η μεγαλύτερη ιδιοτιμή είναι ίση με 2, τότε ο γράφος είναι διμερής ή έχει μια διμερή συνεκτική συνιστώσα, αν δεν είναι συνεκτικός. Επομένως θα πρέπει να σκεφτούμε μήπως η μεγαλύτερη ιδιοτιμή είναι και μια

*robust* εκδοχή της διμερότητας του γράφου, ακριβώς όπως κάναμε με την συνεκτικότητα. Πράγματι αυτή η έννοια υπάρχει στην βιβλιογραφία ([42]). Αν ορίσουμε την διμερότητα ως:

$$\beta(S, A, B) = \frac{2E(A) + 2E(B) + 2E(S, V - S)}{\text{vol}(S)}$$

όπου  $S$  είναι ένα σετ από κόμβους του γράφου που θέλουμε να μοιράσουμε σε δύο μη επικαλυπτόμενα σετ  $A, B$ , τότε καταρχάς παρατηρούμε ότι η μετρική αυτή κοιτάει τι ποσοστό των ακμών του  $S$  πρέπει να διαγράψουμε για να αποσυνδέσουμε το  $S$ , από τον υπόλοιπο γράφο και τα  $A, B$  να αποτελούν ένα *bipartition* για το  $S$ . Επιπλέον αν ορίσουμε την διμερότητα του γράφου ως την μικρότερη τιμή της διμερότητας πάνω σε όλα τα σετ κορυφών και όλα τα *bipartitions*, αποδεικνύεται ότι:

$$1/2 \cdot (2 - \lambda_n) \leq \beta(G) \leq \sqrt{2 \cdot (2 - \lambda_n)}$$

### Γιατί το spectral clustering λειτουργεί

Μπορούμε να πούμε και άλλες ιδιότητες του γράφου που αποκαλύπτονται μέσα από τις ιδιοτιμές. Παραδείγματος χάρη, οι ιδιοτιμές αποκαλύπτουν πληροφορία για τα *mixing times* μιάς μαρκοβιανής αλυσίδας. Επίσης, δίνουν καλά πάνω και κάτω όρια για την διάμετρο του γράφου κλπ. Γιατί όμως οι ιδιοτιμές κρύβουν τόσο μεγάλη πληροφορία; Ποια η διαίσθηση πίσω από την εξαιρετική επίδοση που έχει σε προβλήματα γραφοτομών το πρόσημο του δεύτερου ιδιοδιανύσματος; Ή γενικά ο χώρος που καλύπτουν τα πρώτα  $k$  ιδιοδιανύσματα; Από την πλευρά της θεωρίας αλγορίθμων όπως είδαμε και θα επεκτείνουμε στην συνέχεια, η απάντηση είναι ότι αποτελεί το relaxation ενός προβλήματος συνδυαστικής βελτιστοποίησης με θεωρητικές εγγυήσεις για το χειρότερο λόγο προσέγγισης. Από την περιοχή των statistics ή του machine learning, η εξήγηση δίνεται επειδή βρίσκει τα λανθάνοντα cluster της υπόθεσης παραδείγματος χάρη από stochastic blockmodels ή manifold χαμηλής διάστασης. Τα αποτελέσματα αυτά ενισχύονται από την θεωρία διαταραχών [29], σύμφωνα με την οποία, ξεκινάει κανείς από μια ιδανική κατάσταση, η οποία έχει εύκολη λύση και προχωράει σε μια προσεγγιστική λύση της γενικής περίπτωσης ανάγοντάς την στην ιδανική. Στο πρόβλημά μας ξεκινάμε από ένα γράφο με ήδη διαχωρισμένα cluster. Ο Λαπλάσιαν πίνακας αυτού του γράφου είναι ένας block διαγώνιος πίνακας που εξακολουθεί να έχει τις ιδιότητες ενός γράφου Λαπλάσιαν πάνω σε κάθε συνεκτική συνιστώσα. Το αποτέλεσμα της κατάτμησης αυτής είναι εύκολο να βρεθεί σε γραμμικό χρόνο με αναζήτηση γράφου, αλλά στην περίπτωσή μας, οι ιδιοτιμές θα αποτελούν δείκτριες συναρτήσεις καθεμιάς συνεκτικής συνιστώσας. Αφού κάθε γραμμή θα έχει μόνο ένα μη μηδενικό στοιχείο μπορούμε να επιλέξουμε όλα τα διανύσματα γραμμές να βρίσκονται πάνω σε μια σφαίρα  $K - 1$  διαστάσεων. Κατόπιν, εφαρμόζουμε μια διαταραχή στον πίνακα Λαπλάσιαν, με την προσθήκη θορύβου  $E$ , έτσι ώστε  $\tilde{\mathbf{L}} = \mathbf{L} + \mathbf{E}$ . Αποδεικνύεται λοιπόν ότι αν ο θόρυβος δεν είναι τόσο δυνατός ώστε να αλλοιώνει τις σημαντικές ιδιότητες του γράφου, όπως την καλή διαχωρισιμότητα των κλάσεων (που σχετίζεται με το *eigengap* του γράφου) ή την ισχυρή σύνδεση των εσωτερικών κόμβων κάθε κλάσης, τότε είναι δυνατή η πλήρης ανακατασκευή των κλάσεων από

τον "θορυβώδη γράφο". Μια άλλη σκοπιά που μπορεί να δει κανείς το χώρο που καλύπτουν τα  $k$  πρώτα ιδιοδιανύσματα είναι το dimensionality reduction. Αν από την άλλη γνωρίζουμε ότι ο πίνακας, αφού ο γράφος έχει  $k$  κλάσεις, θα πρέπει να έχει βαθμό  $k$ , προσπαθούμε να βρούμε τον πίνακα εκείνο που είναι πιο κοντά με κάποια έννοια απόστασης στον πίνακα του θορυβώδους γράφου μας και έχει βαθμό  $k$ . Αν επιλέξουμε την *Frobenius* νόρμα σαν απόσταση των πινάκων, το πρόβλημα καταλήγει στην προσέγγιση του πίνακα από τις  $k$  μεγαλύτερες ιδιοτιμές του ως:  $\mathbf{A}_{rank\ k} = \mathbf{U}_{1:k}\mathbf{\Sigma}_{1:k}\mathbf{U}_{1:k}^\top$  (θυμηθείτε ότι μιλάμε για τον πίνακα βαρών).

Υπάρχουν και άλλες εξηγήσεις διάσπαρτες στην βιβλιογραφία οι οποίες ανάγονται σε άλλα γνωστά προβλήματα που έχουν μελετηθεί και έχουν μια πιο διασθητική ερμηνεία για το πως λειτουργούν. Παραδείγματος χάρη, στο [34] ο συγγραφέας συσχετίζει τους αλγόριθμους κατάτμησης γράφων με τον αλγόριθμο K-means ο οποίος εφαρμόζεται όμως σε ένα χώρο διαφορετικό στον οποίο έχουν προβληθεί πρώτα τα δεδομένα μέσω ενός πυρήνα. Επίσης η συνεισφορά κάθε σημείου στο κεντροειδές δεν είναι ομοιόμορφη αλλά εξαρτάται από ένα πίνακα βαρών συμμετοχής. Ο γενικευμένος αλγόριθμος kmeans λειτουργεί όπως φαίνεται στο παρακάτω σχήμα 3.5:

**WEIGHTED\_KERNEL\_KMEANS**( $K, k, w, C_1, \dots, C_k$ )

**Input:**  $K$ : kernel matrix,  $k$ : number of clusters,  $w$ : weights for each point

**Output:**  $C_1, \dots, C_k$ : partitioning of the points

1. Initialize the  $k$  clusters:  $C_1^{(0)}, \dots, C_k^{(0)}$ .
2. Set  $t = 0$ .
3. For each point  $\mathbf{a}$ , find its new cluster index as

$$j^*(\mathbf{a}) = \operatorname{argmin}_j \|\phi(\mathbf{a}) - \mathbf{m}_j\|^2, \text{ using (2).}$$

4. Compute the updated clusters as

$$C_j^{t+1} = \{\mathbf{a} : j^*(\mathbf{a}) = j\}.$$

5. If not converged, set  $t = t + 1$  and go to Step 3; Otherwise, stop.

Σχήμα 3.5: Ο αλγόριθμος του Kernel Kmeans

Η συσχέτιση γίνεται μέσα από την ανακάλυψη ότι τα δύο προβλήματα λύνουν το ίδιο πρόβλημα μεγιστοποίησης ενός ίχνους μιας τετραγωνικής μορφής πινάκων. Για τον αλγόριθμο του Normalized Cuts, ο πυρήνας αυτός είναι ο  $\mathbf{K} = \mathbf{D}^{-1}\mathbf{A}\mathbf{D}^{-1}$  και ο πίνακας βαρών  $\mathbf{W} = \mathbf{D}$ , όπου προφανώς  $A$  είναι ο πίνακας βαρών του γράφου και  $D$  ο πίνακας βαθμών των κορυφών. Θα λέγαμε λοιπόν ότι η επαναληπτική διαδικασία του αλγορίθμου Weighted Kernel Kmeans είναι μια μέθοδος για να επιλύσει κανείς το πρόβλημα των σταθμισμένων τομών χωρίς τον υπολογισμό των ιδιοτιμών. Πράγματι, στο [11] αναλύεται αυτή η ερμηνεία. Μια ακόμη χρήση του αλγορίθμου αυτού είναι σαν *post processing* βήμα τοπικής αναζήτησης



(local search), μετά την γραφοτομή με *normalized cuts*.

Μια ακόμη ερμηνεία συσχετίζει τις σταθμισμένες τομές με τους τυχαίους περιπατητές ([28]). Συγκεκριμένα, παρουσιάζει τις φασματικές τομές σαν την στάσιμη κατανομή της μαρκοβιανής αλυσίδας των περιπατητών, σαν την μόνιμη κατάσταση δηλαδή της μεταβατικής κατάστασης του περιπάτου. Αρχικά, κανείς μπορεί να ορίσει μια κατανομή  $\pi = \frac{d_i}{volG}$  και κατόπιν να αποδείξει εύκολα ότι είναι στάσιμη αν ο πίνακας πιθανοτήτων μετάβασης ήταν ο  $\mathbf{P} = \mathbf{D}^{-1}\mathbf{W}$ , αλλά και ότι η αλυσίδα είναι αντιστρέψιμη (αν είναι πρώτα εργοδική), δηλαδή  $\pi_i P_{ij} = \pi_j P_{ji} = W_{ij}/VolI$ . Αν επιπλέον ορίσουμε ως  $\Pr[A \rightarrow B|A]$ , την πιθανότητα μετάβασης του περιπατητή από ένα σετ  $A$  σε ένα σετ  $B$  με ένα βήμα, δεδομένου ότι αρχική κατάσταση είναι η  $A$ , τότε δείχνεται ότι:

$$NCut(A, A') = \Pr[A \rightarrow A'|A] + \Pr[A' \rightarrow A|A']$$

. Αυτό σημαίνει ότι οι περίπατοι εγκλωβίζονται σε ένα cluster ή όπως αναφέρεται στην βιβλιογραφία των τυχαίων περιπατητών, επηρεάζονται από τα σετ με "χαμηλή αγωγιμότητα" (low conductivity)

Τέλος, αξίζει να αναφέρουμε μια ερμηνεία που χτίζεται ακόμη στην βιβλιογραφία για τις μεθόδους των φασματικών γραφοτομών, που είναι η  $\Gamma$ -σύγκλιση ([7]). Η  $\Gamma$ -σύγκλιση είναι ένα εργαλείο που μελετάει την ασυμπτωτική συμπεριφορά που έχει μια οικογένεια από προβλήματα ελαχιστοποίησης. Ο στόχος είναι η αντικατάσταση της οικογένειας των ελαχιστων από ένα μονο πρόβλημα του οποίου τα ελάχιστα ( $\Gamma$ -όριο) έχουν κάποια ενδιαφέρουσα συμπεριφορά. Αν ορίσουμε την οικογένεια των προβλημάτων ελαχιστοποίησης ως

$$\min \{F_p(x) : x \in X \subset \mathbb{R}^n\}$$

με ελάχιστα τα  $x_p^*$ , τότε ψάχνουμε το :

$$x^* = \lim_{p \rightarrow \infty} \operatorname{argmin}_x F_p(x)$$

Μπορεί να υπάρχουν πολλές τέτοιες ακολουθίες. Αν επιλέξουμε ως προβλήματα τα:

$$F_p(x) = \sum_{i=1}^n \alpha_i^p Q_i(x)$$

όπου  $1 \geq \alpha_n > \alpha_{n-1} > \dots > \alpha_1 > 0$ , και οι  $Q_i(x)$  είναι ομαλές συναρτήσεις και υπάρχει ένα συμπαγές σετ  $C$  στο οποίο ανήκουν όλοι οι ελαχιστοποιητές, τότε αν  $M_n = \operatorname{argmin} Q_n(x), x \in C$  και  $M_k = \operatorname{argmin} Q_k(x) \quad x \in M_{k+1}, \forall k \in \{n-1, \dots, 0\}$ , τότε αποδεικνύεται ότι:  $x^* \in M_1$ . Αυτός ο αλγόριθμος καταλήγει να βρίσκει προσεγγιστικά το  $\Gamma$ -όριο, με την έννοια ότι το εντοπίζει σε ένα συγκεκριμένο σετ. Το πόσο κοντά στην πραγματική λύση είναι εξαρτάται από το συγκεκριμένο πρόβλημα.

Ερχόμενοι στο πρόβλημα των φασματικών γραφοτομών, αν ορίσουμε ένα *level graph*,  $\mathcal{G}_k$  πάνω σε όλες τις ακμές που έχουν βάρος  $w_k$ , τότε ο πίνακας βαρών του συγκεκριμένου επιπέδου έχει τιμές στο μόνο 0 ή 1. Συνεπώς, αν  $w_1 < \dots < w_j, j < |E|$ , τότε το πρόβλημα

του *Ratio Cut* γράφεται ως:

$$\text{Tr}(X^T \mathbf{L} X) = \sum_{i=1}^j w_k \text{Tr}(X^T \mathbf{L}_k X)$$

οπότε αν ορίσουμε τα  $Q_k(X) = \text{Tr}(X^T \mathbf{L}_k X)$ ,  $\alpha_k = w_k$ , θα θέλαμε να υπολογίσουμε το  $\sum_{k=1}^j w_k^p \text{Tr}(X^T \mathbf{L}_k X)$  για  $p \rightarrow \infty$ . Οι συγγραφείς του [7], παρουσιάζουν ένα αλγόριθμο που υπολογίζει προσεγγιστικά το  $\Gamma$ -όριο μέσα από τον υπολογισμό των ιδιοδιανυσμάτων κάθε συνεκτικής συνιστώσας στο *level graph* που δημιουργείται από την ένωση των γράφων με επίπεδο τουλάχιστον  $k$ . Αυτά τα ιδιοδιανύσματα όπως γνωρίζουμε θα είναι δείκτριες συναρτήσεις κάθε συνεκτικής συνιστώσας, οπότε μετά γίνεται ανανέωσή τους με την εισαγωγή των μικρότερων ακμών του επόμενου επιπέδου. Η αντιστοίχιση που κάνουν οι συγγραφείς με τις μεθόδους γραφοτομών από ελάχιστα συνεκτικά δέντρα, δίνει μια βαθύτερη ερμηνεία του γιατί οι μέθοδοι φασματικών γραφοτομών καταφέρνουν να ανακαλύπτουν μη κυρτά clusters.

### 3.3 Επέκταση φασματικής κατάτμησης σε πολλές κλάσεις

#### 3.3.1 Μοντελοποίηση Προβλήματος σε $K$ κλάσεις

Τώρα που είδαμε την διαίσθηση για το πως λειτουργεί ο αλγόριθμος των σταθμισμένων τομών, θα θέλαμε να τον επεκτείνουμε σε περισσότερες κλάσεις. Δεδομένου δηλαδή ενός γράφου  $\mathcal{G}(\mathcal{V}, \mathcal{K})$   $N$  κορυφών, και ενός αριθμού κλάσεων  $K$ , θα θέλαμε να κάνουμε μια διαμέριση του γράφου σε  $K$  υποσύνολα κορυφών. Η μετρική που θα θέλαμε να ελαχιστοποιήσουμε είναι το συνολικό άθροισμα των σταθμισμένων τομών μεταξύ ενός υποσυνόλου κορυφών και του υπόλοιπου γράφου. Φορμαλιστικά, για  $K$  κλάσεις σταθμισμένη τομή ορίζεται ως:

$$Ncut(A_1, \dots, A_K) = \sum_{i=1}^K \frac{\text{links}(A_i, \bar{A}_i)}{\text{vol}(A_i)} = \sum_{i=1}^K \frac{\text{cut}(A_i, \bar{A}_i)}{\text{vol}(A_i)}, \quad (3.5)$$

Με αφορμή την μοντελοποίηση σε δύο κλάσεις μπορεί κανείς να ορίσει ένα πίνακα μεγέθους  $N \times K$ ,  $\mathbf{X} = [X^1 \dots X^K]$  του οποίου οι στήλες θα είναι δείκτριες συναρτήσεις της διαμέρισης, δηλαδή  $X^j = (x_1^j, \dots, x_n^j)$  όπου  $x_i^j \in \{a_j, b_j\}$ , με  $a_j \neq b_j$ . Η στήλη  $X^j$  είναι επομένως μια δείκτρια συνάρτηση του  $A_j$  με την έννοια ότι:

$$x_i^j = \begin{cases} a_j, & u_i \in A_j \\ b_j, & u_i \notin A_j \end{cases}$$

Το σημαντικό είναι η απεικόνιση αυτή να δίνει ένα πίνακά που να έχει τους εξής περιορισμούς. Πρώτον, να δίνει μια *scale invariant* απεικόνιση των διαμερίσεων. Με την έννοια *scale invariance* θα εννοούμε πλέον το εξής: Αν σκεφούμε ότι ο πίνακας  $\mathbf{X}$  έχει ως στήλες  $K$  συναρτήσεις επί των κορυφών του γράφου, θα θέλαμε κάθε μια απο αυτές να παραμένει αμετάβλητη όταν πολλαπλασιαστεί με ένα αριθμό. Αυτή η απαίτηση έρχεται ως αποτέλεσμα του *scale invariance* της μετρικής που θέλουμε να ελαχιστοποιήσουμε. Επομένως τα



στοιχεία των στηλών του πίνακα δεν μπορούν να είναι σταθεροί αριθμοί όπως στην πρώτη εκδοχή της μοντελοποίησης των σταθμισμένων τομών σε πολλές κλάσεις ([49]). Θα περιορίσουμε τους πίνακες αναπαράστασης να έχουν αυτή την μορφή. Ο άλλος περιορισμός που πρέπει να βάλουμε είναι οι πίνακες αυτοί να αποτελούν πράγματι πίνακες διαμέρισης.

Εδώ πρέπει να σημειώσουμε ότι βιβλιογραφικά έχει προταθεί και η επέκταση των σταθμισμένων τομών με επαναληπτική κατάτμηση του μεγάλου κάθε φορά *cluster* σε δύο επιμέρους υποκλάσεις. Η ευρετική αυτή δεν λειτουργεί πολύ καλά στην πράξη σε περιπτώσεις που μπορεί να υπάρχουν παραπάνω από δύο αντικείμενα αλλά κάθε κατάτμηση να αποτυγχάνει να κόψει ένα από αυτά. Στην ιδανική περίπτωση που ο γράφος είχε τις κλάσεις πολύ καλά διαχωρισμένες οι δύο λύσεις θα ήταν ισοδύναμες αλλά εν γένει η επαναληπτική κατάτμηση της μεγάλης κλάσης κάνει τοπική βελτιστοποίηση και δεν γνωρίζουμε ποιά μετρική βελτιστοποιεί στο τέλος.

Παρατηρούμε πως και πάλι το άθροισμα των σταθμισμένων τομών μπορεί να γραφτεί, εντελώς ανάλογα με την περίπτωση όπου  $K=2$  ως:

$$Ncut(A_1, \dots, A_K) = \sum_{i=1}^K \frac{cut(A_i, \bar{A}_i)}{vol(A_i)} = \sum_{i=1}^K \frac{(X^i)^T \mathbf{L} X^i}{(X^i)^T \mathbf{D} X^i} \quad (3.6)$$

αν διαλέξουμε κατάλληλα τα  $a_j, b_j$ . Για να μειώσουμε την εξάρτηση μεταξύ των στηλών, μπορούμε να επιλέξουμε να εξισώσουμε κάθε επιμέρους όρο. Οπότε καταλήγουμε,

$$\frac{cut(A_j, \bar{A}_j)}{vol(A_j)} = \frac{(a_j - b_j)^2 cut(A_j, \bar{A}_j)}{vol(A_j)a_j^2 + vol(\bar{A}_j)b_j^2} \quad (3.7)$$

που μετά από πράξεις γίνεται:  $2a_j b_j (b_j - a_j) = vol(G) b_j^2$ . Επιλέγουμε να διαλέξουμε  $b_j = 0$  γιατί καταλήγει σε αυτόματη ικανοποίηση του περιορισμού, είναι απλή και δίνει αραιή αναπαράσταση. Βλέπουμε πως:  $(X_j)^T \mathbf{D} X_j = vol(A_j) a_j^2$ , οπότε μια καλή επιλογή για την απαλοιφή του  $vol(A_j)$  από την εξίσωση και την απαλοιφή των παρονομαστών είναι:  $a_j = \frac{1}{\sqrt{vol(A_j)}}$ . Παρατηρούμε λοιπόν πως δεν είναι σταθερό, αλλά εξαρτάται από την τελική κατάτμηση!. Αυτό δημιουργεί ένα φαύλο κύκλο γιατί θέλουμε να βρούμε τα στοιχεία των διανυσμάτων για να βρούμε την κατάτμηση όμως, τα ίδια τα στοιχεία εξαρτώνται από την κατάτμηση. Αυτός ο κύκλος σπάει αν καταλάβει κανείς ότι δεν μας ενδιαφέρουν τα ίδια τα στοιχεία για την κατάτμηση αλλά μόνο αν είναι διάφορα του μηδενός. Επομένως, όπως θα δούμε μετά μπορούμε να λύσουμε επακριβώς το πρόβλημα χρησιμοποιώντας μόνο το *shape* του πίνακα. Εμείς ωστόσο δεν θα περιορίσουμε τις τιμές των  $a_j$  οι οποίες θα υπολογιστούν κατά την διαδικασία της επίλυσης. Ένα άλλο σημείο που πρέπει να παρατηρήσουμε είναι ότι η κωδικοποίηση του προβλήματος δεν μοιάζει με την περίπτωση των 2 κλάσεων στο γεγονός ότι για δύο κλάσεις είχαμε μόνο ένα διάνυσμα ενώ εδώ για δύο κλάσεις θα έχουμε ένα πίνακα δύο διανυσμάτων. Βιβλιογραφικά, υπάρχει και η κωδικοποίηση του προβλήματος με  $K-1$  διανύσματα, αλλά ο πίνακας που δημιουργείται είναι αρκετά πυκνός (ειδικά όταν χαλαρώσουμε το πρόβλημα στο συνεχή χώρο) και γι' αυτό επιλέγουμε την αναπαράσταση αυτή.

Για να ολοκληρώσουμε την μοντελοποίηση του προβλήματος χρειάζεται να βρούμε τις αναγκαίες και ικανές συνθήκες για να είναι ο πίνακάς μας, πίνακας διαμέρισης. Καταρχάς, παρατηρούμε ότι αφού επιλέξαμε  $b_j = 0$ , θα πρέπει οι στήλες να είναι ορθογώνιες. Ισχύει όμως και το ανάποδο. Επειδή ο πίνακας μας έχει την ίδια τιμή σε όλα τα μη μηδενικά στοιχεία μιας στήλης ( $a_j$ ) (άρα και το ίδιο πρόσημο), το εσωτερικό γινόμενο  $(X^j)^T X^i = 0$  υπονοεί ότι κάθε επιμέρους γινόμενο  $x_k^i \cdot x_k^j$  θα πρέπει να μηδενίζεται, και επομένως το πολύ ένας από τους όρους μπορεί να είναι διάφορος του μηδενός. Θα πρέπει να προσέξουμε σαυτό το σημείο τα εξής: Πρώτον, θέλουμε να εκφράσουμε το πρόβλημα σε μορφή *Rayleigh* πηλίκων, επομένως θα ήταν πιά χρήσιμο αν γράφαμε τον παραπάνω περιορισμό ως:  $(X^i)^T \mathbf{D} X^j = 0$ , που είναι ισοδύναμος επειδή όλα τα βάρη, και κατέπεκταση οι βαθμοί, είναι θετικά και όπως αναφέραμε πριν τα επιμέρους γινόμενα πρέπει να μηδενίζονται. Δεύτερον, πρέπει να παρατηρήσει κανείς ότι οι προηγούμενες συνθήκες μας εξασφαλίζουν ότι το πολύ ένα στοιχείο κάθε γραμμής θα είναι μη μηδενικό. Δεν μας εξασφαλίζουν όμως ότι ακριβώς ένα στοιχείο θα είναι μη-μηδενικό, δηλαδή ότι κάθε κόμβος του γράφου θα ανήκει σε κάποια κλάση. Μαζί με αυτό θα πρέπει να αντιμετωπίσουμε και το γεγονός ότι καμιά κλάση δεν μπορεί να είναι κενή. Δηλαδή, από την πλευρά των πινάκων καμιά στήλη δεν μπορεί να είναι μηδενική. (Για τον τελευταίο λόγο η καθιερωμένη συνθήκη  $\mathbf{X} \mathbf{1}_K = \mathbf{1}_N$ , που απαντιέται στην βιβλιογραφία δεν είναι σωστή). Μπορούμε όμως, να παρατηρήσουμε τα εξής: πρέπει  $a_j \neq 0$  και επίσης,  $1/a_1 \cdot X^1 + 1/a_2 \cdot X^2 + \dots + 1/a_K \cdot X^K = \mathbf{1}_N$ . Αυτή η σχέση ισχύει γιατί κάθε γραμμή θα έχει μόνο ένα μη μηδενικό στοιχείο και επομένως θα αθροίζει μόνο έναν όρο κάθε φορά. Ανάποδα, αν όλη η γραμμή ήταν μηδενική τότε αυτό το άθροισμα θα έδινε ένα μηδενικό στην αντιστοιχη θέση και όχι το μοναδιαίο διάνυσμα. Τέλος μας μένει να περιορίσουμε τον πίνακα έτσι ώστε κάθε κλάση να είναι μη-κενή. Παρατηρώντας τώρα πως κάθε γραμμή του πίνακα έχει ακριβώς ένα μη-μηδενικό στοιχείο και ότι όλες οι κλάσεις είναι ορθογώνιες μεταξύ τους,  $\mathbf{X}^T \mathbf{X} = \text{diag}(n_1 a_1^2, \dots, n_K a_K^2)$ , μια ισοδύναμη συνθήκη άρα για να είναι οι κλάσεις μη κενές είναι ο πίνακας αυτός να είναι αντιστρέψιμος. Ανάποδα, αν ο πίνακας αυτός είναι αντιστρέψιμος και ισχύουν οι υπόλοιπες συνθήκες, τότε οι στήλες είναι μη-μηδενικές. Επειδή, δεν ξέρουμε τα  $a_i$  για να επιβάλλουμε ρητά την προηγούμενη συνθήκη, θα ενσωματώσουμε τις δύο τελευταίες συνθήκες λέγοντας ότι:  $\mathbf{X}^T \mathbf{1}_N = (n_1 a_1, \dots, n_K a_K)^T$  και επομένως:

$$(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{1}_N = \left( \frac{1}{a_1}, \dots, \frac{1}{a_K} \right)^T \iff \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{1}_N = \mathbf{1}_N$$

Να σημειωθεί ότι επειδή οι στήλες του  $X$  είναι γραμμικά ανεξάρτητες,  $(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T$  είναι ο ψευδοαντίστροφος του πίνακα  $\mathbf{X}$ , και επομένως, η συνθήκη γράφεται ως:

$$\mathbf{X} \mathbf{X}^+ \mathbf{1}_N = \mathbf{1}_N$$

Επειδή όμως ο  $\mathbf{X} \mathbf{X}^+$  είναι η ορθογώνια προβολή του  $\mathbf{X}$ , πάνω στο χώρο στηλών του  $\mathbf{X}$ , η τελευταία συνθήκη είναι ισοδύναμη με την συνθήκη ότι το  $\mathbf{1}_N$  ανήκει στο χώρο στηλών του

**X.** Αν ορίσουμε

$$\mathcal{X} = \{[X^1, \dots, X^K] | X^j = a_j(x_1^j, \dots, x_N^j), x_i^j \in \{1, 0\}, a_j \in \mathbb{R}, X^j \neq 0\}$$

( $X^j \neq 0 \rightarrow a_j \neq 0$ ), τότε το σύνολο των πινάκων που αναπαριστούν διαμέριση των κορυφών σε  $K$  υποσύνολα είναι:

$$\mathcal{K} = \{\mathbf{X} = [X^1 \dots X^K] | \mathbf{X} \in \mathcal{X}, (X^i)^\top \mathbf{D} X^j = 0, \forall i, j \in \{1, \dots, K\}, i \neq j, \mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{1}_N = \mathbf{1}_N\}$$

(Επειδή όπως εξηγήσαμε οι στήλες είναι μη μηδενικές και ορθογώνιες ο αντίστροφος έχει νόημα). Εκμεταλλευόμενοι και την *scale invariant* αναπαράσταση μπορούμε να πούμε ότι οι λύσεις μας είναι  $K$ -τούπλες από σημεία στο  $\mathbb{R}\mathbb{P}^{N-1}$ . Άρα το σύνολο των λύσεών μας είναι στην πραγματικότητα:

$$\mathbb{P}(\mathcal{K}) = \{[\mathbb{P}(X^1), \dots, \mathbb{P}(X^K)] | [X^1 \dots X^K] \in \mathcal{K}\}$$

Το πρόβλημα βελτιστοποίησης που πρέπει να λύσουμε λοιπόν είναι:

$$\begin{aligned} \min_{\mathbf{X}} \quad & \sum_{i=1}^K \frac{(X^i)^\top \mathbf{L} X^i}{(X^i)^\top \mathbf{D} X^i} \\ \text{s.t.} \quad & (X^i)^\top \mathbf{D} X^j = 0, \quad \forall i, j \in \{1, \dots, K\}, i \neq j \\ & \mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{1}_N = \mathbf{1}_N, \quad \mathbf{X} \in \mathcal{X} \end{aligned} \quad (3.8)$$

### 3.3.2 Επίλυση για $K > 2$ κλάσεις

Επειδή όπως είπαμε και προηγουμένως το πρόβλημα είναι εκθετικής πολυπλοκότητας θα προσπαθήσουμε να βρούμε έναν προσεγγιστικό αλγόριθμο λύνοντας το πρόβλημα σε ένα συνεχή χώρο και μετά διακριτοποιώντας το. Εύκολα παρατηρεί κανείς ότι:

$$\text{tr}(\mathbf{X}^\top \mathbf{A} \mathbf{X}) = \sum_{j=1}^K (X^j)^\top \mathbf{A} X^j$$

Επιπλέον, κάθε ορθογώνιος πίνακας που πολλαπλασιάζει από δεξιά τον πίνακά μας δεν αλλάζει το παραπάνω αποτέλεσμα:  $\Delta \delta$ ,

$$\text{tr}(\mathbf{X}^\top \mathbf{A} \mathbf{X}) = \text{tr}(\mathbf{R}^\top \mathbf{X}^\top \mathbf{A} \mathbf{X} \mathbf{R})$$

Τέλος, το *scale invariance* του προβλήματος μας επιτρέπει να πολλαπλασιάσουμε κάθε στήλη με ένα αριθμό και να αφήσουμε παρόλαυτά αναλλοίωτο το αποτέλεσμα. Αν δηλαδή κάνουμε την αλλαγή,  $X^i \leftarrow X^i / \sqrt{(X^i)^\top \mathbf{D} X^i}$ , τα *Rayleigh* πηλίκα  $\frac{(X^i)^\top \mathbf{L} X^i}{(X^i)^\top \mathbf{D} X^i}$  μένουν αναλλοίωτα, οπότε:

$$\begin{aligned} \text{NCut}([X^1, \dots, X^K]) &= \text{NCut}([X^1 / \sqrt{(X^1)^\top \mathbf{D} X^1}, \dots, X^K / \sqrt{(X^K)^\top \mathbf{D} X^K}]) = \\ & \sum_{j=1}^K ((X^j)^\top \mathbf{D} X^j)^{-1/2} (X^j)^\top \mathbf{L} ((X^j)^\top \mathbf{D} X^j)^{-1/2} X^j = \\ & \text{tr}(\mathbf{\Lambda}^{-1/2} \mathbf{X}^\top \mathbf{L} \mathbf{X} \mathbf{\Lambda}^{-1/2}) = \text{tr}(\mathbf{\Lambda}^{-1} \mathbf{X}^\top \mathbf{L} \mathbf{X}) \end{aligned}$$

όπου

$$\mathbf{\Lambda} = \text{diag}((X^1)^\top \mathbf{D} X^1, \dots, (X^K)^\top \mathbf{D} X^K) = \mathbf{X}^\top \mathbf{D} \mathbf{X}$$

γιατί περιορίσαμε τις στήλες πριν να είναι  $D$ -ορθογώνιες. Αν

$$\mathbf{\Lambda} = c^2 \mathbf{I}_K$$

που μπορεί να γίνει επιλέγοντας την τιμή των  $a_j = \frac{1}{\sqrt{\text{vol}(A_j)}}$  όπως είδαμε πριν, τότε

$$NCut(\mathbf{X}) = \text{tr}(\mathbf{\Lambda}^{-1} \mathbf{X}^\top \mathbf{L} \mathbf{X}) = \frac{1}{c^2} \text{tr}(\mathbf{X}^\top \mathbf{L} \mathbf{X}) = NCut(\mathbf{X} \mathbf{R})$$

Εδώ πρέπει να παρατηρήσουμε ότι όλοι οι περιορισμοί είναι επίσης αναλλοίωτοι στις από δεξιά περιστροφές.

**Σημείωση:** Ο πολλαπλασιασμός του πίνακα διαμέρισης από δεξιά με ένα πίνακα περιστροφής είναι σαν να περιστρέφουμε τους άξονες πάνω στους οποίους κατηγοριοποιούμε τα δεδομένα· τους άξονες των κλάσεων. Γιατί μπορούμε να το κάνουμε αυτό; Μια ερμηνεία είναι ότι από την στιγμή που έχουμε βρει τον χώρο  $K$ -διαστάσεων, έχουμε το δικαίωμα να τον καλύψουμε με οποιαδήποτε  $K$  διανύσματα όχι μόνο με τα κάθετα ιδιοδιανύσματα του Λαπλάσιαν. Μάλιστα, ο υπόχωρος που καλύπτουν αυτά τα ιδιοδιανύσματα είναι πιό σταθερός σε αριθμητικές ανακρίβειες από τον υπολογισμό των ίδιων των ιδιοδιανυσμάτων. Αυτό μας δίνει την ελευθερία, όταν το πρόβλημα είναι χαλαρωμένο σε όλο το  $\mathbb{R}$ , και βρούμε μια λύση που δεν δίνει καλή διακριτή αναπαράσταση, να βρούμε μια άλλη ισοδύναμη λύση, η οποία δίνει καλύτερη αναπαράσταση όταν διακριτοποιηθεί, απλά περιστρέφοντας τους άξονες των κλάσεων (ή καθρεπτίζοντας κάποιους) για να πέσουν περισσότερα σημεία πάνω ή κοντά τους. Να σημειώσουμε ότι στον διακριτό χώρο, μόλις κάνουμε την διαμέριση, όλα τα σημεία πέφτουν πάνω στον άξονα κάποιας κλάσης. (Επίσης, από τους περιορισμούς οι άξονες είναι κάθετοι μεταξύ τους). Όταν είμαστε στον διακριτό χώρο οι μόνες λύσεις που είναι ισοδύναμες και ικανοποιούν όλους τους περιορισμούς είναι όταν κάποιοι άξονες ανταλαχθούν μεταξύ τους, δηλαδή όταν αλλάζουμε τα ονόματα των κλάσεων.

Τελικά, το πρόβλημα που καλούμαστε να λύσουμε είναι ισοδύναμο με το:

$$\begin{aligned} \min_{\mathbf{X}} \quad & \text{tr}(\mathbf{X}^\top \mathbf{L} \mathbf{X}) \\ \text{s.t.} \quad & \mathbf{X}^\top \mathbf{D} \mathbf{X} = \mathbf{I}, \quad \mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{1}_N = \mathbf{1}_N, \quad \mathbf{X} \in \mathcal{X} \end{aligned} \tag{3.9}$$

Ξανατονίζουμε ότι η ισοδυναμία των προβλημάτων έγκειται στο ότι για κάθε λύση του πρώτου  $(X^1, \dots, X^K)$  δίνει μια λύση για το δεύτερο αν σταθμίσουμε κάθε στήλη με το  $((X^i)^\top \mathbf{D} X^i)^{-1/2}$ . Από την άλλη, για κάθε λύση του ισοδύναμου προβλήματος μπορούμε να βρούμε μια οικογένεια λύσεων του πρώτου αν πολλαπλασιάσουμε κάθε στήλη με έναν (ίσως διαφορετικό αριθμό), δηλ αν  $(Z^1, \dots, Z^K)$  μια λύση του δεύτερου τότε όλες οι λύσεις της μορφής  $(\lambda_1 Z^1, \dots, \lambda_K Z^K)$  είναι λύσεις του πρώτου προβλήματος για κάθε επιλογή των μη-μηδενικών συντελεστών. Και αυτό είναι μια ελευθερία που μας επιτρέπει να προσδιορίσουμε την καλύτερη κατάτμηση από την διακριτοποίηση μιας μεγάλης οικογένειας

λύσεων του συνεχούς χώρου.

**Σημείωση:** Το φυσικό νόημα είναι ότι τα δύο προβλήματα έχουν ισοδύναμες λύσεις στον προβολικό χώρο που τα σημεία μια ακτίνας θεωρούνται ίδια και εκφρασμένα από τις ομογενείς συντεταγμένες. Απλά το ισοδύναμο πρόβλημα λυνεται στο προβολικό επίπεδο που διώχνει τους παρονομαστές. Μετά όμως κερδίζουμε πίσω την εκφραστικότητα του να επιλέξουμε κλίμακα διαφορετική για κάθε άξονα κλάσεων. Στον διακριτό χώρο, που όλα τα σημεία είναι πάνω σε άξονες δεν θα αλλάξει το αποτέλεσμα της κατάτμησης, στον συνεχή χώρο όμως το να συμπιέζουμε ή να μεγενθύνουμε ή ακόμη και να αλλάζουμε πρόσημο σε κάποιες κατευθύνσεις-άξονες κλάσεων θα μας δώσει καλύτερη προσέγγιση της βέλτιστης διακριτής λύσης.

Μέχρι τώρα η ανάλυση που κάναμε ακουμπούσε και τους δύο χώρους (με εξαίρεση το *rotation invariance* που στον διακριτό χώρο αναφέρεται μόνο σε συγκεκριμένες περιστροφές που εναλλάσσουν τις στήλες του πίνακα διαμέρισης). Τώρα όμως η δυσκολία επίλυσης του *combinatorial* προβλήματος θα μας οδηγήσει στο συνεχή χώρο κρατώντας μόνο την απαίτηση κάθε στήλη να είναι μη-κενή. Με το τελευταίο εξασφαλίζουμε ότι ο πίνακας  $\mathbf{X}^T \mathbf{X}$  θα είναι αντιστρέψιμος γιατί έχει βαθμό  $K$  αφού οι στήλες του είναι μη-κενές και  $\mathbf{D}$ -ορθογώνιες και άρα, γραμμικά ανεξάρτητες. Επίσης, υπενθυμίζουμε από πριν ότι οι ιδιότητες του αναλλοίωτου της περιστροφής και του αναλλοίωτου της κλίμακας που μοντελοποιήσαμε το πρόβλημά μας μας επιτρέπουν, αν βρούμε μια λύση  $\mathbf{X}$  στο ισοδύναμο πρόβλημα, να μπορέσουμε να προσδιορίσουμε της βέλτιστη διακριτή λύση από μια οικογένεια ισοδύναμων λύσεων της μορφής  $\mathbf{XRA}$ , όπου ο  $\mathbf{R}$  είναι ορθογώνιος και ο  $\mathbf{A}$  διαγώνιος.

Θυμίζουμε από διαφορική γεωμετρία ότι το *Stiefel manifold*  $St(k, n)$ , περιέχει στοιχεία με  $k$ -tuples από ορθομοναδιαία διανύσματα του  $\mathbb{R}^n$ . Άρα μπορεί να το δει κανείς και σαν το σύνολο όλων των  $n \times k$  ορθογώνιων πινάκων. Επιπλέον, το *Grassmannian manifold*  $G(k, n)$  αποτελείται από όλους τους γραμμικούς  $k$ -διάστατους υποχώρους του  $\mathbb{R}^n$ . Ο δεξιός πολλαπλασιασμός ενός στοιχείου του  $St(K, N)$  με ένα πίνακα που ανήκει στο  $\mathbf{O}(K)$ , δημιουργεί ένα *manifold*, ισομορφικό με το  $G(K, N)$ . Εμάς οι πίνακες όμως δεν είναι ορθογώνιοι αλλά  $D$ -ορθογώνιοι. Αυτό σημαίνει ότι δεν ανήκουν στο *Stiefel manifold* αλλά αποτελούν την εικόνα του *Stiefel manifold* κάτω από τη γραμμική απεικόνιση  $\mathcal{D}(X) = D^{1/2}X$ . Και πάλι το *coset manifold* που δημιουργείται από την δεξιά δράση του *group*  $\mathbf{O}(K)$  είναι ισομορφικό με το  $G(K, N)$ . Σκεπτόμενοι ότι οι λύσεις μας λοιπόν ανήκουν σε αυτό το *manifold* θα μπορούσαμε να ορίσουμε τις αποστάσεις των σημείων με κατάλληλο τρόπο ώστε να ελαχιστοποιήσουμε μια μετρική που θα μας δώσει καλύτερη προσέγγιση στην βέλτιστη διακριτή λύση([13]).

Κάνοντας αλλαγή μεταβλητής στο πρόβλημά μας:  $\mathbf{Y} = \mathbf{D}^{1/2}\mathbf{X}$ , οι περιορισμοί γίνονται:

$$1. \mathbf{X}^T \mathbf{D} \mathbf{X} = \mathbf{I} \iff \mathbf{Y}^T \mathbf{Y} = \mathbf{I} \iff \mathbf{Y}^T = \mathbf{Y}^+.$$

$$2. \mathbf{X}^\top \mathbf{LX} = \mathbf{Y}^\top \mathbf{L}_{sym} \mathbf{Y}.$$

$$3. \mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{1}_N = \mathbf{X} \mathbf{X}^+ \mathbf{1}_N = \mathbf{1}_N \iff \mathbf{1}_N \in \text{range}(\mathbf{X}) \iff \mathbf{D}^{1/2} \mathbf{1}_N \in \text{range}(\mathbf{Y}) \iff \mathbf{Y} \mathbf{Y}^+ \mathbf{D}^{1/2} \mathbf{1}_N = \mathbf{D}^{1/2} \mathbf{1}_N \iff \mathbf{Y} \mathbf{Y}^\top \mathbf{D}^{1/2} \mathbf{1}_N = \mathbf{D}^{1/2} \mathbf{1}_N.$$

Επομένως, το πρόβλημά μας γίνεται:

$$\begin{aligned} \min_{\mathbf{Y}} \quad & tr(\mathbf{Y}^\top \mathbf{L}_{sym} \mathbf{Y}) \\ \text{s.t.} \quad & \mathbf{Y}^\top \mathbf{Y} = \mathbf{I}, \quad \mathbf{Y} \mathbf{Y}^\top \mathbf{D}^{1/2} \mathbf{1}_N = \mathbf{D}^{1/2} \mathbf{1}_N. \end{aligned} \quad (3.10)$$

Γνωρίζουμε από το *Poincare separation theorem* ότι το ελάχιστο του  $tr(\mathbf{Y}^\top \mathbf{L}_{sym} \mathbf{Y})$  πάνω σε όλους τους  $N \times K$  πίνακες  $\mathbf{Y}$  που ικανοποιούν τη σχέση  $\mathbf{Y}^\top \mathbf{Y} = \mathbf{I}$ , είναι κάτω φραγμένο από το άθροισμα των  $K$ -μικρότερων ιδιοτιμών του  $\mathbf{L}_{sym}$ . Επιπλέον, αγνοώντας προς το παρόν το τελευταίο περιορισμό, μπορούμε εύκολα να δούμε ότι η τιμή του κάτω άκρου του ελαχίστου λαμβάνεται όταν επιλέξουμε για λύση τα  $K$  ιδιοδιανύσματα του  $\mathbf{L}_{sym}$  που αντιστοιχούν στις μικρότερες ιδιοτιμές.  $tr(\mathbf{U}^\top \mathbf{L}_{sym} \mathbf{U}) = tr(\mathbf{U}^\top \mathbf{U} \mathbf{\Lambda}) = tr(\mathbf{\Lambda}) = \sum_{i=1}^K \lambda_i$ . Επειδή ο πίνακας είναι συμμετρικός αυτά θα είναι κάθετα μεταξύ τους, και μπορούμε να τα επιλέξουμε και μοναδιαία. Τι συμβαίνει όταν π.χ. η  $K$ -οστή και η  $(K+1)$ -οστή ιδιοτιμές είναι ίσες; Το γεγονός ότι τα ιδιοδιανύσματα είναι κάθετα θα μπορούσε να δώσει αρκετά διαφορετικές λύσεις. Στην πράξη, αν ο γράφος δεν έχει πολλές συμμετρίες αυτό δεν συμβαίνει, αλλά σε περίπτωση που τύχει θα πρέπει να ελέγξουμε όλες τις πιθανές λύσεις. Η λογική λέει ότι αν ο γράφος έχει τόσες συμμετρίες ώστε να έχει ίσες ιδιοτιμές, λογικά θα έχει και πολλές ισοδύναμα ερμηνεύσιμες κατατμήσεις. Παραδείγματος χάρη η κλίκα έχει πολλές ιδιοτιμές ίσες, αλλά και όλες οι κατατμήσεις που προκύπτουν από αυτή είναι όμοιες αν περιστρέψουμε το γράφο.

Ας κοιτάξουμε τώρα τον τελευταίο περιορισμό. Αν θεωρήσουμε ότι ο γράφος είναι συνεκτικός και άρα  $\lambda_2 > 0$ , από τα θεωρήματα του προηγούμενου κεφαλαίου ξέρουμε πως το πρώτο μοναδιαίο ιδιοδιάνυσμα του  $\mathbf{L}_{sym}$ , είναι το  $\mathbf{Y}^1 = \frac{\mathbf{D}^{1/2} \mathbf{1}_N}{\|\mathbf{D}^{1/2} \mathbf{1}_N\|}$ . Αυτό σημαίνει ότι ανήκει στο  $\text{range}(\mathbf{Y})$  το  $\mathbf{D}^{1/2} \mathbf{1}_N$ .

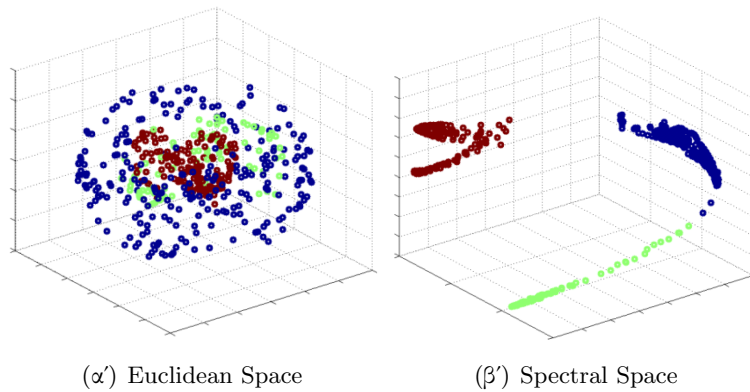
Τελικά από το ισοδύναμο πρόβλημα μπορούμε να μεταφερθούμε στο αρχικό ξέροντας ότι η  $\mathbf{Z} = \mathbf{D}^{-1/2} \mathbf{Y} = \mathbf{D}^{-1/2} \mathbf{U}_{1:K}$ , είναι βέλτιστη λύση του. Το  $\mathbf{Z}$  είναι όπως γνωρίζουμε από τα θεωρήματα του προηγούμενου κεφαλαίου, το σύνολο των  $K$  μικρότερων ιδιοδιανυσμάτων του  $\mathbf{L}_{rw}$ . Αυτά είναι  $\mathbf{D}$ -ορθογώνια, άρα (επειδή ο  $\mathbf{D}$  είναι θετικά ορισμένος) και γραμμικά ανεξάρτητα. Επιπλέον, υπάρχουν πίνακες περιστροφής που μπορούν να κάνουν τις στήλες του  $\mathbf{Z}$  ταυτόχρονα  $\mathbf{D}$ -ορθογώνιες και ορθογώνιες. Π.χ. αν διαγωνιοποιήσουμε τον  $\mathbf{Z}^\top \mathbf{Z} = \mathbf{R} \mathbf{\Sigma} \mathbf{R}^\top \iff (\mathbf{Z} \mathbf{R})^\top (\mathbf{Z} \mathbf{R}) = \mathbf{\Sigma}$ , που σημαίνει ότι οι στήλες του  $\mathbf{Z} \mathbf{R}$  είναι ορθογώνιες. Επειδή όπως είπαμε πριν οι στήλες του  $\mathbf{Z}$ , είναι επίσης γραμμικά ανεξάρτητες, ο πίνακας  $\mathbf{Z}^\top \mathbf{Z}$  θα είναι θετικά ορισμένος και όλα τα στοιχεία του  $\mathbf{\Sigma}$  θα είναι θετικά. Να σημειωθεί ότι μπορεί ο  $\mathbf{R}$  να υπολογιστεί με καλύτερη αριθμητική ευστάθεια από την *SVD* του  $\mathbf{Z}$ . Αυτό που απομένει τώρα είναι να βρει κάποιος κατάλληλους πίνακες  $\mathbf{R}, \mathbf{\Lambda}$  που θα δώσουν λύση κοντά, με κάποια έννοια απόστασης στην βέλτιστη διακριτή λύση.

### 3.3.3 Διαφορετικά *Rounding Scemes*

Στην παρούσα φάση θα θέλαμε να βρούμε ένα τρόπο να περάσουμε από τις λύσεις του συνεχούς χώρου σε διακριτές λύσεις. Μπορούμε να παρατηρήσουμε και πάλι ότι αφού τα ιδιοδιανύσματα του  $\mathbf{L}_{rw}$  είναι κάθετα στο  $1_N$  που είναι το πρώτο, θα έχουν και θετικές και αρνητικές τιμές. Αυτό μπορεί να μας βοήθησε όταν είχαμε να χωρίσουμε τον γράφο σε δύο κλάσεις αλλά τώρα η πληροφορία αυτή δεν είναι αρκετή.

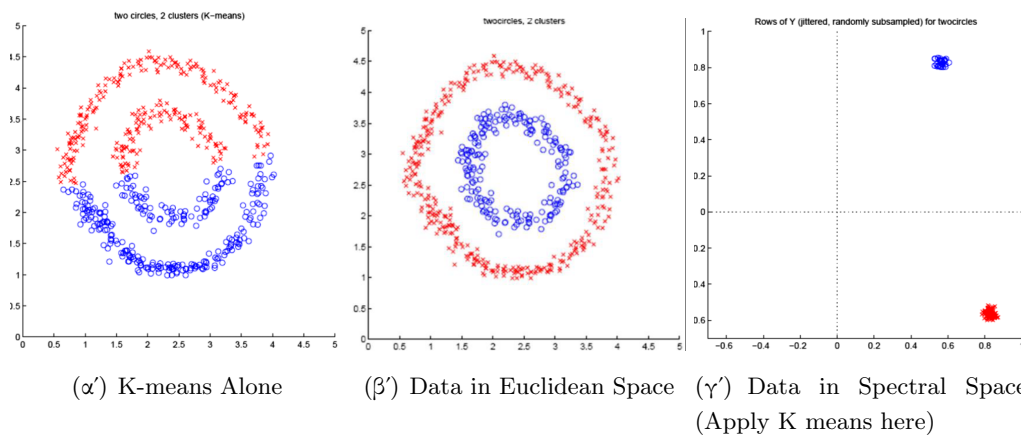
#### Πρώτη προσέγγιση: Χρήση του *k-means* αλγορίθμου σαν *rounding scheme*

(Andrew Ng [29]) Αφορμή για την ιδέα αυτή αποτέλεσε η ιδανική περίπτωση ενός γράφου, διαχωρισμένου εξαρχής σε  $K$  συνεκτικές συνιστώσες. Όπως είδαμε τα  $K$  πρώτα ιδιοδιανύσματα ενός τέτοιου γράφου είναι δείκτριες συναρτήσεις των  $K$  κλάσεων. Επομένως, αν απεικονίζαμε κάθε γραμμή του πίνακα στο χώρο των  $K$  ιδιοδιανυσμάτων, με την έννοια ότι κάθε κόμβος τώρα έχει  $K$  συνιστώσες, μια από κάθε ιδιοδιάνυσμα, τότε τα σημεία που ανήκαν στην ίδια συνεκτική συνιστώσα θα συνέπεφταν μαζί. Τα στοιχεία που ανήκαν σε άλλη συνεκτική συνιστώσα θα συνέπεφταν επίσης μαζί πάνω σε έναν άλλο άξονα κάθετο στον πρώτο. Επομένως, οι ευκλείδειες αποστάσεις των σημείων θα ήταν αρκετά καλή μετρική για τον πλήρη διαχωρισμό των σημείων. Τώρα, όταν φύγουμε από την ιδανική περίπτωση και ο γράφος μας είναι πιά συνεκτικός αλλά με καλά διαχωρισμένες κλάσεις, εξαιτίας της συνέχειας των ιδιοδιανυσμάτων του πίνακα  $\mathbf{L}_{sym}$  περιμένουμε ότι τα σημεία απεικονισμένα σε αυτόν τον χώρο ιδιοδιανυσμάτων δεν θα απέχουν πολύ μεταξύ τους. Αν επιλέξουμε λοιπόν όλα να τα ομαδοποιήσουμε σε  $K$  κλάσεις που έχουν  $K$  κέντρα αντιπροσώπους τότε ο *k-means*, θα δώσει αρκετά καλά αποτελέσματα. Να σημειωθεί ότι παρότι ο *k-means* θα δώσει μόνο γραμμικές περιοχές απόφασης, επειδή είμαστε σε ένα εξαιρετικά μη γραμμικό χώρο ακόμη και αυτές οι περιοχές μπορεί να έχουν μη γραμμικά όρια στον ευκλείδειο χώρο. Μπορεί κανείς να σκεφτεί αυτή την διαδικασία της επίλυσης στον συνεχή χώρο, ως μια μη-γραμμική απεικόνιση από τον ευκλείδειο χώρο σε ένα χώρο ιδιοδιανυσμάτων όπου οι αποστάσεις των σημείων είναι πραγματικά ευκλείδειες και επομένως η χρήση των κλασικών αλγορίθμων είναι ικανοποιητική. Στα παρακάτω σχήματα (3.6,3.7) φαίνεται πως λειτουργεί το *mapping* των κόμβων στο *spectral space* και κατόπιν πως ο αλγόριθμος *k-κέντρων* κάνει την κατάτμηση. Να σημειωθεί ότι συχνά μια ακόμη ευρετική που χρησιμοποιείται είναι να σταθμίζουμε όλες τις γραμμές του πίνακα κατάτμησης, που σημαίνει ότι στο χώρο των ιδιοδιανυσμάτων όλα τα σημεία θα βρίσκονται πάνω σε μια *hypersphere*. Επίσης, αφού γνωρίζουμε ότι αν έχουμε πολύ καλά διαχωρισμένες κλάσεις, αυτές θα βρίσκονται σε περίπου 90 μοίρες απόσταση πάνω στην σφαίρα



Σχήμα 3.6: Embedding των σημείων σε ένα χώρο τριών διαστάσεων.

Συχνά εφαρμογές χρησιμοποιούν τη λύση του *Normalized Cut* στο συνεχή χώρο για να κάνουν *dimensionality reduction* σε διαστάσεις όσες τα ιδιοδιανύσματα ([5]). Μπορεί κανείς να σκέφτεται γενικά το *Spectral Clustering* σαν μια προσπάθεια να διαχωρίσουμε τα δεδομένα προβάλλοντάς τα ταυτόχρονα σε ένα χώρο χαμηλών διαστάσεων. Γιαυτό το λόγο υπάρχουν αντιστοιχίσεις με μεθόδους *Kernel PCA* ([2]), *Kernel Kmeans* ([34]), καθώς και με την *LDA* ([41]). Μια ακόμη χρήση της λύσης στο συνεχή χώρο είναι ότι μας δίνει ένα *confidence score* για το κάθε σημείο ως προς το σε ποιά κλάση ανήκει.



Σχήμα 3.7: Προβολή σημείων σε ένα χώρο ιδιοδιανυσμάτων και κατάτμησή τους με χρήση *k - means*

### Μια καλύτερη προσέγγιση: Μοντελοποίηση του προβλήματος διακριτοποίησης σαν ένα πρόβλημα βελτιστοποίησης

(Yu and Shi [49]). Μια γενικότερη μέθοδος που συνήθως δίνει καλύτερα αποτελέσματα στην πράξη, είναι να μοντελοποιήσουμε το πρόβλημα διακριτοποίησης σαν ένα πρόβλημα ελαχιστοποίησης της απόστασης μεταξύ των λύσεών μας, που αντιστοιχούν σε σημεία  $\mathbf{Z} = (Z^1, \dots, Z^K) \in G(K, N)$  και των διακριτών λύσεων που αντιστοιχούν σε σημεία  $\mathbb{P}(X) = (\mathbb{P}(X^1), \dots, \mathbb{P}(X^K)) \in \mathbb{P}(K)$ . Μπορούμε να πούμε ότι το  $\mathcal{K}$  είναι ένα υποσύνολο



του  $G(K, N)$  αν σκεφτούμε το χώρο που καλύπτουν τα διανύσματα  $(X^1, \dots, X^K) \in \mathcal{K}$ , ωστόσο πρέπει να θυμόμαστε ότι το σύνολο αυτό είναι κλειστό μόνο κάτω από πολύ συγκεκριμένους ορθογώνιους μετασχηματισμούς του  $\mathbf{O}(K)$ , αυτούς που αντιμεταθέτουν τις στήλες ενός πίνακα). Αν επιλέξουμε να δούμε έτσι το  $\mathcal{K}$ , τότε για να ελαχιστοποιήσουμε την *Riemannian* απόσταση  $d(X, Z)$  ένα λογικό μέτρο απόστασης θα ήταν το  $\text{tr}(\mathbf{X}^\top \mathbf{Y})$ . Η επίλυση του προβλήματος αυτού δεν είναι καθόλου μελετημένη στην βιβλιογραφία. Αντάυτου θα επιλέξουμε να δούμε το  $\mathcal{K}$  σαν ένα υποσύνολο του  $(\mathbb{R}\mathbb{P}^{N-1})^K$ . Όπως αναλύσαμε στην περίπτωση των 2 κλάσεων, η ελαχιστοποίηση της απόστασης 2 σημείων στο προβολικό χώρο  $N-1$  διαστάσεων είναι ισοδύναμη με την ελαχιστοποίηση της ευκλείδειας νόρμας των διανυσμάτων των ομογενών συντεταγμένων τους, αρκεί αυτά να είναι πάνω στην μοναδιαία σφαίρα. Οπότε, αν χρησιμοποιήσουμε το *product distance*, στο  $(\mathbb{R}\mathbb{P}^{N-1})^K$  που δίνεται ως:

$$d((\mathbb{P}(X^1), \dots, \mathbb{P}(X^K)), (\mathbb{P}(Z^1), \dots, \mathbb{P}(Z^K))) = \sum_{j=1}^K d(\mathbb{P}(X^j), \mathbb{P}(Z^j))$$

τότε καλούμαστε να ελαχιστοποιήσουμε την *Frobenius* απόσταση των δύο πινάκων, υπό τον περιορισμό οι στήλες να έχουν ίσα μέτρα. Επειδή δεν υπάρχουν γνωστοί αλγόριθμοι βελτιστοποίησης που να λύνουν αυτό το πρόβλημα θα πετάξουμε τον τελευταίο περιορισμό και δωσμένης μιας λύσης  $\mathbf{Z}$  του συνεχούς χώρου θα αναζητήσουμε  $(\mathbf{X}, \mathbf{Q} = \mathbf{R}\mathbf{A})$ , όπου  $\mathbf{X} \in \mathcal{X}$  ενώ ο  $\mathbf{Q}$  είναι ένας  $K \times K$  πίνακας με κάθετες μεταξύ τους στήλες (μη μηδενικές) τέτοιους ώστε:

$$\begin{aligned} \min_{(\mathbf{X}, \mathbf{R}, \mathbf{A})} \quad & \varphi(\mathbf{X}, \mathbf{R}, \mathbf{A}) = \|\mathbf{X} - \mathbf{Z}\mathbf{R}\mathbf{A}\|_F \\ \text{s.t.} \quad & \mathbf{R} \in \mathbf{O}(K), \quad \mathbf{X} \in \mathcal{K}, \quad \mathbf{A} = \text{diag}(\lambda_1, \dots, \lambda_K), \lambda_i > 0 \\ & \|\mathbf{X}\|_F = \|\mathbf{Z}\|_F \end{aligned} \quad (3.11)$$

Εξαιτίας του τελευταίου περιορισμού η μετρική γίνεται:

$$\varphi(\mathbf{X}, \mathbf{R}, \mathbf{A}) = \|\mathbf{X} - \mathbf{Z}\mathbf{R}\mathbf{A}\|_F = \text{tr}((\mathbf{X} - \mathbf{Z}\mathbf{R}\mathbf{A})^\top (\mathbf{X} - \mathbf{Z}\mathbf{R}\mathbf{A})) \quad (3.12)$$

$$= \|\mathbf{Z}\|_F^2 - 2\text{tr}((\mathbf{R}\mathbf{A})^\top \mathbf{Z}^\top \mathbf{X}) + \text{tr}(\mathbf{Z}^\top \mathbf{Z}\mathbf{R}\mathbf{A}(\mathbf{R}\mathbf{A})^\top) \quad (3.13)$$

Η δυσκολία του προβλήματος έγγειται στο ότι είναι μη-γραμμικό με αγνώστους και τους τρεις πίνακες. Γιαυτό θα εφαρμόσουμε μια επαναληπτική μέθοδο που θα κρατάει κάθε φορά δύο από τους τρεις πίνακες σταθερούς. Δηλαδή,

1. minimize  $\varphi_1(\mathbf{X}) = \|\mathbf{X} - \mathbf{Z}\mathbf{R}\mathbf{A}\|_F$ , κρατώντας τους  $\mathbf{R}, \mathbf{A}$  σταθερούς (Εστω ότι είναι αρχικοποιημένοι στον μοναδιαίο), αλλά υπό τον περιορισμό  $\|\mathbf{X}\|_F = \|\mathbf{Z}\|_F$ .
2. minimize  $\varphi_2(\mathbf{R}) = \|\mathbf{X} - \mathbf{Z}\mathbf{R}\|_F$ , χρησιμοποιώντας το προηγούμενο  $\mathbf{X}$ .
3. minimize  $\varphi_3(\mathbf{A}) = \|\mathbf{X} - \mathbf{Z}\mathbf{R}\mathbf{A}\|_F$ , χρησιμοποιώντας τις πιο πρόσφατες τιμές για κάθε πίνακα.

Αν θέλουμε να είμαστε πολύ αυστηροί, ο πίνακας  $\mathbf{R}\mathbf{A}$  δεν θα είναι ελαχιστοποιητής αυτής της μετρικής αλλά σίγουρα αποτελεί βελτίωση σε σχέση με το να είχαμε μόνο του

το  $\mathbf{R}$ . (Η βελτίωση αυτή οφείλεται στο *J.Gallier* ([13]), που έκανε μια *scale invariant* αναπαράσταση του πίνακα διαμέρισης. Πώς λύνουμε όμως κάθε βήμα ξεχωριστά;

- Το πρώτο βήμα από την εξίσωση 3.12 είναι ισοδύναμο με την μεγιστοποίηση του  $tr((\mathbf{Z}\mathbf{R}\mathbf{\Lambda})^\top \mathbf{X}) = tr(\mathbf{X}(\mathbf{Z}\mathbf{R}\mathbf{\Lambda})^\top)$ , αφού όλες οι άλλες παράμετροι είναι σταθερές. Ο περιορισμός που έχει ο πίνακας  $\mathbf{X}$ , μας επιτρέπει να σκεφτούμε ότι αρκεί να βρούμε μόνο ποιές τιμές του πίνακα είναι μη-μηδενικές σε πρώτη φάση (σχήμα του πίνακα) και κατόπιν να σταθμίσουμε τις στήλες κατάλληλα. Επειδή το  $\mathbf{X}$  έχει σε κάθε γραμμή μόνο ένα στοιχείο θετικό  $a_j$ , στην θέση  $j_i$  ( $1 \leq j_i \leq K$ ), θα ισχύει:  $tr(\mathbf{X}\mathbf{Y}^\top) = \sum_{i=1}^N a_{j_i} y_{ij_i}$ , όπου  $\mathbf{Y} = \mathbf{Z}\mathbf{R}\mathbf{\Lambda}$ . Προφανώς, η εξίσωση αυτή μεγιστοποιείται όταν διαλέξουμε για την  $i$ -οστή γραμμή του πίνακα την στήλη  $l$ , για την οποία το  $y_l$  είναι μέγιστο. Αν το κάνουμε αυτό βέβαια δημιουργούνται τα εξής προβλήματα: Το πρώτο είναι ότι δεν εξασφαλίζεται ότι ο τελικός πίνακας δεν θα έχει μηδενικές στήλες. Κατά μια έννοια αυτό κάνει την μέθοδο πιο *adaptive*, αφού δίνει το πολύ  $K$  κλάσεις. Αν συμβεί αυτό τότε κάνουμε μεταφορά κάποιων σημείων από την κλάση με τα περισσότερα προς μια κλάση χωρίς σημεία. Προφανώς, αυτό πλέον σημαίνει ότι δεν έχουμε την βέλτιστη λύση ως προς το άθροισμα. Το άλλο πρόβλημα είναι ότι τελικά πρέπει οι πίνακες  $\mathbf{Z}, \mathbf{X}$  να έχουν ίδια *Frobenius* νόρμα. Επειδή όμως έχουμε βρει το σχήμα κάθε πίνακα, κάνοντας στάθμιση στις στήλες του μπορούμε να το πετύχουμε αυτό.
- Το δεύτερο βήμα από την εξίσωση 3.12 και με γνώση ότι  $\mathbf{\Lambda} = \mathbf{I}$  και ότι  $\mathbf{R}\mathbf{R}^\top = \mathbf{I}$  καταλήγει και πάλι στην μεγιστοποίηση του  $tr(\mathbf{R}^\top \mathbf{Z}^\top \mathbf{X})$ , μόνο που τώρα ο άγνωστος είναι ο πίνακας  $\mathbf{R}$ , ο οποίος περιορίζεται να είναι ορθογώνιος. Η λύση αυτού του προβλήματος αποδεικνύεται ότι είναι:  $\mathbf{R}^* = \mathbf{V}\mathbf{U}^\top$ , για οποιαδήποτε *SVD* διάσπαση  $\mathbf{U}\mathbf{\Sigma}\mathbf{V}^\top = \mathbf{Z}^\top \mathbf{X}$ , του  $\mathbf{Z}^\top \mathbf{X}$ .
- Το τρίτο βήμα από την εξίσωση 3.12 γράφεται ως ( $\mathbf{Y} = \mathbf{Z}\mathbf{R}$ ):

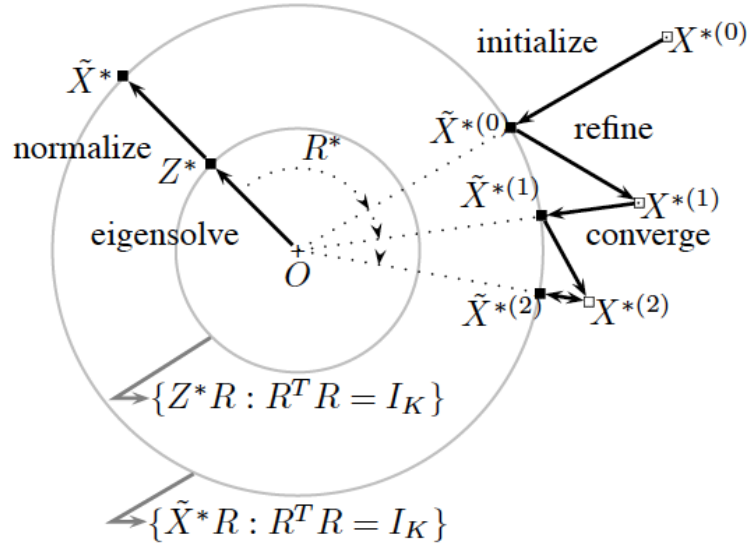
$$\begin{aligned} \|\mathbf{X} - \mathbf{Y}\mathbf{\Lambda}\|^2 &= \|\mathbf{X}\|^2 - 2tr(\mathbf{\Lambda}^\top \mathbf{Y}^\top \mathbf{X}) + tr(\mathbf{Y}^\top \mathbf{Y}\mathbf{\Lambda}\mathbf{\Lambda}^\top) \\ &= \|\mathbf{X}\|^2 - 2tr(\mathbf{Y}^\top \mathbf{X}\mathbf{\Lambda}) + tr(\mathbf{Y}^\top \mathbf{Y}\mathbf{\Lambda}^2) \\ &= \|\mathbf{X}\|^2 - \sum_{j=1}^K (\mathbf{Y}^\top \mathbf{X})_{jj} \lambda_j + \sum_{j=1}^K \|Y^j\|_2^2 \lambda_j^2 \end{aligned}$$

Πρόκειται για μια τετραγωνική παράσταση με θετικούς συντελεστές στους μεγιστοβάθμιους όρους που αν πάρουμε τις μερικές παραγώγους της μηδενίζεται στα σημεία

$$\lambda_j = \frac{((\mathbf{Z}\mathbf{R})^\top \mathbf{X})_{jj}}{\|Z^j\|_2^2}$$

που είναι και τα ελάχιστα της παράστασης. Υπάρχει θεωρητικά η πιθανότητα κάποιος από τους όρους να είναι ίσος με 0, που θα σήμαινε ότι ο πίνακας δεν είναι πια αντιστρέψιμος. Αν συμβεί αυτό προσπερνάμε το βήμα 2 γιαυτή την επανάληψη.

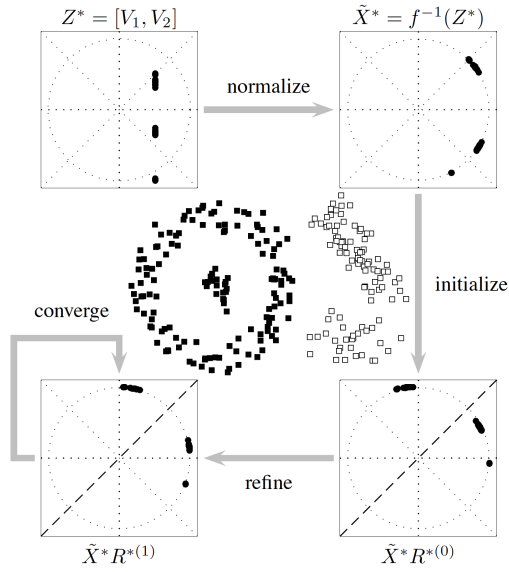
Στην πρώτη απόπειρα επίλυσης αυτού του προβλήματος [49] το τρίτο βήμα είχε παραλειφθεί και ο αλγόριθμος σχηματικά έμοιαζε ως εξής:



**Figure 1.** Schematic diagram of our algorithm. (1) We first obtain eigenvectors  $Z^*$ . Shown as the inner circle,  $Z^*$  generates the whole family of global optima through orthonormal transform  $R$ . After length normalization, each optimum corresponds to a partitioning solution in the continuous domain (the outer circle). (2) We then obtain a discrete solution closest to the continuous optima in an iterative fashion. Starting from discrete solution  $X^{*(0)}$ , we find  $\tilde{X}^{*(0)}$  by computing  $R^*$  that brings  $\tilde{X}^*$  closest to  $X^{*(0)}$ . Given the continuous optimum  $\tilde{X}^{*(0)}$ , we compute its closest discrete solution  $\tilde{X}^{*(0)}$ ; so on and so forth. The algorithm converges at the solution pair  $(X^{*(2)}, \tilde{X}^{*(2)})$ , which are the closest to each other. The optimality of  $\tilde{X}^{*(2)}$  guarantees that  $X^{*(2)}$  is nearly global-optimal.

Σχήμα 3.8: Αλγόριθμος Yu-Shi

όπου  $\mathbf{X} \in \{0, 1\}^{N \times K}$  είναι ο πίνακας κατάτμησης και  $\mathbf{Z} = \mathbf{X}(\mathbf{X}^\top \mathbf{D} \mathbf{X})^{-1/2}$  ο πίνακας του ισοδύναμου προβλήματος. Κρατάμε το σύμβολο αυτό και μετά την χαλάρωση για το συνεχή χώρο ενώ, ενώ  $\mathbf{Z}^*$  είναι οι ιδιοτιμές του  $\mathbf{L}_{rw}$ . Επειδή η αναπαράσταση δεν επιλέχθηκε εξαρχής να είναι *scale invariant*, πρέπει να επανέλθουμε στον συνεχή χώρο του αρχικού προβλήματος ως:  $\tilde{\mathbf{X}} = \text{Diag}(\text{diag}(\mathbf{Z} \mathbf{Z}^\top)^{-1/2}) \mathbf{Z}$ , που είναι μια στάθμιση των γραμμών του  $\mathbf{Z}^*$  που φέρνει τα σημεία πάνω σε μια *hypersphere*. Αν και το τελευταίο δεν είναι απαραίτητα λύση του προβλήματος είναι μια ευρετική που λειτουργεί καλά στην πράξη. Παρακάτω φαίνεται και ο τρόπος με τον οποίο γίνεται η σύγκλιση του αλγορίθμου.



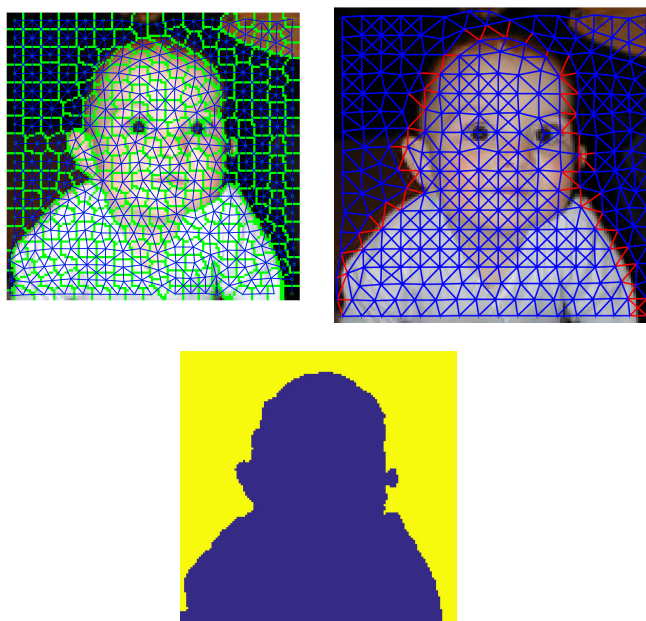
**Figure 2.** Progression of our algorithm. Each plot shows an  $N \times 2$  matrix, with each row taken as  $(x, y)$  coordinates of a point in the plane. Though  $N = 245$ , many of them are mapped to the same planar points, resulting in three visible clusters. 1) Normalize: starting with the eigenvectors  $Z^*$ , we first map it back to the  $X$ -space by normalizing their lengths so that all of them lie on the unit circle. 2) Initialize: two points with almost orthogonal phases are selected to form  $R^{*(0)}$ .  $\tilde{X}^* R^{*(0)}$  is the projection of all the points to the two chosen directions. An initial clustering  $X^{*(0)}$  is obtained by non-maximum suppression: points are divided according to the dashed line  $x = y$ : points below the line assigned to  $(1, 0)$  hence  $\mathbb{V}_1$ , those above the line assigned to  $(0, 1)$  hence  $\mathbb{V}_2$ . 3) Refine: we find the closest continuous optimal to  $X^{*(0)}$  by adjusting the rotation matrix  $R^{*(1)}$ . Non-maximum suppression produces its closest discrete solution  $X^{*(1)}$ , which is exactly the same as  $X^{*(0)}$ . The algorithm converges and stops. The final clustering is shown in the center, with  $\varepsilon(X^*) = 0.9997 < \varepsilon(\tilde{X}^*) = 0.9998$ .

Σχήμα 3.9: Αλγόριθμος *Yu, Shi*: Κατάτμηση σημείων

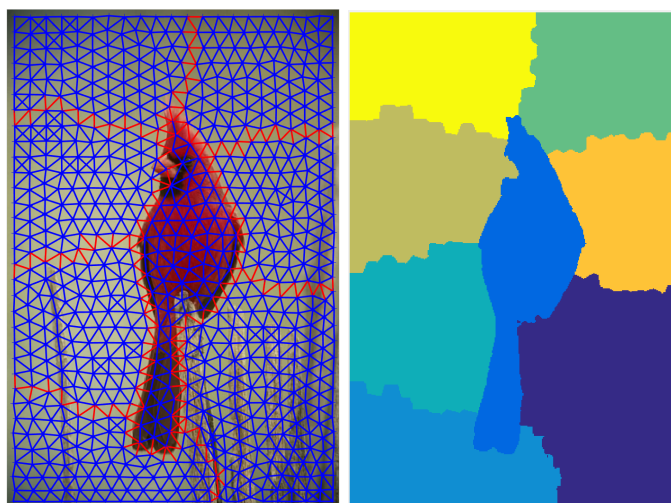
Πειραματικά, έχειδειχθεί ότι το  $\varphi$  μειώνεται από επανάληψη σε επανάληψη των αλγορίθμων και ότι η επιλογή  $\lambda = I$ , από τους *Shi, Yu* μειώνει την αποτελεσματικότητα και την ταχύτητα του αλγορίθμου ([13]).

### 3.4 Εφαρμογή και πρακτικά χαρακτηριστικά

Εφαρμόσαμε τον παραπάνω αλγόριθμο σε προβλήματα κατάτμησης εικόνων για εντοπισμό αντικειμένων και σε *point cloud* δεδομένα. Πειραματιστήκαμε με αρκετούς τρόπους κατασκευής γράφων και αρκετά χαρακτηριστικά κόμβων και αποστάσεων και αυτά που δούλεψαν καλύτερα είναι: Ο *RAG – SLIC* γράφος, που περιγράψαμε στο προηγούμενο κεφάλαιο, με τις ακμές να δημιουργούνται είτε από απλά χρωματικά χαρακτηριστικά, είτε με την μέθοδο των *intervening contours* και να πέφτει το βάρος τους εκθετικά με την απόσταση των κόμβων, με παράμετρο  $\sigma \in [30, 50]$ . Κάποια ποιοτικά αποτελέσματα φαίνονται στην συνέχεια. Για την επιτάχυνση των αλγορίθμων χρησιμοποιήθηκαν αραιοί γράφοι και *Krylov* τεχνικές στον υπολογισμό των ιδιοτιμών, ώστε να μπορούμε να λύσουμε το πρόβλημα σε μεγάλες εικόνες. Οι εικόνες είναι από το *Berkeley Dataset*.



Σχήμα 3.10: Κατάτμηση με χρήση intervening contours



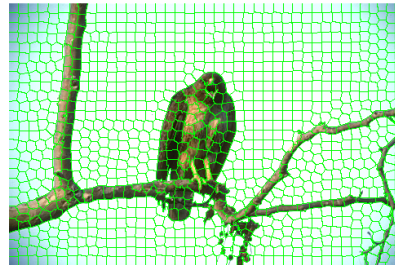
Σχήμα 3.11: Χρωματική κατάτμηση σε πολλές κλάσεις

Στην εικόνα 3.10 βλέπουμε πως ο αλγόριθμος κατάτμησης με την χρήση intervening contours καταφέρνει να κάνει κατάτμηση χρησιμοποιώντας μόλις δύο κλάσεις. Στην αντίστοιχη εικόνα που χρησιμοποιούσε μόνο χρωματική πληροφορία (2.8) η κατάτμηση του μωρού από το background ήταν εφικτή μόνο όταν απαιτούσαμε τρία cluster. Στην επόμενη εικόνα 3.11 φαίνεται πόσο σταθερή είναι η κατάτμηση όταν χρησιμοποιήσουμε πολλές κλάσεις, αφού ο αλγόριθμος προτιμάει να κατατμήσει το background από το να διασπάσει το foreground. Αυτό δεν είναι προφανές, γιατί ανεβάζοντας τον αριθμό των κλάσεων η κατάτμηση δεν γίνε-

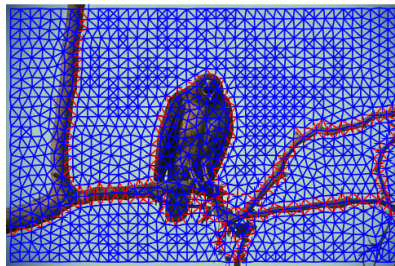
ται απαραίτητα στα προυπάρχοντα αντικείμενα που δημιούργησε ο αλγόριθμος για μικρότερο αριθμό κλάσεων, αλλά μπορεί να είναι πολύ διαφορετική.



(α') original



(β') RAG-SLIC graph



(γ') NCut



(δ') Clusters



(ε') More Clusters

Σχήμα 3.12: Κατάτμηση αντικειμένων με πολύ διαφορετική δομή. Ο αλγόριθμος σταθμισμένων τομών δεν κάνει καμία υπόθεση για το σχήμα των clusters.

Αυτή η ιεραρχία που περιλαμβάνει ανακατασκευή και ανασύνθεση όμως των αντικειμένων που δημιουργούνται από τομές με λιγότερες κλάσεις, φαίνεται και στο παραπάνω σχήμα 3.12.

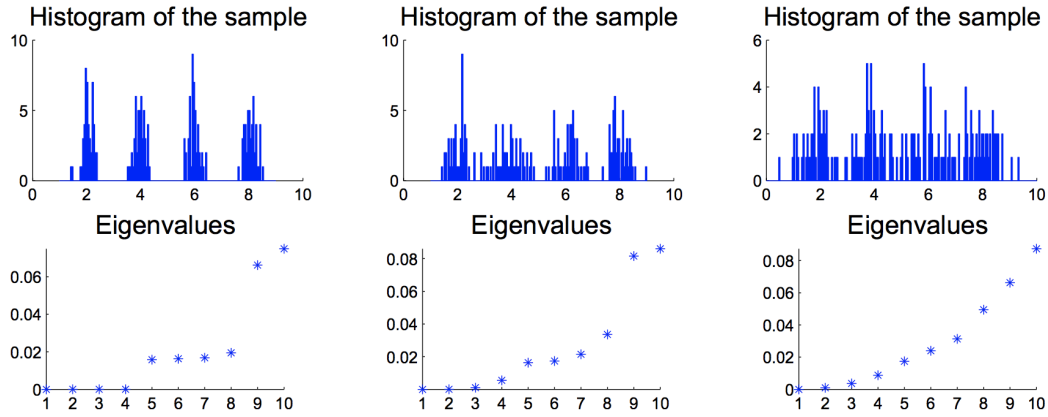


Παρότι στο (δ) η μπλε και η πορτοκαλί περιοχή είναι διαχωρισμένες, όταν απαιτούμε την διάσπαση σε μικρότερες δομές και αρχίζουν να διαχωρίζονται ο αετός και τα κλαδιά, ένα κομμάτι της πορτοκαλί και ένα κομμάτι της σκούρας μπλε περιοχής ενώνονται στο (ε) για να συνθέσουν το αριστερό κομμάτι του κλαδιού.

### 3.5 Προβλήματα που δημιουργούνται και μια πρόταση για να λυθούν

1. **Ο αριθμός των κλάσεων.** Σε όλη την ανάλυση ένα πράγμα που εξ αρχής θεωρήθηκε γνωστό είναι ο αριθμός των κλάσεων που αναζητούμε. Σε πραγματικά δεδομένα όμως, εικόνων ή άλλων δικτύων, δεν γνωρίζουμε εκ των προτέρων πόσα αντικείμενα ή clusters αναζητούμε. Το πρόβλημα εντοπισμού του αριθμού αυτού είναι ακόμη ανοιχτό για μη επιβλεπόμενες μεθόδους. Σε προβλήματα επιβλεπόμενης μάθησης, μπορούμε να εφαρμόσουμε μια μάσκα ή περισσότερες σε όλη την εικόνα και κατόπιν να αντιστοιχίσουμε από μια γνωστή βάση με παραδείγματα εκπαίδευσης μια πιθανότητα εύρεσης αντικειμένου. Μια παρόμοια τεχνική περιγράφεται στο [19]. Μια ευρετική για το πρόβλημα αυτό μπορεί να αναζητηθεί από την δομή των ιδιοτιμών του γράφου. Όπως έχουμε δει πολλές φορές, στην ιδανική περίπτωση που έχουμε  $K$  συνεκτικές συνιστώσες θα έχουμε επίσης  $k$  μη μηδενικές ιδιοτιμές, ενώ η επόμενη θα είναι θετική. Αν προστεθούν 'θορυβώδεις' ακμές μεταξύ των κλάσεων τότε παύουμε να έχουμε όλες τις  $K$  πρώτες ιδιοτιμές μηδενικές, αλλά αν η  $K + 1$  ιδιοτιμή του γράφου ήταν πολύ μεγάλη, τότε και η νέα σύνθεση του διαγράμματος των ιδιοτιμών παρουσιάζει μεγάλο *eigengap*. Ως *eigengap* θα ορίζουμε την διαφορά δύο διαδοχικών ιδιοτιμών του πίνακα Λαπλάσιαν. Στο [25], παρουσιάζεται η μορφή των ιδιοτιμών όσο αρχίζει να μεγαλώνει ο θόρυβος μεταξύ των κλάσεων και να επικαλύπτονται μεταξύ τους. Για μικρά επίπεδα θορύβου, όπως είδαμε και στην μέθοδο θεωρίας διαταραχών, το *eigengap* είναι ανθεκτικό και μπορούμε έτσι να αναζητήσουμε το πλήθος των κλάσεων κοιτώντας απλά το μεγαλύτερο *eigengap* στο γράφο των ιδιοτιμών. Σε μεγαλύτερα επίπεδα θορύβου η καμπύλη των ιδιοτιμών μοιάζει συνεχής (3.13) γιατί χρειάζεται κατάλληλο χώρο χαρακτηριστικών και μια σωστή αναπαράσταση του γράφου για να καταδείξουν αυτό το διάκενο. Αυτή η πληροφορία δεν έρχεται εύκολα σε προβλήματα κατάτμησης εικόνων και γιαυτό πολλές φορές η καμπύλη των ιδιοτιμών είναι συνεχής χωρίς μεγάλα άλματα και πρέπει να βρεθεί μια πιο αποτελεσματική μετρική για να περιγράψει με ακρίβεια την αυτόματη επιλογή της υπερπαραμέτρου  $K$ . Τέτοιες τεχνικές, τις οποίες αφήνουμε για μελλοντική επέκταση της διπλωματικής, μπορούν να αναζητηθούν σε γνωστά *quality measures* από την θεωρία του *machine learning*, όπως το cross entropy, αλλά με την προσθήκη κάποιων model regularizer, όπως το Minimum Description Length, για να αποφύγουμε το overfitting, που θα συμβαίνει για μεγάλες τιμές του  $K$ . Επίσης, το πρόβλημα του αριθμού των κλάσεων μπορεί να λυθεί και με την ιεραρχική τροποποίηση του προβλήματος σε

ένα framework για multiscale normalized cuts, που αναπτύσσεται σαν επέκταση της διπλωματικής και με την χρήση του framework των ημιεπιβλεπόμενων τεχνικών που αναπτύσσουμε στο επόμενο κεφάλαιο.



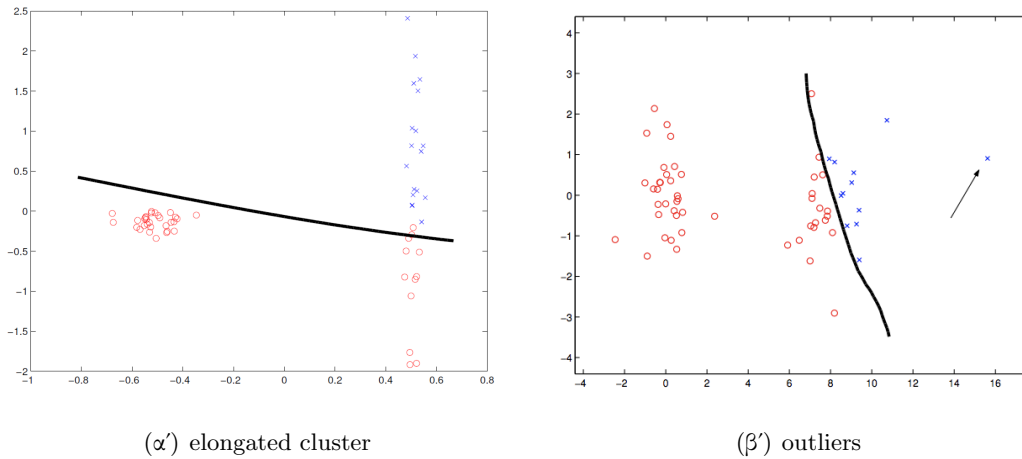
Σχήμα 3.13: Η μεταβολή του *eigengap* καθώς αυξάνονται τα επίπεδα θορύβου

- Η ευαισθησία σε outliers.** Οι σταθμισμένες τομές είναι ευαίσθητες σε outliers, ακριβώς όπως είναι κάθε μέσος όρος. Στο [31] οι συγγραφείς αναλύουν το φαινόμενο αυτό αντιστοιχίζοντας την λειτουργία του αλγορίθμου των Normalized Cuts, με ένα γραμμικό ταξινομητή που μεγιστοποιεί κάποιου είδους "gap" (στο πλαίσιο ανάλυσης των SVM), σε ένα χώρο άπειρης διάστασης. Αποδεικνύεται ότι το *gap* αυτό δίνει μεγαλύτερο βάρος σε σημεία που απέχουν πολύ από τα κεντροειδή των κλάσεων (σε κάθε κατεύθυνση, και προς αλλά και από το υπερεπίπεδο διαχωρισμού). Πιο επίσημα το *gap* που μεγιστοποιεί ο αλγόριθμος είναι:

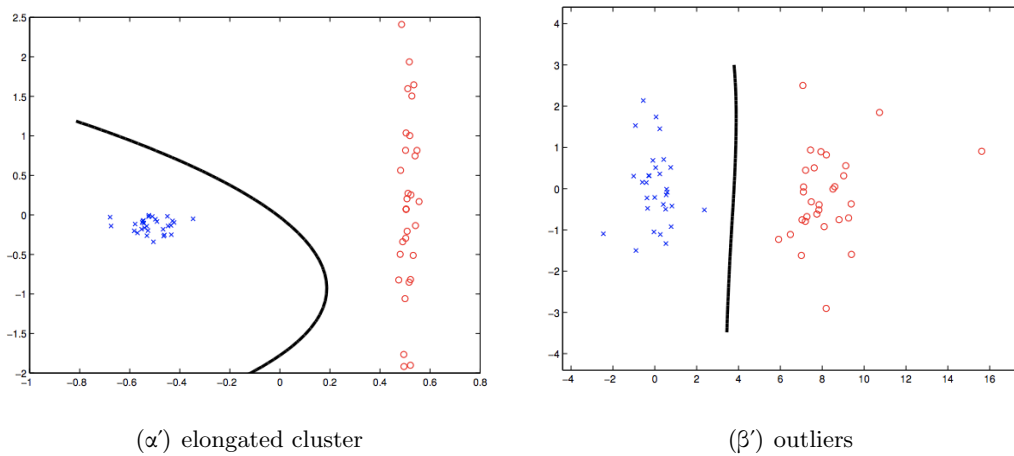
$$Gap(\beta) = \frac{1}{N} \sum_{i=1}^n \frac{1}{\cos \theta_i} (\beta^\top X_i)^2$$

όπου το  $\beta$  είναι το κάθετο διάνυσμα στο υπερεπίπεδο, το  $X_i$  είναι το (άπειρης διάστασης) διάνυσμα χαρακτηριστικών του  $i$ -οστού δεδομένου, και  $\theta_i$  η γωνία του διανύσματος αυτού με το μέσο όρο όλων των δεδομένων. Επειδή τα δεδομένα σταθμίζονται ώστε να βρίσκονται στην μοναδιαία σφαίρα η γωνία αυτή είναι και μια μετρική της απόστασης των σημείων από το μέσο όρο όλων των ιδιοδιανυσμάτων. Παρατηρούμε πως ο πολλαπλασιαστικός όρος  $\frac{1}{\cos \theta_i}$  είναι προβληματικός γιατί σταθμίζει σημεία με μεγαλύτερη απόσταση, με μεγαλύτερο βάρος. Αυτός είναι και ο λόγος που οι σταθμισμένες γραφοτομές δεν μπορούν να διαχωρίσουν καλά *elongated clusters*, όπως φαίνεται και στο σχήμα των συγγραφέων 3.14. Σαν λύση προτείνεται η αφαίρεση αυτού του όρου και η δημιουργία ενός *average gap* που όπως φαίνεται στο παρακάτω σχήμα 3.15



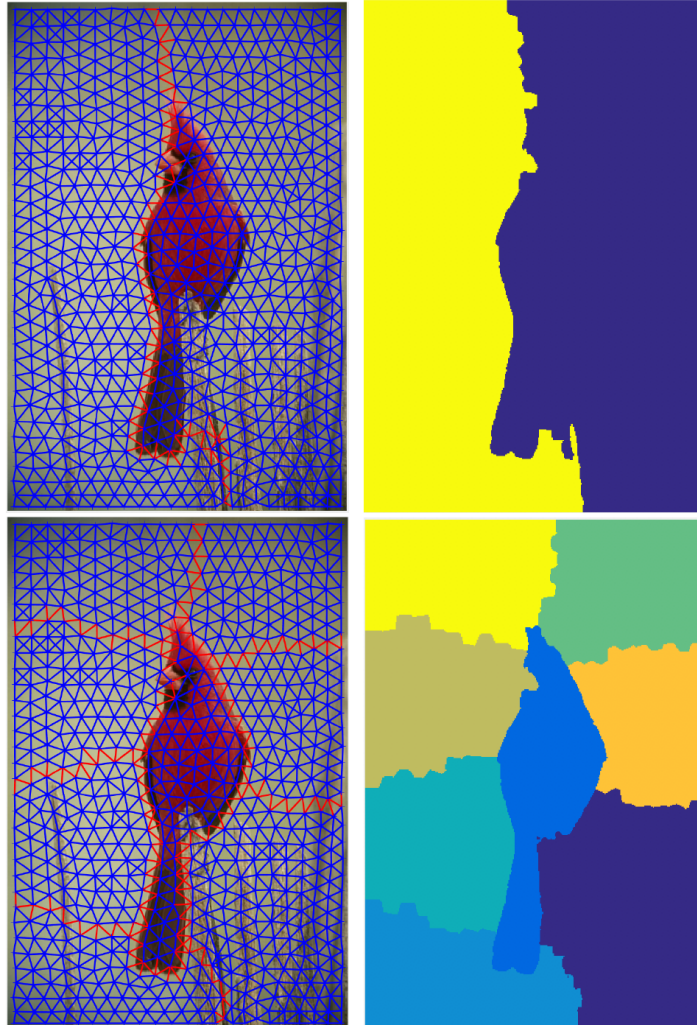


Σχήμα 3.14: Η αδυναμία του Normalized Cuts να διαχωρίσει elongated clusters και η ευαισθησία του σε outliers.



Σχήμα 3.15: Το average gap [31]

3. Η "υπερβολική" κανονικοποίηση των κλάσεων. Από τα παραπάνω σχήματα ένα πρόβλημα που είναι εμφανές στην κατάτμηση με σταθμισμένες τομές είναι ότι δεν μπορούν να ανακατασκευάσουν clusters που είναι σε πολύ διαφορετικές κλίμακες. Στο παρακάτω σχήμα 3.16, παρατηρούμε ότι παρότι έχουμε ένα αντικείμενο και το περιβάλλον του, με 2 κλάσεις δεν διαχωρίζεται το αντικείμενο. Χρειάζονται τουλάχιστον 8 κλάσεις για να γίνει σωστή σημασιολογική κατάτμηση. Από την άλλη τα ξύλα στα οποία κάθετα το πουλί δεν μπορούν να διαχωριστούν παρά με τόσα πολλά clusters που χάνεται η σημασιολογική κατάτμηση και πάλι πρέπει να συγχωνευτούν κάποιες κλάσεις για να γίνουν ολόκληρα αντικείμενα. Η δυνατότητα διαχωρισμού clusters που βρίσκονται σε πολύ διαφορετικές κλίμακες θα αποτελέσει μια επέκταση της διπλωματικής με την χρήση ενός multiscale framework.



Σχήμα 3.16: Η ανάγκη για κατάτμηση αντικειμένων σε πολλαπλές κλίμακες

## Κεφάλαιο 4

# Επέκταση των σταθμισμένων τομών με τεχνικές ημιεπιβλεπόμενης μάθησης

### 4.1 Εισαγωγή a-priori γνώσης για την βελτίωση του clustering

Εξαιτίας της μη-επιβλεπόμενης φύσης της διαδικασίας της κατάτμησης, εξαιρετικό βάρος για την απόδοση της λύσης επωμίζεται η κατασκευή του γράφου ως προς την τοπολογία και ως προς τον ορισμό της ομοιότητας δύο κόμβων που χαρακτηρίζονται από ένα διάνυσμα χαρακτηριστικών. Επιπλέον αρκετές φορές οι κλάσεις που δημιουργεί ένας μη-επιβλεπόμενος αλγόριθμος δεν ανταποκρίνονται στην σημασιολογική πληροφορία που περιέχεται στο γράφο, πρόβλημα το οποίο στην βιβλιογραφία αναφέρεται ως σημασιολογικό κενό (semantic gap). Από την άλλη υπάρχει συχνά και το φαινόμενο του clustering ambiguity, κατά το οποίο τα δεδομένα μπορεί να επεξηγούνται με αρκετές διαφορετικές κατατμήσεις και ανάλογα με την εφαρμογή εμείς να χρειαζόμαστε μια μόνο από αυτές. Παραδείγματος χάρη αν έχουμε ένα σύνολο ανθρώπων διαφορετικών φύλων, διαφορετικών ηλικιών, σκουρόχρωμων ή όχι κλπ. Συχνά λοιπόν για την μετρίαση των παραπάνω προβλημάτων οι αλγόριθμοι κατάτμησης ενσωματώνουν εξωτερική πληροφορία που δίνεται από τον χρήστη, από γνώση για την φύση της εφαρμογής ή από ετικέτες των κλάσεων κλπ.

Ιδιαίτερα στο πρόβλημα της κατάτμησης εικόνων και του εντοπισμού αντικειμένων, επειδή τυπικά οι αλγόριθμοι δουλεύουν *bottom – up* βρίσκοντας αρχικά, χαρακτηριστικά χαμηλού επιπέδου (χρώμα, φωτεινότητα, υφή, ακμές) και κατόπιν συνδυάζοντας *pixel* μεταξύ τους, συχνά έχουν το πρόβλημα του να υπερβούν το σημασιολογικό κενό που δημιουργείται μεταξύ χαρακτηριστικών χαμηλού επιπέδου και αντικειμένων αφαιρετικού επιπέδου. Παραδείγματος χάρη, στο 4.1 δεν θα ήταν λογικό να απαιτήσει κάποιος από έναν αλγόριθμο κατάτμησης να συμπεριλάβει στο ίδιο *cluster* και την κίτρινη φούστα και το κορμό του

άντρα (Σχήμα 4.1α). Για να γίνει αυτό θα πρέπει να δοθεί κάποια εξωτερική πληροφορία όπως φαίνεται στην εικόνα (Σχήμα 4.1δ: *unary constraints*). Επομένως, το καλύτερο που μπορούμε να περιμένουμε από τέτοιους αλγορίθμους χωρίς επίβλεψη είναι κάποια ιεραρχική συσταδοποίηση (Σχήμα 4.1β,γ).



Σχήμα 4.1: Χρειάζεται η ενσωμάτωση εξωγενούς πληροφορίας στο πρόβλημα της κατάτμησης, γιατί μόνο έτσι μπορεί να διαχωριστεί η αντανάκλαση του βράχου και να ενοποιηθεί η κλάση του ανθρώπου. Αλλιώς, μια ιεραρχική κατάτμηση είναι αρκετά ικανοποιητική.

Τέλος αξίζει κανείς να παρατηρήσει ότι η επίλυση *semi-supervised* προβλημάτων, ανοίγει το ‘μαύρο κουτί’ του *Normalized Cut* και κατ’ επέκταση όλων των φασματικών γραφοθεωρητικών μεθόδων, ώστε να εφαρμοστούν σε *classification* προβλήματα μέσα από *learning* τεχνικές. (Ειδικότερα, *transduction learning* για εκμάθηση από ταξινομημένα και αταξινομητα δεδομένα, όπως εκμάθηση του πίνακα βαρών κλπ). Το να σχεδιάζει κανείς *full – supervised* αλγορίθμους (που δεν λαμβάνουν υπόψη μη ταξινομημένα δεδομένα) είναι λιγότερο χρήσιμο όταν δεν έχουμε πολλά κατηγοριοποιημένα δεδομένα, πράγμα πολύ συχνό στις περισσότερες εφαρμογές, ειδικά στις πιο πρωτότυπες.

## 4.2 Εισαγωγή a-priori πληροφορίας στον αλγόριθμο *Normalized Cuts*

Μια κατεύθυνση προς την οποία θα μπορούσαμε να βελτιώσουμε την ποιότητα του αλγορίθμου των *Normalized Cuts* και να απαλύνουμε προβλήματα που δημιουργούνται από την κατασκευή του γράφου ή τη φύση του προβλήματος (*NP-complete*: όχι ολικό ελάχιστο, πολύ ισορροπημένα *clusters* ακόμη και αν τα *objects* είναι διαφορετικών μεγεθών), είναι να ενσωματώσουμε *prior information* κατά την διαδικασία της κατάτμησης. Αυτή η πληροφορία θα δίνεται από τον χρήστη (*interactive segmentation*) για βελτίωση της ποιότητας της κατάτμησης ή από άλλη εξωτερική πηγή (π.χ. έξοδος ενός *object detector*) και θα ενσωματώνεται στον πρόβλημα βελτιστοποίησης σαν περιορισμός του χώρου επιχτών λύσεων (*hard constraint*). Οι περιορισμοί αυτοί θα λαμβάνουν δυαδικές τιμές. Εναλλακτικά μπορεί να ενσωματωθούν οι περιορισμοί σαν *soft constraints*. Υπό αυτή την έννοια δημιουργούνται 2 ειδών *soft* κατευθύνσεις. Σύμφωνα με την πρώτη αντί να ικανοποιούμε όλους τους

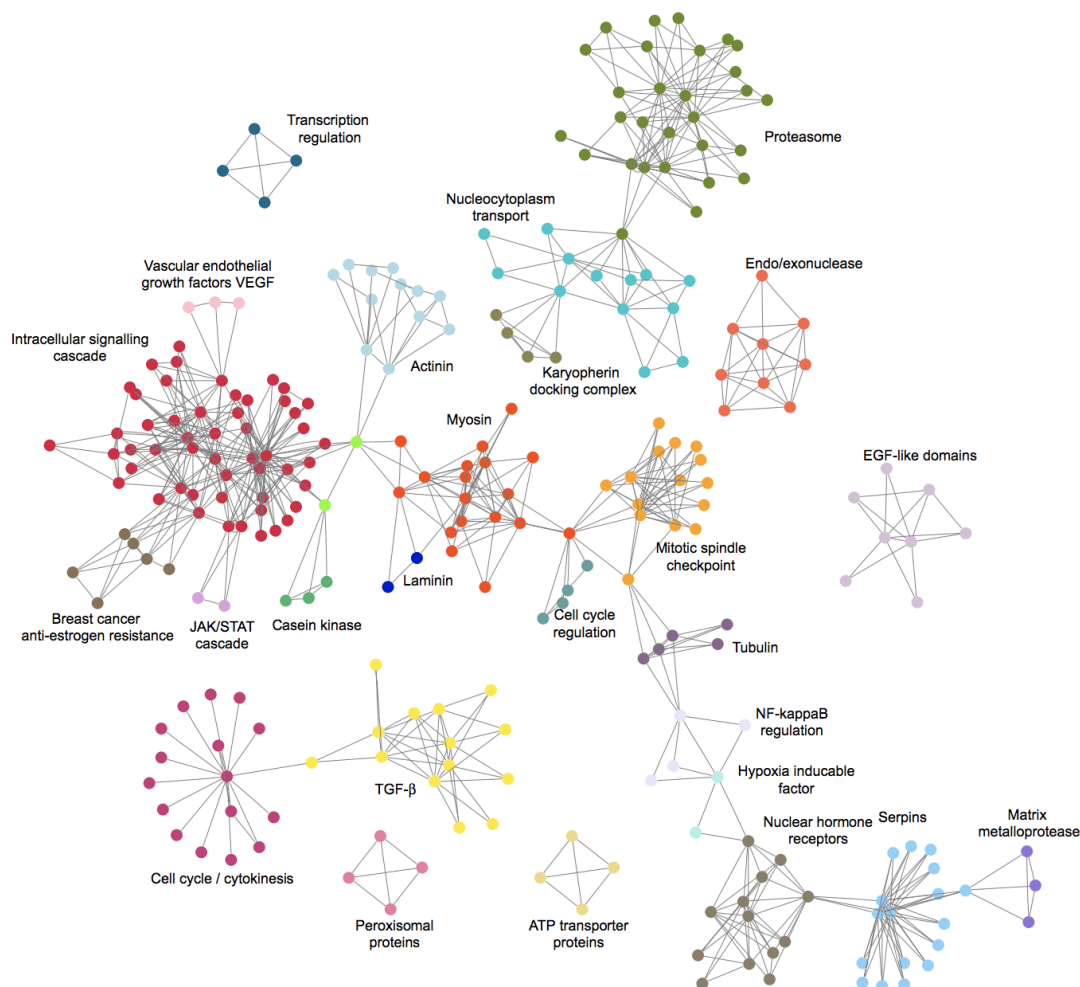
περιορισμούς, προσπαθούμε να βρούμε μια αρκετά καλή σταθμισμένη τομή που ικανοποιεί αρκετούς περιορισμούς, μια όσο το δυνατόν μικρότερη σταθμισμένη τομή που έχει όσο το δυνατόν μεγαλύτερη συσχέτιση με τους περιορισμούς. Έτσι οι περιορισμοί ικανοποιούνται μεν, αλλά σε κάποιο βαθμό. Η άλλη κατεύθυνση που εννοούμε τους *soft* περιορισμούς είναι μέσα από την εισαγωγή μιας αβεβαιότητας για την ισχύ των επιμέρους περιορισμών (*degree of belief*). Συνήθως στην δεύτερη κατηγορία μας ενδιαφέρει να ικανοποιήσουμε αθροιστικά και όχι μεμονωμένα όσο το δυνατόν μεγαλύτερο κομμάτι των περιορισμών. Γι' αυτό και επιβάλλονται με την μορφή μιας καθολικής συνάρτησης είτε σαν περιορισμοί στο πρόβλημα βελτιστοποίησης ([43]) είτε σαν *regularizer* στην αντικειμενική συνάρτηση. Επειδή σε καμία κατεύθυνση δεν μας νοιάζει η τήρηση μεμονωμένων περιορισμών τους λέμε και τους δύο *soft*. Και οι 2 αυτές μορφές *soft* περιορισμών είναι πίο εύρωστες σε λάθη από ότι οι αντίστοιχοι *hard* περιορισμοί που μπορεί να δοθούν από το χρήστη. Επίσης, συνήθως οδηγούν σε γρηγορότερους αλγορίθμους αφού δεν ικανοποιούν όλους τους περιορισμούς.

Η εξωτερική πληροφορία μπορεί να δίνεται στο πρόβλημά μας μέσω περιορισμών για κάθε κόμβο (*unary constraints*), μέσω περιορισμών σε ζεύγη κόμβων (*pairwise constraints*), μέσω περιορισμών του μεγέθους των κλάσεων κλπ. Οι περιορισμοί που αφορούν ζεύγη κόμβων ουσιαστικά αποτελούν μια επιπλέον πηγή ομοιότητας για δύο κόμβους που πρέπει να συνδυαστεί κατάλληλα με την υπάρχουσα πληροφορία από την ομοιότητα των χαρακτηριστικών των δύο αυτών κόμβων. Μια άλλη κατεύθυνση επομένως που θα ακολουθήσουμε είναι να επέμβουμε άμεσα στους πίνακες ομοιότητας και άρα στο γράφο για να επιβάλλουμε τους περιορισμούς που αφορούν ζεύγη κόμβων. Στα προβλήματα που αυτό είναι δυνατό θα ερμηνεύουμε το περιορισμό που επιβάλλαμε στο πρόβλημα βελτιστοποίησης και θα τον συσχετίζουμε με μετασχηματισμό στον αρχικό γράφο. Έτσι θα ενοποιήσουμε αυτές τις δύο λογικές που ακολουθούνται στην βιβλιογραφία του *semi-supervised graph clustering* και θα θέσουμε σε πίο στέρεες θεωρητικές βάσεις τους μετασχηματισμούς γράφων για ενσωμάτωση περιορισμών.

#### 4.2.1 Τοποθέτηση 2 κόμβων στο ίδιο cluster (**Must Link Constraint**)

Χωρίς να ξέρουμε σε ποιό cluster ανήκουν οι κόμβοι, περιορίζουμε τις πιθανές τομές σε αυτές που δεν τους χωρίζουν. Με αυτό τον τρόπο θα μπορούσαμε να ομαδοποιήσουμε από συστάδες pixel μέχρι ολόκληρα αντικείμενα, ανάλογα με τον αριθμό κόμβων που δεσμεύουμε. Εκτός από *hard ML constraints* μπορεί κάποιος να έχει στην διάθεσή του μια τιμή που δείχνει πόσο πιθανό είναι να ανήκουν μαζί αυτοί οι δύο κόμβοι. Μάλιστα αυτή η τεχνική (με την οποία δεν θα ασχοληθώ εδώ) είναι πίο εύρωστη σε θόρυβο που προέρχεται από λάθος περιορισμούς του χρήστη. Συνήθως, *hard ML constraints* δίνονται από το χρήστη, ενώ *soft ML constraints* μπορεί να αποκτηθούν αυτόματα αν έχει κάποιος στην διάθεση του μια ακολουθία εικόνων χρησιμοποιώντας *optical flow* και παίρνοντας τροχιές από σημεία, τις οποίες μετά συγκρίνει ως προς την πιθανότητα να ανήκουν στην ίδια κλάση [17]. Επιπλέον, οι *hard ML* περιορισμοί είναι μερικές φορές απαραίτητοι, όπως σε

περιπτώσεις διάσπασης πρωτεϊνών που κάποιοι κόμβοι που συγκροτούν ένα μακρομόριο είναι απαραίτητο να συνδεθούν στην τελική κατάτμηση αλλιώς χάνεται η σημασιολογία της δομής. Τα cluster που σχηματίζονται από την κατάτμηση αυτών των δικτύων μπορεί να αποτελούν ένα μακρομόριο ή σύνδεση πολλών μακρομορίων. Αν ισχύει το δεύτερο, όπως φαίνεται και στο σχήμα 4.2, θα θέλαμε ιδανικά όλοι οι κόμβοι ενός μακρομορίου να βρίσκονται στο ίδιο cluster (χρωματισμένοι κόμβοι). Μέσα από την διάσπαση τέτοιων δικτύων μπορεί να ανακαλυφθούν σημαντικές ακολουθίες μακρομορίων που κατόπιν χρησιμοποιούνται από βιολόγους και φαρμακοποιούς για την ερμηνεία ανθρώπινων λειτουργιών, την κατασκευή φαρμάκων και λοιπά.



Σχήμα 4.2: Δίκτυο Πρωτεϊνών. Ιδανικά, θα θέλαμε ένα εξωτερικό ειδήμονα ή ένα προστάδιο του αλγορίθμου, να επιβάλλει στους όμοια χρωματισμένους κόμβους του μακρομορίου να ανήκουν στην ίδια κλάση



### Μοντελοποίηση Περιορισμού

$$K_{u_i} = K_{u_j} \Leftrightarrow X_t(i) = X_t(j), \forall t \in [1, \dots, K] \Leftrightarrow (e_i - e_j)^T \cdot \mathbf{X} = 0$$

Όταν έχουμε ένα σύνολο  $S = (S_1, \dots, S_C)$  από  $C$  δυάδες κόμβων δεσμευμένων να ανήκουν στο ίδιο cluster τότε ο περιορισμός γίνεται:

$$\mathbf{B}_\sigma^T \cdot \mathbf{X} = 0,$$

όπου ο  $\mathbf{B}$  είναι ο πίνακας πρόσπτωσης μεγέθους  $N \times C$  με τιμές (για διατεταγμένα ζεύγη περιορισμών):

$$b_{il} = \begin{cases} +1 & \exists j \ S_l = (i, j) \\ -1 & \exists j \ S_l = (j, i) \\ 0 & otherwise \end{cases} \quad (4.1)$$

ενώ  $\sigma$  είναι μια συνάρτηση κατεύθυνσης που μετατρέπει τους περιορισμούς σε διατεταγμένα ζεύγη ώστε κάθε στήλη του  $\mathbf{B}_\sigma$  να παίρνει ένα +1 και ένα -1 στις κατάλληλες θέσεις.

Επειδή η εξίσωση είναι ομογενής, ο περιορισμός είναι (column) scale-invariant και επομένως περνάει αυτοούσιος στο ισοδύναμο πρόβλημα. Επίσης, περνάει αυτοούσιος στο relaxed πρόβλημα με την ίδια ερμηνεία. Τέλος είναι και rotation-invariant με την έννοια ότι για κάθε πίνακα  $\mathbf{X}$  που τον ικανοποιεί, ο πίνακας  $\mathbf{X}\mathbf{R}$  όπου  $\mathbf{R}^T\mathbf{R} = \mathbf{I}$  τον ικανοποιεί επίσης.

### Κατασκευή/Ελαχιστοποίηση πίνακα $\mathbf{B}_\sigma^T$

Κατά την ανάθεση περιορισμών από τον χρήστη υπάρχει μεγάλη πιθανότητα να προκύψουν πλεονασματικοί περιορισμοί. Παραδείγματος χάρη αν επιβληθεί οι κόμβοι 1 και 2 να ανήκουν στο ίδιο cluster, και μετά η ίδια συνθήκη για τους κόμβους 2,3 τότε είναι προφανές ότι οι κόμβοι 1,3 θα μπουν στο ίδιο cluster και δεν θα χρειαστεί νέος περιορισμός γι' αυτό. Ή αν δοθεί ένα σύνολο κόμβων από τον χρήστη, με τον περιορισμό να ανήκουν στο ίδιο cluster, τότε εμείς θα πρέπει να κατασκευάσουμε τον πίνακα  $\mathbf{B}_\sigma^T$  χωρίς πλεονασματικούς περιορισμούς (που αντιστοιχούν σε γραμμές του πίνακα) ώστε να μειώσουμε την πολυπλοκότητα του αλγορίθμου κατάτμησης.

Ένας πίνακας  $\mathbf{B}'_\sigma^T$  θα λέγεται *ισοδύναμος* με τον  $\mathbf{B}_\sigma^T$  αν ο χώρος λύσεων της εξίσωσης  $\mathbf{B}'_\sigma^T \mathbf{X} = 0$  είναι ίδιος με τον αρχικό. Ο πίνακας  $\mathbf{B}'_\sigma^T$  θα λέγεται *ελάχιστος* αν είναι ισοδύναμος με τον αρχικό και έχει το μικρότερο πλήθος γραμμών από όλους τους ισοδύναμους με τον αρχικό πίνακες. Θα αποδείξουμε ένα συστηματικό τρόπο για την κατασκευή ενός ελάχιστου  $\mathbf{B}_\sigma^T$  ή την ελαχιστοποίηση ενός γνωστού πίνακα  $\mathbf{B}_\sigma^T$ .

Η είσοδος μπορεί να δίνεται από τον χρήστη είτε με την μορφή ενός συνόλου από δυάδες κόμβων, είτε με την μορφή ενός πίνακα  $\mathbf{B}_\sigma^T$  (του οποίου η αραιή αναπαράσταση είναι

όμοια με το σύνολο από δυάδες), είτε με την μορφή  $N$ -άδων κόμβων. Θα ασχοληθούμε με την πρώτη μορφή. Η τρίτη αποτελεί μια πολύ μικρή τροποποίηση στην αρχικοποίηση του αλγορίθμου.

Θεωρούμε ότι το σύνολο των περιορισμών δεν έχει δυάδες της μορφής  $(i, i)$ , και έχει το πολύ μια εκ των  $(i, j), (j, i)$ . Αυτή η υπόθεση είναι λογική γιατί η διμελής σχέση σύμφωνα με την οποία 2 κόμβοι ανήκουν στο ίδιο cluster, είναι ανακλαστική και συμμετρική επομένως μπορούμε να απλοποιήσουμε τα συγκεκριμένα ζεύγη χωρίς να χάσουμε πληροφορία. Μάλιστα, η σχέση είναι και μεταβατική, δηλαδή συνολικά είναι σχέση ισοδυναμίας, επομένως αν απλοποιήσουμε και τις μεταβατικές ακμές θα έχουμε ένα σύνολο που είναι ισοδύναμο και ελάχιστο. Αυτό το σύνολο ψάχνει ο αλγόριθμός μας.

Αρχικά, θα χρειαστούμε την έννοια του *γράφου περιορισμών*. Έτσι θα ονομάζουμε τον γράφο του οποίου οι κόμβοι το σύνολο των κόμβων του αρχικού γράφου, ενώ μια ακμή μεταξύ 2 κόμβων προκύπτει από τον αντίστοιχο περιορισμό. Από τις υποθέσεις που κάναμε για το σύνολο των περιορισμών ο γράφος θα είναι απλός, μη-κατευθυνόμενος και χωρίς ανακυκλώσεις.

#### Αλγόριθμος ελαχιστοποίησης:

1. Φτιάξε τον γράφο των περιορισμών από το σύνολο  $S$ .
2. Διέσχισε το γράφο με Depth First Search.
3. Οι ακμές των DFS δέντρων (γιατί πιθανότατα ο γράφος έχει πολλές συνεκτικές συνιστώσες) συγκροτούν ακριβώς το ελάχιστο σύνολο περιορισμών. Κάθε ακμή  $(i, j)$  του δέντρου γίνεται μια στήλη  $e_i - e_j$  στον πίνακα  $\mathbf{B}$ .

*Σημείωση:* Υποθέτουμε ότι ο χρήστης φροντίζει να δώσει έτσι τους περιορισμούς ώστε να υπάρχει η δυνατότητα για σχηματισμό  $K$  cluster. Μπορούμε σίγουρα να το ελέγξουμε αυτό μετρώντας τις συνεκτικές συνιστώσες στον γράφο των περιορισμών μόλις τελειώσει ο DFS και να καταστήσουμε το σύνολο περιορισμών άκυρο αν αυτές είναι κάτω από  $K$ .

Για την απόδειξη ορθότητας του αλγορίθμου θα αποδείξω πρώτα το παρακάτω θεώρημα.

**Θεώρημα 4.7.** *Ο πίνακας  $\mathbf{B}_\sigma^T$  είναι ελάχιστος αν και μόνο αν όλες οι γραμμές του είναι γραμμικά ανεξάρτητες.*

*Παράδειγμα:* Όπως προηγουμένως αν συμπεριλάβουμε τους περιορισμούς  $(1,2), (2,3), (1,3)$  τότε οι 3 πρώτες γραμμές του πίνακα  $\mathbf{B}_\sigma^T$  είναι  $(e_1 - e_2)^T, (e_2 - e_3)^T, (e_1 - e_3)^T$ . Παρατηρούμε ότι οποιεσδήποτε 2 από αυτές είναι γραμμικά ανεξάρτητες, ενώ και οι 3 είναι γραμμικά εξαρτημένες αφού η τελευταία π.χ. είναι το άθροισμα των 2 πρώτων. Όπως βλέπουμε 2 γραμμές αρκούν για να ικανοποιηθεί ο περιορισμός και όπως γνωρίζαμε 2 περιορισμοί ήταν αρκετοί. Ο τρίτος προκύπτει.



Απόδειξη. Ο περιορισμός  $\mathbf{B}_\sigma^T \mathbf{X} = \mathbf{B}_\sigma^T [X_1, \dots, X_K] = [\mathbf{B}_\sigma^T X_1, \dots, \mathbf{B}_\sigma^T X_K] = \mathbf{0}$  είναι ισοδύναμος με το να πούμε ότι θέλουμε όλες οι στήλες του πίνακα  $\mathbf{X}$  να ανήκουν στο *nullspace* του  $\mathbf{B}_\sigma^T$ . Δηλαδή, από αυτή μόνο την εξίσωση ο χώρος λύσεων είναι ακριβώς το σύνολο  $(\text{null}(\mathbf{B}_\sigma^T))^K$ . (Παρεπιπτόντως, επειδή απαιτήσαμε οι στήλες του  $\mathbf{X}$  είναι γραμμικά ανεξάρτητες, η διάσταση του *nullspace* του  $\mathbf{B}_\sigma^T$  πρέπει να είναι τουλάχιστον  $K$ .) Αρκεί τώρα που ερμηνεύσαμε έτσι τον περιορισμό να παρατηρήσουμε ότι κάθε φορά που μπαίνει μια γραμμικά ανεξάρτητη γραμμή στον πίνακα, το *rank* του ανεβαίνει (αφού ανεβαίνει το *rank* του ανάστροφου) και άρα η διάσταση του *nullspace* κατεβαίνει. Άρα δεν θα μπορούσαμε ποτέ να αφαιρέσουμε μια γραμμικά ανεξάρτητη γραμμή. Αντίθετα, αν συμπεριλάβουμε ένα περιορισμό που είναι γραμμικά εξαρτημένος από τους υπόλοιπους τότε ο χώρος στηλών και το *nullspace* μένουν ίδια. Οπότε αφαιρώντας μια τέτοια γραμμή δημιουργούμε έναν ισοδύναμο πίνακα με λιγότερες γραμμές. Συνολικά επομένως, αφαιρώντας όλες τις γραμμικά εξαρτημένες γραμμές δημιουργούμε τον ελάχιστο πίνακα.  $\square$

**Θεώρημα 4.8.** Όταν ο γράφος περιορισμών έχει πολλές συνεκτικές συνιστώσες τότε ο βαθμός του συνολικού πίνακα περιορισμών είναι το άθροισμα των βαθμών των πινάκων περιορισμών κάθε συνεκτικής συνιστώσας.

Απόδειξη. Δύο ακμές-περιορισμοί που ανήκουν σε διαφορετικές συνεκτικές συνιστώσες έχουν τα άκρα τους σε τελείως διαφορετικούς κόμβους. Επομένως, οι περιορισμοί  $(i, j)$ ,  $(i', j')$  αντιστοιχούν σε γραμμικά ανεξάρτητες γραμμές του πίνακα γιατί έχουν μη επικαλυπτόμενο *support*. Μπορούμε επομένως να ομαδοποιήσουμε τους κόμβους κάθε συνεκτικής συνιστώσας και επειδή κάθε γραμμή του πίνακα θα έχει μηδενικά στις θέσεις όλων των κόμβων που ανήκουν σε διαφορετικές συνεκτικές συνιστώσες (έστω  $l$ ) τελικά ο συνολικός πίνακας γράφεται ως:

$$\mathbf{B}_\sigma^T = \begin{bmatrix} \mathbf{B}_1 & 0 & \dots & 0 \\ 0 & \mathbf{B}_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \mathbf{B}_l \end{bmatrix}$$

Ο βαθμός του συνολικού πίνακα όπως γνωρίζουμε είναι το άθροισμα των βαθμών των επιμέρους πινάκων.  $\square$

Το επόμενο θεώρημα ολοκληρώνει την απόδειξη ορθότητας του αλγορίθμου μας.

**Θεώρημα 4.9.** Ο πίνακας  $\mathbf{B}_\sigma^T$  είναι ελάχιστος αν και μόνο αν ο γράφος περιορισμών που δημιουργεί αποτελείται από ένα σύνολο συνεκτικών δέντρων, ένα σε κάθε συνεκτική συνιστώσα.

Θα αποδείξω το παραπάνω θεώρημα με 3 τρόπους, έναν που θα χρησιμοποιεί θεωρία γραμμικής άλγεβρας, έναν που θα χρησιμοποιεί γραφοθεωρητικές μεθόδους και έναν που θα χρησιμοποιεί θεωρία διμελών σχέσεων, δείχνοντας έτσι την εντυπωσιακή σύνδεση αυτών των τριών τομέων.

Απόδειξη. (Γραφοθεωρητική) Ισοδύναμα θα αποδείξω ότι σε συνεκτικό γράφο περιορισμών οι γραμμικά εξαρτημένοι περιορισμοί είναι ακριβώς οι απλοί κύκλοι του γράφου. Παίρνοντας κατόπιν την ένωση των συνεκτικών συνιστωσών όπως είδαμε και στο προηγούμενο θεώρημα, αποδεικνύουμε το ζητούμενο. Από την μια αν έχουμε κύκλο στο γράφο, τότε έχουμε μια ακολουθία ακμών  $\{(i_1, i_2), \dots, (i_{r-1}, i_r), (i_r, i_1)\}$  δηλαδή στον πίνακα έχουμε γραμμές  $(e_1 - e_2)^T, \dots, (e_{r-1} - e_r)^T, (e_r - e_1)^T$ . Αν τις αθροίσουμε όλες μαζί έχουν άθροισμα το μηδενικό διάνυσμα που σημαίνει ότι είναι γραμμικά εξαρτημένες. Αυτό σημαίνει ότι μπορούμε να "σπάσουμε" κάθε κύκλο αφαιρώντας οποιαδήποτε ακμή του και να διατηρήσουμε ένα ισοδύναμο γράφο περιορισμών. Μπορούμε μήπως να σπάσουμε και ένα μονοπάτι βγάζοντας μια ακμή του; Η απάντηση είναι όχι. Κάθε μονοπάτι (μη-κλειστό)  $n$  κόμβων (άρα  $n-1$  ακμών) δημιουργεί ένα πίνακα βαθμού  $n-1$ . Αυτό συμβαίνει επειδή μόνο ένα άκρο είναι κοινό, οπότε ο πίνακας μεγέθους  $(n-1) \times n$  που αφορά τους κόμβους του μονοπατιού θα έχει στην κεντρική διαγώνιά του  $+1$  και στην από πάνω  $-1$ . Αν αφαιρέσουμε την τελευταία στήλη πέρνουμε έναν υποπίνακα  $(n-1) \times (n-1)$  άνω τριγωνικό με  $+1$  στην κεντρική διαγώνιο, άρα βαθμού  $n-1$ . Αυτός είναι και ο βαθμός του αρχικού υποπίνακα. Αν αφαιρέσουμε μια ακόμη ακμή από το μονοπάτι, θα καταλήξουμε με 2 μονοπάτια ανεξάρτητα (έστω  $x, n-x$  κόμβων). Επειδή είναι ανεξάρτητα ο βαθμός του συνολικού πίνακα που δημιουργούν είναι, από το προηγούμενο θεώρημα, το άθροισμα των βαθμών των υποπίνακων κάθε μονοπατιού. Αυτός είναι ακριβώς  $(x-1) + ((n-x)-1) = n-2$  (!) Άρα αν σπάσουμε ακμή από μονοπάτι ο πίνακας που δημιουργείται δεν είναι ισοδύναμος με τον αρχικό. Άρα κάθε μονοπάτι συγκροτεί ένα γραμμικά ανεξάρτητο σύνολο περιορισμών. Τελικά, το πλήθος των απλών κύκλων είναι σε αμφιμονοσήμαντη αντιστοιχία με το πλήθος των απλών κύκλων του γράφου. Επειδή ο γράφος είναι συνεκτικός υπάρχουν μονοπάτια σε όλους τους κόμβους. Άρα το ελάχιστο υπογράφημα θα ήταν ένα συνεκτικό γράφημα χωρίς κύκλους. Γνωρίζουμε από την θεωρία γράφων ότι το μέγιστο ακυκλικό γράφημα είναι ένα συνεκτικό δέντρο.  $\square$

Απόδειξη. (Θεωρία διμελών σχέσεων) Όπως αναφέραμε και πριν η σχέση  $\mathcal{R}(a, b) = "$ τα σημεία  $a, b$  βρίσκονται στο ίδιο cluster" είναι ανακλαστική, συμμετρική και μεταβατική. Άρα με άλλα λόγια είναι σχέση ισοδυναμίας. Γνωρίζουμε ότι η κλειστότητα μιας σχέσης ισοδυναμίας απεικονίζεται σε γράφημα ως μια ένωση από ανεξάρτητες κλίκες. Αρκεί λοιπόν να βρούμε την σχέση εκείνη με τις ελάχιστες ακμές τέτοια ώστε η κλειστότητά της να είναι ίδια με την κλειστότητα της αρχικής. Αρκεί και πρέπει να διατηρήσουμε την συνεκτικότητα των συνιστωσών γιατί αν την χαλάσουμε και σπάσουμε παραδείγματος χάρη μια από αυτές σε περισσότερες δεν θα υπάρχει μονοπάτι που θα συνδέει τους αντίστοιχους κόμβους και άρα αποκλείεται η κλειστότητα της τελικής σχέσης να έχει τον ίδιο αριθμό κλικών με την αρχική. Από την άλλη, αν διατηρήσουμε την συνεκτικότητα τότε διατηρούμε τουλάχιστον ένα μονοπάτι από κάθε κόμβο της συνιστώσας σε κάθε άλλον και συνεπώς, από την μεταβατική ιδιότητα η κλειστότητα θα περιέχει όλες τις απευθείας ακμές μεταξύ αυτών των κόμβων. Επειδή θέλουμε την ελάχιστη σχέση, τελικά ψάχνουμε ένα γράφημα ελάχιστα συνεκτικό σε κάθε συνιστώσα. Από θεωρία γράφων γνωρίζουμε ότι αυτό είναι οποιοδήποτε σύνολο, συνεκτικών σε κάθε συνιστώσα, δέντρων.  $\square$

Απόδειξη. (Από γραμμική Άλγεβρα) Όπως είδαμε πριν κάθε πίνακας ισοδύναμος με τον αρχικό θα πρέπει να έχει το ίδιο *nullspace*. Ας βρούμε λοιπόν το *nullspace* του αρχικού πίνακα. Είναι,

$$\mathbf{B}_\sigma^T \mathbf{x} = 0 \iff x_i = x_j, \quad \forall (i, j) \in S \iff x_i = x_j, \quad \forall (i, j) \quad s.t. \quad G_i = G_j \quad (4.2)$$

όπου  $G_i$  είναι η συνεκτική συνιστώσα στην οποία ανήκει ο κόμβος  $i$ . Άρα το *nullspace* είναι ακριβώς το σύνολο που κάνουν *span* οι δείκτριες συναρτήσεις κάθε συνεκτικής συνιστώσας. Επομένως, ο τρόπος να ελαχιστοποιήσουμε τον πίνακα είναι να αφαιρούμε γραμμές σεβόμενοι την συνεκτικότητα σε κάθε συνιστώσα. Έτσι, ο πυρήνας θα παραμείνει ίδιος και άρα ο νέος πίνακας θα είναι ισοδύναμος. Τελικά, ζητάμε πάλι το γράφημα εκείνο που είναι ελάχιστα συνεκτικό σε κάθε συνεκτική συνιστώσα. Αυτό είναι ακριβώς ένα σύνολο από δέντρα συνεκτικά σε κάθε συνιστώσα, ένα σε κάθε μια.  $\square$

Το πλήθος των γραμμών του πίνακα  $\mathbf{B}_\sigma^T$  μπορεί έτσι να μειωθεί από  $\frac{(N-K)^2 + (N-K)}{2}$  σε  $N - K$ .

### Επίλυση του περιορισμού

Το πρόβλημα βελτιστοποίησης που καλούμαστε να λύσουμε (μετά την χαλάρωση) είναι:

$$\begin{aligned} & \underset{\mathbf{X}}{\text{minimize}} && \text{tr}(\mathbf{X}^T \mathbf{L} \mathbf{X}), \\ & \text{subject to} && \mathbf{X}^T \mathbf{D} \mathbf{X} = \mathbf{I}, \\ & && \mathbf{X} \mathbf{X}^\dagger \mathbf{1}_N = \mathbf{1}_N, \\ & && \mathbf{B}_\sigma^T \mathbf{X} = \mathbf{0} \end{aligned} \quad (4.3)$$

Τώρα που φτιάξαμε τον περιορισμό ώστε να είναι ελάχιστος πρέπει να βρούμε ένα τρόπο να τον επιβάλλουμε στο πρόβλημα. Όπως είδαμε και πριν θα πρέπει όλες οι στήλες του πίνακα  $\mathbf{X}$  να ανήκουν στον πυρήνα του  $\mathbf{B}_\sigma^T$ . Βιβλιογραφικά, αυτό το πρόβλημα έχει λυθεί στη γενική μορφή του από τους V.Golub et. al [14] και η λύση του έχει χρησιμοποιηθεί από τους J.Shi, S.Yu στο συγκεκριμένο πρόβλημα που σκοπεύουμε να λύσουμε ([47],[48],[50]). Ωστόσο, η λύση αυτή είναι αρκετά ακριβή υπολογιστικά γιατί δεν λαμβάνει καθόλου υπόψη την ειδική μορφή που έχει ο πίνακας περιορισμών.

Θα παρουσιάσω τις 2 λύσεις του Ξ.Γολυβ που χρησιμοποιήθηκαν από τον Θ.Σημ και κατόπιν την δικιά μου λύση που λαμβάνει υπόψη την μορφή του πίνακα περιορισμών και δίνει έτσι μια ενδιαφέρουσα νέα ερμηνεία στο πρόβλημα.

Καταρχάς, ο πίνακας  $\mathbf{B}_\sigma^T$  (μεγέθους  $C$ -επί- $N$ ) έχει μετά τον αλγόριθμο ελαχιστοποίησής του βαθμό  $C$ , γιατί όπως είδαμε όλες του οι γραμμές είναι γραμμικά ανεξάρτητες. Η διάσταση του πυρήνα του επομένως θα είναι  $N - C$ .

1. (Βιβλιογραφική Λύση 1) Βρίσκουμε τον πυρήνα του πίνακα περιορισμών κόνοντας  $QR$  διάσπαση στον πίνακα  $\mathbf{B}_\sigma$ .

$$\mathbf{B}_\sigma = \mathbf{Q}\mathbf{R}\mathbf{c} = [\mathbf{Q}_1(N \times C), \mathbf{Q}_2(N \times (N-C))] \begin{bmatrix} \mathbf{R}_{C \times C} \\ \mathbf{0}_{(N-C) \times C} \end{bmatrix} \quad (4.4)$$

όπου ο  $\mathbf{Q}$  είναι ορθομοναδιαίος και ο  $\mathbf{R}$  είναι άνω τριγωνικός. Κάθε πίνακας  $\mathbf{X}$  που ανήκει στον πυρήνα του  $\mathbf{B}_\sigma$  γράφεται σαν  $\mathbf{X} = \mathbf{Q}_2\mathbf{Y}$  για κάποιο πίνακα  $\mathbf{Y}$  μεγέθους  $(N-C) \times K$ . Πράγματι,

$$\begin{aligned} \mathbf{B}^T\mathbf{X} &= [\mathbf{R}^T_{C \times C}, \mathbf{0}^T_{C \times (N-C)}] \begin{bmatrix} \mathbf{Q}_1^T_{C \times N} \\ \mathbf{Q}_2^T_{(N-C) \times N} \end{bmatrix} \mathbf{Q}_2_{N \times (N-C)} \mathbf{Y} \\ &= [\mathbf{R}^T_{C \times C}, \mathbf{0}^T_{C \times (N-C)}] \begin{bmatrix} \mathbf{Q}_1^T \mathbf{Q}_2_{C \times (N-C)} \\ \mathbf{Q}_2^T \mathbf{Q}_2_{(N-C) \times (N-C)} \end{bmatrix} \mathbf{Y} \\ &= [\mathbf{R}^T_{C \times C}, \mathbf{0}^T_{C \times (N-C)}] \begin{bmatrix} \mathbf{0}_{C \times (N-C)} \\ \mathbf{I}_{(N-C) \times (N-C)} \end{bmatrix} \mathbf{Y} = \mathbf{0} \end{aligned}$$

Το πρόβλημα είναι ότι η  $QR$  διάσπαση είναι βαριά υπολογιστικά και επίσης ο πίνακας  $\mathbf{Q}_2$  δεν είναι μοναδικός. Υπάρχει η περίπτωση να μας βγει πολύ πυκνός και κατόπιν να πυκνώσει πολύ τον νέο *Laplacian* οπότε να κάνει ακόμη πιο ακριβή υπολογιστικά την διαδικασία. Η λύση μας όχι μόνο θα βρίσκει σε κλειστή μορφή τον πυρήνα σε γραμμικό χρόνο αλλά θα αποτελεί και την πιο αραιή αναπαράσταση του πυρήνα. Μπορούμε να λύσουμε έτσι πολύ μεγαλύτερα προβλήματα (εικόνες ή περιορισμούς) από παλιά.

2. (Βιβλιογραφική Λύση 2) Αυτή η λύση μοιάζει με γενίκευση των *Householder* μετασχηματισμών. Άρα θα μπορούσε κανείς να πει ότι είναι μια πιο έξυπνη υλοποίηση της προηγούμενης σκέψης. Ουσιαστικά προβάλλει τις λύσεις του προηγούμενου προβλήματος, στο χώρο επικτών λύσεων του παρόντος προβλήματος. Αποδεικνύεται ότι αν :

- $P = D^{-1}W$  είναι ο ανά γραμμή σταθμισμένος πίνακας βαρών.
- Ενώ ο  $Q = I - D^{-1}B(B^T D^{-1}B)^{-1}B^T$  είναι ο πίνακας προβολής στο χώρο επικτών λύσεων.

αποδεικνύεται ότι η βέλτιστη λύση του προβλήματος είναι τα  $K$  ιδιοδιανύσματα που αντιστοιχούν στις  $K$  μεγαλύτερες ιδιοτιμές του πίνακα  $QPQ$ .

Το ένα μειονέκτημα είναι ότι και αυτή η λύση είναι ακριβή υπολογιστικά. Με κάποια έξυπνα τρικ οι συγγραφείς στο [50] καταφέρνουν να διατηρήσουν την αύξηση της πολυπλοκότητας σε  $O(NC + C^2)$  το οποίο είναι αποδεκτό αν οι περιορισμοί είναι λίγοι γιατί αποτελεί μια γραμμική αύξηση στην πολυπλοκότητα, αλλά είναι ανεφάρμοστο αν οι περιορισμοί είναι της τάξης του  $O((N-K)^2)$ . Βέβαια, όπως είδαμε μετά την ελαχιστοποίηση το πολύ οι περιορισμοί να είναι της τάξης του  $O(N-K)$  οπότε η αύξηση της πολυπλοκότητας τετραγωνική. Επίσης, είναι παράλογο να έχει γίνει ήδη αρκετή δουλειά εξωτερικά

για να ομαδοποιηθούν κάποια *pixel* και αυτό να καταλήγει να δυσκολεύει περισσότερο τον αλγόριθμο. Στην ακραία περίπτωση που οι περιορισμοί σχηματίζουν μόνοι τους  $K$  *cluster* ο αλγόριθμος θα έπρεπε να τερματίζει, αλλά παρόλαυτά με αυτή την τεχνική θα καταλήξει να χρειάζεται πάρα πολύ ώρα ακόμη και για μέτριου μεγέθους εικόνες. Η δικιά μας λύση θα επιβαρύνει με καθαρά γραμμική αύξηση την πολυπλοκότητα του αλγορίθμου ( $O(N + C)$ ) λόγω προεπεξεργασίας, αλλά όσο πιά πολλοί είναι οι περιορισμοί τόσο θα βελτιώνει την πολυπλοκότητα του υπόλοιπου αλγορίθμου επειδή θα δουλεύει και αυτή σε μειωμένο χώρο λύσεων.

παρ

Επιπλέον, αυτή η λύση εκτός από το ότι είναι ακριβή υπολογιστικά, όπως αναφέρουν και οι συγγραφείς, χάνει εντελώς την ερμηνεία ότι η διαδικασία κατάτμησης με *Normalized Cuts* αποτελεί την κατάσταση ισορροπίας σε μια διαδικασία διάχυσης, επειδή ο πίνακας  $QPQ$  έχει και αρνητικά στοιχεία και δεν μπορεί να ερμηνευθεί ως πίνακας πιθανοτήτων μετάβασης. Αντίθετα, η λύση μας συνεχίζει, όπως είναι φυσικό, να διατηρεί την ίδια ερμηνεία.

Τελικά, και οι 2 προηγούμενες λύσεις μπορούν να βελτιωθούν ως προς την πολυπλοκότητά τους πολύ με την ελαχιστοποίηση του πίνακα περιορισμών. Ωστόσο, θέλουμε να κάνουμε περαιτέρω βελτίωση στην λύση του προβλήματος προτείνοντας ένα νέο αλγόριθμο μικρότερης πολυπλοκότητας.

### Νέος βελτιωμένος αλγόριθμος

Από εδώ και πέρα μπορούμε να θεωρούμε ότι ο πίνακας περιορισμών μεγέθους  $C \times N$  έχει  $C \leq N - K$  και είναι πλήρους βαθμού (κάτι που πετυχαίνουμε μετά την ελαχιστοποίηση). Άρα και ο πυρήνας είναι βαθμού  $N - C$  και οι συνεκτικές συνιστώσες του γράφου περιορισμών είναι επίσης  $N - C$ .

Όπως είδαμε και στην προηγούμενη απόδειξη (με γραμμική άλγεβρα), το σύνολο επιφύτων λύσεων έχει γνωστή λύση με κλειστή μορφή. Αυτή είναι ότι οι στήλες είναι δείκτριες συναρτήσεις των συνεκτικών συνιστωσών. Επομένως μπορούμε να βρούμε μια ορθοκανονική βάση  $\mathbf{B}^\perp$  για τον πυρήνα (που είναι μεγέθους  $N \times (N - C)$ ), το φυσικό νόημα της οποίας είναι σε ποιά συνεκτική συνιστώσα ανήκει κάθε κόμβος. Αυτό γίνεται προφανώς με χρήση του προηγούμενου *DFS*. Το εντυπωσιακό είναι ότι ενώ μέχρι τώρα χρησιμοποιούσαμε πίνακες για να λύσουμε προβλήματα γράφων, τώρα λύσαμε ένα πρόβλημα πινάκων (εύρεση πυρήνα πίνακα) χρησιμοποιώντας ένα αλγόριθμο γράφων. Το σημαντικό είναι ότι έτσι η εύρεση βάσης πυρήνα γίνεται σε γραμμικό χρόνο και όχι σε τουλάχιστον τετραγωνικό, όπως αν χρησιμοποιήσουμε *QR*-διάσπαση.

Έχοντας μια (ορθοκανονική) βάση μπορούμε να εκφράσουμε κάθε διάνυσμα που ανήκει στον πυρήνα ως:  $\mathbf{X} = \mathbf{B}^\perp \mathbf{Y}$ , όπου ο πίνακας  $\mathbf{Y}$  είναι μεγέθους  $(N - C) \times K$ .

Επειδή οι πίνακες  $\mathbf{B}^\perp$ ,  $\mathbf{X}$  έχουν σταθερές τιμές σε κάθε στήλη (στις μη-μηδενικές θέσεις) και ακριβώς μια μη-μηδενική τιμή σε κάθε γραμμή, το ίδιο θα ισχύει και για τον πίνακα  $\mathbf{Y}$ . Επόμενως, και ο πίνακας  $\mathbf{Y}$  ικανοποιεί τον 3ο περιορισμό και είναι πίνακας διαμέρισης. Η φυσική ερμηνεία του πίνακα αυτού είναι επομένως, σε ποιό cluster ανήκει κάθε συνεκτική συνιστώσα του γράφου περιορισμών. Ο πίνακας  $\mathbf{X}$  επομένως προκύπτει από τον πίνακα  $\mathbf{Y}$  αν αναπαράγουμε αυτούσια κάθε γραμμή του δεύτερου στον πρώτο για όλους τους κόμβους που ανήκουν στην συνεκτική συνιστώσα που αναπαριστά η γραμμή του δεύτερου.

Το πρόβλημά μας γίνεται,

$$\begin{aligned} & \underset{\mathbf{Y}}{\text{minimize}} && \text{tr}(\mathbf{Y}^T \mathbf{L}_r \mathbf{Y}), \\ & \text{subject to} && \mathbf{Y}^T \mathbf{D}_r \mathbf{Y} = \mathbf{I}, \\ & && \mathbf{Y} \mathbf{Y}^\dagger \mathbf{1}_{N-C} = \mathbf{1}_{N-C}, \end{aligned} \quad (4.5)$$

όπου  $\mathbf{L}_r = (\mathbf{B}^\perp)^T \mathbf{L} (\mathbf{B}^\perp)$  και  $\mathbf{D}_r = (\mathbf{B}^\perp)^T \mathbf{D} (\mathbf{B}^\perp)$ .

Επομένως είναι πρόβλημα γενικευμένων ιδιοτιμών για τους πίνακες  $\mathbf{L}_r, \mathbf{D}_r$ . Με την επόμενη ερμηνεία που θα δώσουμε δεν χρειάζεται καν να υπολογίσουμε αυτούς τους πίνακες μέσα από τους πολλαπλασιασμούς.

### Ερμηνεία πινάκων και περαιτέρω βελτίωση αλγορίθμου με αναγωγή σε ελαττωμένο γράφο

Μπορούμε να δουλέψουμε εξαρχής στον διακριτό χώρο, με την αρχική εκδοχή του προβλήματος μας και να παρατηρήσουμε μια διαισθητική ερμηνεία του προβλήματος όπως προκύπτει μετά την απαλοιφή του περιορισμού.

Το αρχικό πρόβλημα όπως είδαμε στο Κεφάλαιο 2 είναι:

$$\begin{aligned} & \underset{\mathbf{X}}{\text{minimize}} && \sum_{i=1}^K \frac{(X^i)^T \mathbf{L} X^i}{(X^i)^T \mathbf{D} X^i} \\ & \text{subject to} && (X^i)^T \mathbf{D} X^j = 0, \quad 1 \leq i, j \leq K, i \neq j \\ & && \mathbf{X} \mathbf{X}^\dagger \mathbf{1}_N = \mathbf{1}_N, \quad X \in \mathcal{X} \end{aligned} \quad (4.6)$$

Θέτοντας  $\mathbf{X} = \mathbf{B}^\perp \mathbf{Y}$  και γνωρίζοντας ότι και ο πίνακας  $\mathbf{Y}$  είναι πίνακας διαμέρισης όπως δείξαμε παραπάνω το πρόβλημα μετασχηματίζεται σε:

$$\begin{aligned} & \underset{\mathbf{Y}}{\text{minimize}} && \sum_{i=1}^K \frac{(Y^i)^T \mathbf{L}_r Y^i}{(Y^i)^T \mathbf{D}_r Y^i} \\ & \text{subject to} && (Y^i)^T \mathbf{D}_r Y^j = 0, \quad 1 \leq i, j \leq K, i \neq j \\ & && \mathbf{Y} \mathbf{Y}^\dagger \mathbf{1}_{N-C} = \mathbf{1}_{N-C}, \quad Y \in \mathcal{X} \end{aligned} \quad (4.7)$$

Θα προσπαθήσουμε να ερμηνεύσουμε τους αριθμητές και παρονομαστές των κλασμάτων. Αν ο γράφος  $\Gamma$  έχει κόμβους  $\{1, 2, \dots, N\}$  και συνεκτικές συνιστώσες  $\{G_1, G_2, \dots, G_{N-C}\}$  τότε αν ο πίνακας  $\mathbf{B}^\perp$  έχει σε κάθε στήλη του σταθερές  $\{(c_1 = 1, c_2 = 1, \dots, c_{N-C} = 1)\}$ :

- Ο πίνακας  $\mathbf{D}_r = (\mathbf{B}^\perp)^T \mathbf{D} (\mathbf{B}^\perp)$  είναι ένας πίνακας  $(N-C) \times (N-C)$  που προκύπτει ως εξής: Ο αριστερός πολλαπλασιασμός αθροίζει τις γραμμές του πίνακα  $\mathbf{D}$  που ανήκουν στην ίδια συνεκτική συνιστώσα πολλαπλασιάζοντάς τες με την αντίστοιχη σταθερά, ενώ ο δεξιός πολλαπλασιασμός αθροίζει τις στήλες του νέου πίνακα που ανήκουν στην ίδια συνεκτική συνιστώσα, ξαναπολλαπλασιάζοντάς τες με την ίδια σταθερά, με αποτέλεσμα ο τελικός πίνακας να είναι:

$$\mathbf{D}_r = \begin{bmatrix} vol(G_1) & 0 & \dots & 0 \\ 0 & vol(G_2) & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & vol(G_{N-C}) \end{bmatrix},$$

όπου  $vol(G_i) = \sum_{j \in G_i} d(u_j)$ .

- Αφού ο πίνακας  $\mathbf{Y}$  είναι πίνακας διαμέρισης όπως και ο  $\mathbf{X}$ , ενώ ο  $\mathbf{B}^\perp$  έχει άσσους στις στήλες, μπορούμε να θεωρήσουμε ότι οι τιμές των στηλών του  $\mathbf{Y}$  είναι ίδιες με τις τιμές των στηλών του  $\mathbf{X}$ . Ο παρονομαστής

$$(Y^i)^T \mathbf{D}_r Y^i = \alpha_i^2 \cdot \sum_{j: G_j \in K_i} vol(G_j) = \alpha_i^2 \cdot vol(K_i)$$

- Όμοια ο πίνακας  $\mathbf{L}_r = (\mathbf{B}^\perp)^T \mathbf{L} (\mathbf{B}^\perp)$  είναι ένας πίνακας  $(N-C) \times (N-C)$  που στις διαγώνιες θέσεις έχει τα στοιχεία του  $\mathbf{D}_r$  ελαττωμένα όμως κατά το *association* της κάθε συνεκτικής συνιστώσας. Στις μη διαγώνιες οι τιμές του πίνακα, έστω  $l_{ij}$  είναι με αρνητικό πρόσημο το αθροιστικό βάρος των ακμών που έχουν το ένα άκρο τους σε κόμβο που ανήκει στην  $i$ -οστή συν.συνιστώσα και το άλλο σε κόμβο που ανήκει στην  $j$ -οστή.
- Ο πίνακας  $\mathbf{D}_r$  είναι πίνακας βαθμών στον νέο γράφο αν συμπεριλάβουμε και τις ανακυκλώσεις.
- Ο πίνακας  $\mathbf{L}_r$  είναι πίνακας *Laplacian* είτε τον σκεφτούμε στον γράφο με τις ανακυκλώσεις, είτε χωρίς αυτές. Γιατί ακόμη και αν προστεθούν οι ανακυκλώσεις στον  $\mathbf{D}_r$  θα αφαιρεθούν από τον πίνακα βαρών  $\mathbf{W}_r$ . Επίσης, επειδή ο  $\mathbf{B}^\perp$  είναι πίνακας διαμέρισης με άσσους στις στήλες ισχύει:  $\mathbf{B}^\perp \mathbf{1}_{N-C} = \mathbf{1}_N \Rightarrow \mathbf{L}_r \mathbf{1}_{N-C} = 0$ .
- Επομένως ο αριθμητής όπως έχουμε δει από το κεφάλαιο 2 για τις τετραγωνικές μορφές *Laplacian* πινάκων είναι:

$$(Y^i)^T \mathbf{L}_r Y^i = \sum_{(k,j): G_k \sim G_j} w_{kj}^{(r)} (Y_k^i - Y_j^i)^2 = \alpha_i^2 \sum_{G_j \in K_i, G_k \notin K_i} w_{ij}^{(r)} = \alpha_i^2 \cdot cut(K_i)$$

Άρα κάθε κλάσμα αντιπροσωπεύει τον λόγο  $\frac{cut(K_i)}{vol(K_i)}$  και στον αρχικό γράφο αν συνυπολογίσουμε τους περιορισμούς αλλά και στον αντίστοιχο *reduced* γράφο  $G_r$ , που κάθε κόμβος του είναι η αντίστοιχη συνεκτική συνιστώσα του γράφου περιορισμών και κάθε ακμή προς άλλο κόμβο είναι το άθροισμα των βαρών όλων των ακμών που έχουν ένα άκρο στην μια και ένα στην άλλη συν. συνιστώσα. Επίσης, ο  $G_r$  έχει ανακυκλώσεις σε κάθε κόμβο ίσες με το *association* κάθε συνεκτικής συνιστώσας του αρχικού γράφου.

Καταφέραμε και απαλείψαμε τον περιορισμό αναγάγοντας το πρόβλημα σε ένα ισοδύναμο για ένα μειωμένο γράφο. Αυτό είναι πιά λογικό από τις προτεινόμενες λύσεις εξαιτίας της φύσης του περιορισμού που επιβάλλαμε, γιατί όσο ομαδοποιούμε *pixel* θα πρέπει να περιμένουμε να κάνει λιγότερη δουλειά ο αλγόριθμος, όχι περισσότερη. Στην ακραία περίπτωση που οι περιορισμοί δημιουργούν μόνοι τους *K cluster* ο αλγόριθμος δεν έχει καθόλου φόρτο, πέρα από το *overhead* της κατασκευής του γράφου.

### Τελικός αλγόριθμος

1. Φτιάξε τον γράφο περιορισμών.
2. Βρες τις συνεκτικές συνιστώσες του γράφου, χρησιμοποιώντας κάποιο αλγόριθμο διάσχισης.
3. Κατασκεύασε τον *reduced* γράφο βάζοντας για κόμβους τις συνεκτικές συνιστώσες του βήματος 2) και για ακμές το άθροισμα των βαρών των ακμών του αρχικού γράφου που συνδέουν κόμβους από αυτές τις συνεκτικές συνιστώσες.
4. Πρόσθεσε στον γράφο και ακμές ανακύκλωσης με βάρη ίσα με το *association* κάθε συνεκτικής συνιστώσας. (τα βήματα 3,4 γίνονται με έναν αλγόριθμο διάσχισης στον κανονικό γράφο).
5. Εφάρμοσε τον αλγόριθμο του *NCut* σε αυτό το γράφο και κάνε την διακριτοποίηση του  $\mathbf{Y}$
6. Ό,τι κατάτμηση βρεις για τις συνεκτικές συνιστώσες επέχτεινέ την για την κατάτμηση των κόμβων του αρχικού γράφου.

Τελικά,

$$NCut(G, K) + MUST LINK CONSTRAINTS \iff NCut(G_r, K) \quad (4.8)$$

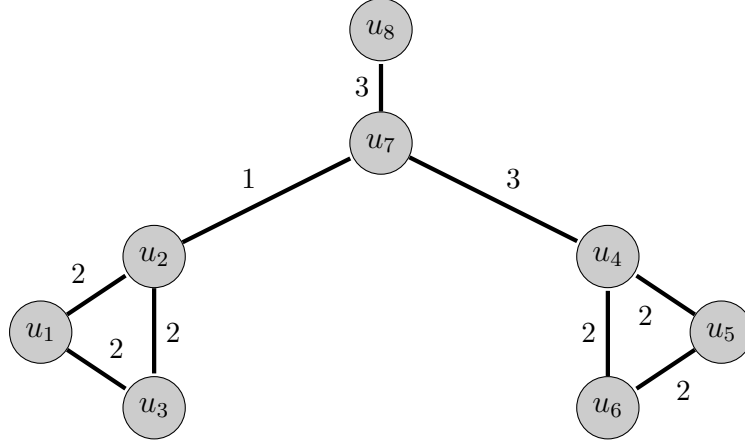
*Σημείωση:* Ο λόγος που στα [47],[48] δεν βρέθηκε αυτή η λύση είναι ότι οι συγγραφείς χρησιμοποιούσαν *reduced* αναπαράσταση (με 1 διάνυσμα για 2 κλάσεις), οπότε έπρεπε να εισάγουν και έναν ακόμη περιορισμό σχετικά με τις *non – trivial* λύσεις. Ωστόσο, προφανώς το αποτέλεσμα δεν εξαρτάται από την αναπαράσταση.

*Σημείωση 2:* Μέσα από την επίλυση αυτών των περιορισμών ενοποιούμε μεθόδους που επεμβαίναν απευθείας στον γράφο [21] με μεθόδους που περιορίζαν τον εφικτό χώρο λύσεων [50]



### A Toy Example: Οπτική ανάλυση των βημάτων του αλγορίθμου

Θα δείξουμε πως λειτουργεί ο αλγόριθμος μέσα από ένα απλό παράδειγμα. Έστω ο παρακάτω γράφος τον οποίο θέλουμε να χωρίσουμε σε 2 κλάσεις:



Οι πίνακες που περιγράφουν τον γράφο είναι:

$$\mathbf{W} = \begin{bmatrix} 0 & 2 & 2 & 0 & 0 & 0 & 0 & 0 \\ 2 & 0 & 2 & 0 & 0 & 0 & 1 & 0 \\ 2 & 2 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 2 & 2 & 3 & 0 \\ 0 & 0 & 0 & 2 & 0 & 2 & 0 & 0 \\ 0 & 0 & 0 & 2 & 2 & 0 & 0 & 0 \\ 0 & 1 & 0 & 3 & 0 & 0 & 0 & 3 \\ 0 & 0 & 0 & 0 & 0 & 0 & 3 & 0 \end{bmatrix}, \quad \mathbf{D} = \text{diag} \left( \begin{bmatrix} 4 \\ 5 \\ 4 \\ 7 \\ 4 \\ 4 \\ 7 \\ 3 \end{bmatrix} \right)$$

Όπως μπορούμε εύκολα να δούμε η βέλτιστη σταθμισμένη τομή είναι :

$$\{K_1, K_2\} = \{\{u_1, u_2, u_3\}, \{u_4, u_5, u_6, u_7, u_8\}\}$$

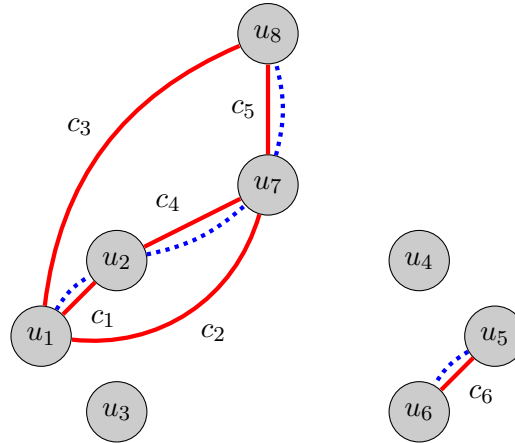
Και άρα ο πίνακας κατάτμησης είναι :

$$\mathbf{X} = \begin{bmatrix} \alpha_1 & 0 \\ \alpha_1 & 0 \\ \alpha_1 & 0 \\ 0 & \alpha_2 \\ 0 & \alpha_2 \\ 0 & \alpha_2 \\ 0 & \alpha_2 \\ 0 & \alpha_2 \end{bmatrix}$$

Ωστόσο, αν για κάποιο λόγο γνωρίζουμε ότι κάποιοι κόμβοι πρέπει να τοποθετηθούν στην ίδια κλάση, μπορούμε να προσθέσουμε *must link* περιορισμούς.

$$\text{Π.χ. } \Pi = \{(1,2),(1,7),(1,8),(2,7),(7,8),(5,6)\}$$

Άρα ο γράφος των περιορισμών είναι:



Σχήμα 4.3: Γράφος περιορισμών πριν (κόκκινο) και μετά (μπλε) την ελαχιστοποίηση

$$\mathbf{B}_\sigma^{\text{init}} = \begin{bmatrix} 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & -1 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & -1 \\ 0 & 1 & 0 & 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 \end{bmatrix}, \quad \mathbf{B}_\sigma^{\text{min}} = \begin{bmatrix} 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 \end{bmatrix}$$

Σχήμα 4.4: Ελαχιστοποίηση πίνακα περιορισμών.

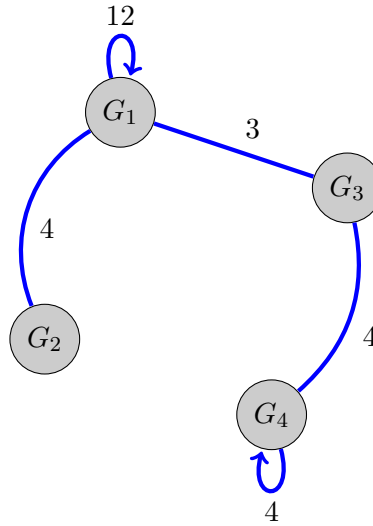
Παρατηρούμε ότι οι γραμμικά εξαρτημένες γραμμές του αρχικού πίνακα περιορισμών είναι 2, όσοι και οι απλοί κύκλοι του γράφου περιορισμών. Απαλείφοντας αυτές τις γραμμές μας μένει ο τελικός πίνακας περιορισμών που είναι ισοδύναμος με τον πρώτο και έχει βαθμό 4, όσοι οι γραμμικά ανεξάρτητοι περιορισμοί. Επίσης, οι συνεκτικές συνιστώσες του γράφου είναι  $8-4 = 4$ . Επομένως και η διάσταση του πυρήνα θα είναι 4.

$$\text{Μάλιστα, ο πίνακας του πυρήνα είναι (από τον } DFS \text{): } \mathbf{B}^\perp = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix}$$

Οπότε οι πίνακες που περιγράφουν το πρόβλημα στο *reduced space* είναι:

$$\mathbf{L}_r = \mathbf{B}^\perp \mathbf{T} \mathbf{L} \mathbf{B}^\perp = \begin{bmatrix} 7 & -4 & -3 & 0 \\ -4 & 4 & 0 & 0 \\ -3 & 0 & 7 & -4 \\ 0 & 0 & -4 & 4 \end{bmatrix}, \quad \mathbf{D}_r = \mathbf{B}^\perp \mathbf{T} \mathbf{D} \mathbf{B}^\perp = \begin{bmatrix} 19 & 0 & 0 & 0 \\ 0 & 4 & 0 & 0 \\ 0 & 0 & 7 & 0 \\ 0 & 0 & 0 & 8 \end{bmatrix}$$

Παρατηρούμε ότι αυτοί οι πίνακες αντιστοιχούν στον *Laplacian* και βαθμών αντίστοιχα, στον *reduced* γράφο του παρακάτω σχήματος:



Σχήμα 4.5: *Reduced* γράφος (ή γράφος συνεκτικών συνιστωσών)

Παρατηρούμε ότι οι κόμβοι μειώθηκαν κατά το πλήθος των γραμμικά ανεξαρτητων περιορισμών και ότι ο νέος γράφος έχει ως κόμβους τις συνεκτικές συνιστώσες του γράφου των περιορισμών. Τα βάρη των ακμών μεταξύ 2 κόμβων ισούνται με το άθροισμα των επιμέρους βαρών των ακμών που συνέδεαν κόμβους από τις 2 αυτές συνεκτικές συνιστώσες στον αρχικό γράφο. Τέλος, τα βάρη των ακμών ανακύκλωσης ισούνται με το *association* κάθε συνεκτικής συνιστώσας του γράφου περιορισμών.

Τώρα προφανώς η ελάχιστη τομή στο γράφο είναι :

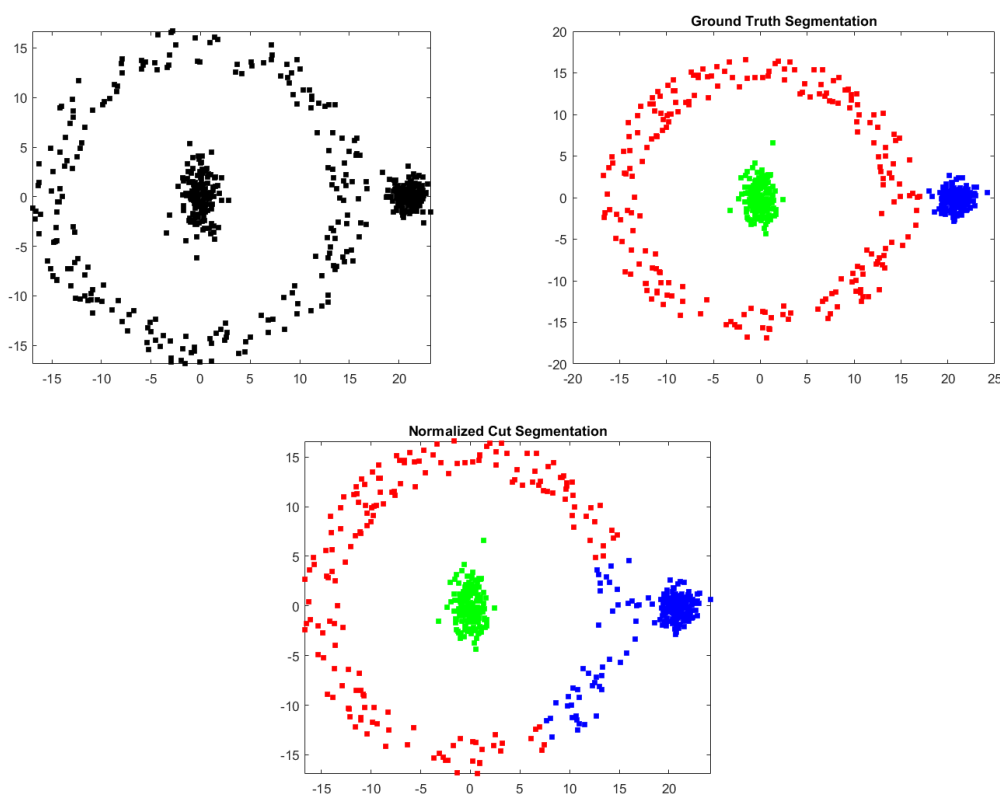
$$\{K_1, K_2\} = \{\{G_1, G_2\}, \{G_3, G_4\}\} = \{\{u_1, u_2, u_3, u_7, u_8\}, \{u_4, u_5, u_6\}\}$$

Ενώ οι πίνακες της λύσης είναι:

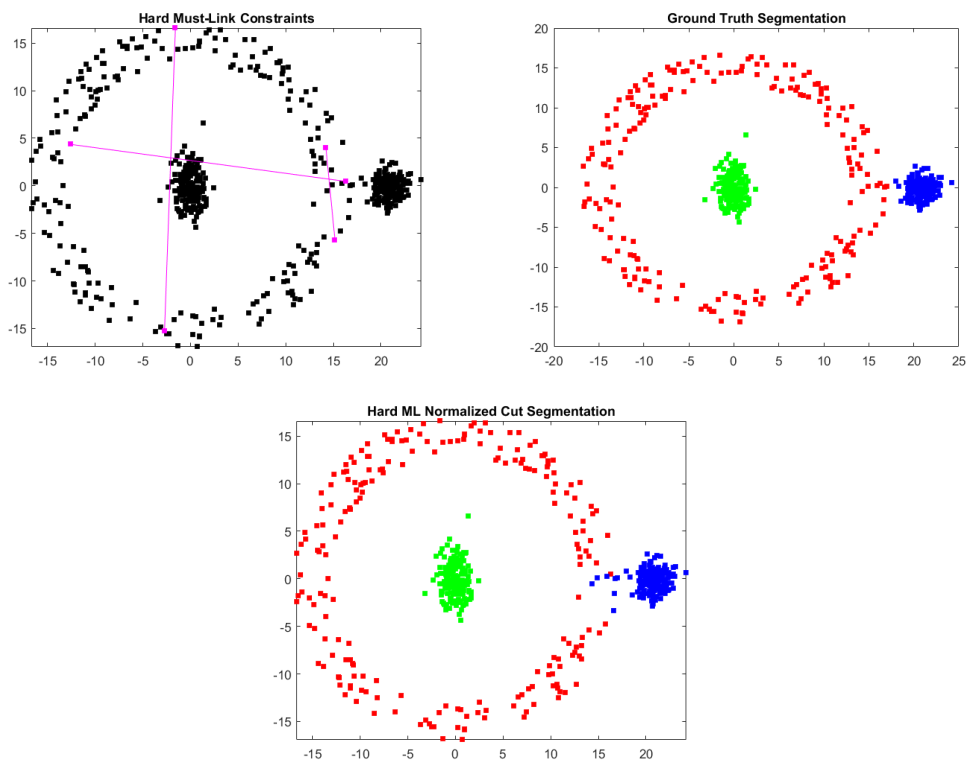
$$\mathbf{Y} = \begin{bmatrix} \alpha_1 & 0 \\ \alpha_1 & 0 \\ 0 & \alpha_2 \\ 0 & \alpha_2 \end{bmatrix} \xrightarrow{\mathbf{B}^\perp} \mathbf{X} = \begin{bmatrix} \alpha_1 & 0 \\ \alpha_1 & 0 \\ \alpha_1 & 0 \\ 0 & \alpha_2 \\ 0 & \alpha_2 \\ 0 & \alpha_2 \\ \alpha_1 & 0 \\ \alpha_1 & 0 \end{bmatrix}$$

### Data Clustering: Sampling Known Distributions

Θα χρησιμοποιήσουμε ένα συνθετικό παράδειγμα για να επιβεβαιώσουμε την αποτελεσματικότητα του αλγορίθμου μας, τόσο στην ποιότητα κατάτμησης όσο και στον χρόνο εκτέλεσης. Επίσης, θα δούμε εποπτικά πως βελτιώνεται η κατάτμηση με την αύξηση των  $ML$  περιορισμών. Στο παράδειγμά μας υποδειγματοληπούμε από 3 κατανομές, 2 γκαουσιανές και μια δακτυλίου, όπως φαίνεται στα σχήματα 4.6, 4.7.



Σχήμα 4.6: Πρώτο συνθετικό παράδειγμα. Βλέπουμε την αδυναμία του unsupervised αλγορίθμου να εντοπίσει το Ground Truth Segmentation

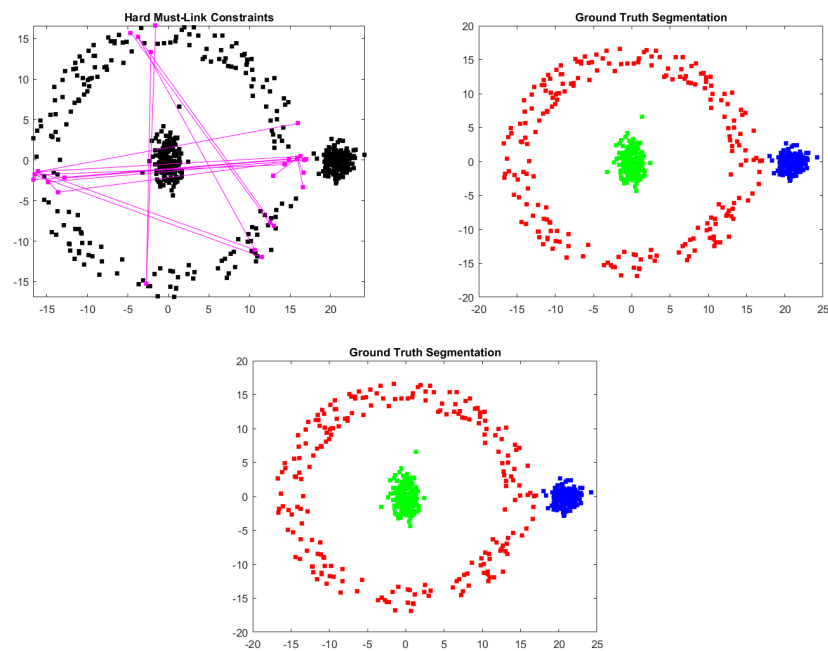


Σχήμα 4.7: Η αποτελεσματικότητα του αλγορίθμου φαίνεται στο γεγονός ότι μόνο 3 *ML* περιορισμοί βελτιώσαν αισθητά την κατάτμηση.

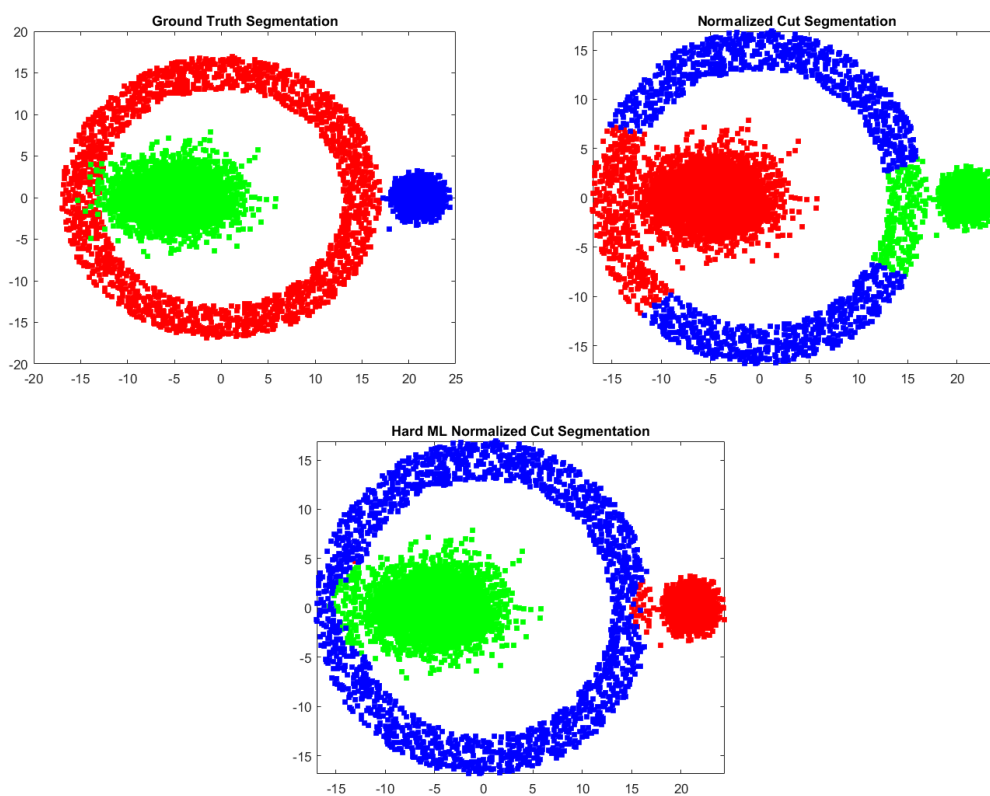
### Περιορισμοί και χρόνος. Περιορισμοί και αποτελεσματικότητα.

Είναι λογικό, όχι όμως πάντα αληθές, ότι όσο αυξάνουμε τους περιορισμούς μεγαλώνει η αποτελεσματικότητα του αλγορίθμου. Παραδείγματος χάρη στο επόμενο σχήμα βλέπουμε πως μόνο με 20 περιορισμούς, καλά μοιρασμένους στο χώρο, πετυχαίνουμε την αληθινή κατάτμηση.

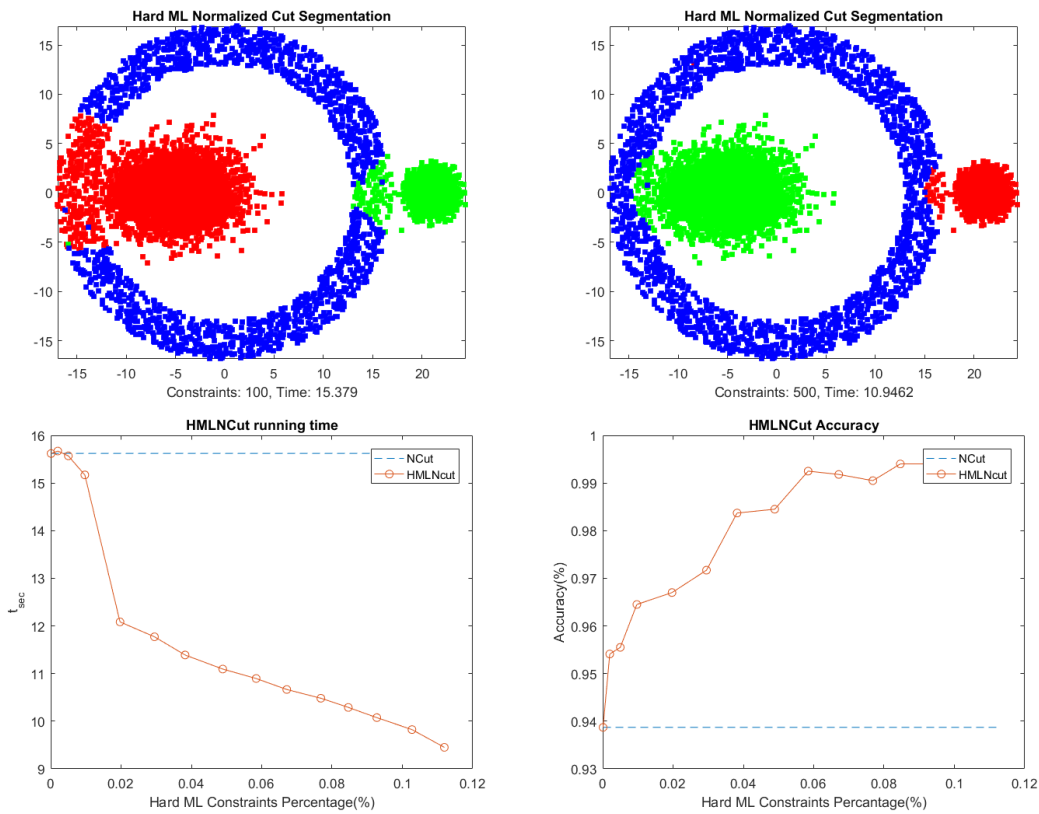
Θα δούμε τώρα πως αλλάζει η ποιότητα την κατάτμησης με την εισαγωγή περισσότερων περιορισμών. Γιαυτό το λόγο θα επεκτείνουμε την δεύτερη γκαουσιανή ώστε να δυσκολέψουμε την κατάτμηση και θα εισάγουμε τυχαία τους περιορισμούς. Επιπλέον, θα δούμε και πως αλλάζει ο χρόνος εκτέλεσης με την εισαγωγή μεγάλου αριθμού περιορισμών. Γιαυτό θα αυξήσουμε τα δείγματα από τις κατανομές.



Σχήμα 4.8: Με 20 περιορισμούς πετυχαίνουμε την πιο καλή κατάτμηση.



Σχήμα 4.9: Έχουμε επιβάλλει 200 τυχαίους  $ML$  περιορισμούς ενώ έχουμε 10000 δείγματα συνολικά (Δηλαδή υπάρχουν 10000-3 γραμμικά ανεξάρτητοι περιορισμοί).



Σχήμα 4.10: Επίδοση και χρόνος συναρτήσει του αριθμού των ανεξάρτητων περιορισμών. Τα ποσοστά έχουν υπολογιστεί με βάση τους μέγιστους ανεξάρτητους περιορισμούς που είναι  $(\#Κομβων) - (\#cluster)$ .

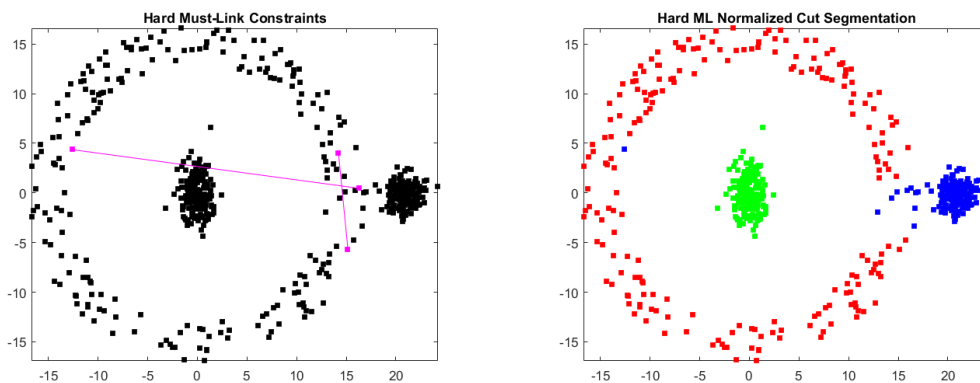
Στα πειράματα δημιουργήσαμε ανεξάρτητους τυχαίους  $ML$  περιορισμούς. Τα ποσοστά των  $ML$  περιορισμών έχουν υπολογιστεί ως προς τον μέγιστο αριθμό ανεξάρτητων περιορισμών που μπορούμε να έχουμε, που είναι όπως είδαμε  $size(Cluster) - 1$  για κάθε κλάση.

Στα σχήματα 4.10 επιβεβαιώνουμε την τεράστια επιτάχυνση που προκαλεί στον αλγόριθμο του  $NCut$  η εισαγωγή  $Hard ML$  περιορισμών. Όχι μόνο δεν καθυστερεί η προβολή στον εφικτό χώρο λύσεων αλλά ενσωματώνεται η πληροφορία φυσικά στο πρόβλημα, βοηθώντας την επίλυσή του. Παρατηρούμε επίσης πως μόνο με το 2% των περιορισμών η χρονική βελτίωση είναι πολύ μεγάλη. Τέλος, η μικρή αύξηση του χρόνου όταν υπάρχουν λίγοι περιορισμοί οφείλεται στο *overhead* που προκαλεί η σύντηξη των συνεχτικών συνιστωσών. Μάλιστα, με 10% περιορισμούς, παραπάνω από το μισό χρόνο εκτέλεσης παίρνει αυτό το *overhead*.

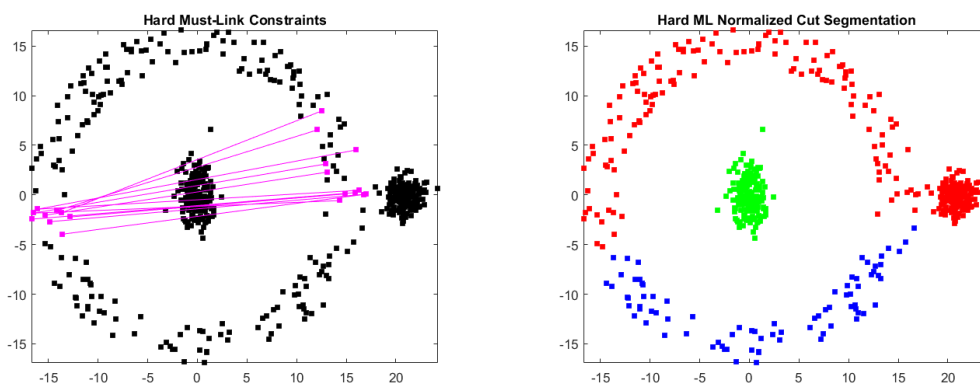
Ως προς την ποιότητα της κατάτμησης παρατηρούμε πως εν γένει αυξάνεται με την αύξηση των περιορισμών και μάλιστα με πολύ γρήγορους ρυθμούς, καθώς το 2% των περιορισμών προκαλεί αισθητή βελτίωση (97%). Ωστόσο κάτι που θα αναλύσουμε μετά είναι ότι δεν είναι πάντα βέβαιο ότι περισσότεροι  $HML$  περιορισμοί θα οδηγήσουν σε καλύτερη κατάτμηση.

### Πότε οι $ML$ περιορισμοί δεν είναι αρκετοί

Στην πράξη οι  $ML$  περιορισμοί είναι αραιοί στο σύνολο των δεδομένων. Σκεφτείτε π.χ. να δίνονται από τον χρήστη. Πολλές φορές λοιπόν, όταν είναι τόσο αραιοί μπορεί να μην συνεισφέρουν στην τελική κατάτμηση ή ακόμη και να την χειροτερεύουν. Κοιτάξτε π.χ. τα επόμενα σχήματα 4.11,4.12.



Σχήμα 4.11: Η κατάτμηση χειροτερεύει γιατί ο  $ML$  περιορισμός είναι πολύ αδύναμος από μόνος του για να τραβήξει όλα τα λάθος κατανομημένα δεδομένα στην πρώτη κλάση.



Σχήμα 4.12: Παρότι προσθέτουμε πολλούς  $ML$  περιορισμούς, επειδή δεν είναι στοχευμένοι τελικά η κατάτμηση χειροτερεύει.

Το πρώτο πρόβλημα οφείλεται στην αραιότητα των  $ML$  περιορισμών, και στην αδυναμία να συμπερασύρουν όλους τους λάθος κατατμημένους κόμβους. Μια λύση σε αυτό το πρόβλημα είναι με κάποιον τρόπο να διαδώσουμε την πληροφορία του  $ML$  περιορισμού και στους γείτονες των δύο κόμβων. Απαιτώντας μια "ομαλότητα" στην τελική τομή, τελικά καταφέρνουμε να δυναμώσουμε την ισχύ των περιορισμών αυτών παρότι μπορεί να είναι αραιοί. Αφήνουμε αυτό το πρόβλημα για μελλοντική ενασχόληση.

Από την άλλη βέβαια βλέπουμε πως και πληθώρα  $ML$  περιορισμών ίσως δεν δώσει βελ-



τιωμένα αποτελέσματα, αν αυτοί οι περιορισμοί δεν κρύβουν σημαντική πληροφορία. Ίδανικά, θα θέλαμε να δηλώσουμε με κάποιον τρόπο ότι οι λάθος ταξινομημένοι κόμβοι δεν μπορούν να ανήκουν μαζί με τους κόμβους του δεξιού cluster. Αυτό το φαινόμενο γεννάει την ανάγκη για ενσωμάτωση Cannot Link περιορισμών. Με αυτό το πρόβλημα θα ασχοληθούμε στην συνέχεια.

### Μελλοντικές επεκτάσεις και εφαρμογές

- Να μελετηθεί η *online* εκδοχή του περιορισμού. Δηλαδή, αν γίνει κατάτμηση και ξαφνικά ο χρήστης αποφασίσει να βελτιώσει το αποτέλεσμα ομαδοποιώντας 2 *pixel*, πως αλλάζει το πρόβλημα; Χρήσιμο για *interactive segmentation*. Αλλάζουμε στην ουσία τον γράφο επεμβαίνοντας άμεσα στην προηγούμενη κατάτμηση.
- Να μελετηθούν οι *soft* περιορισμοί. Το να μην ικανοποιούνται όλοι οι περιορισμοί είναι μερικές φορές χρήσιμο γιατί το *interactive segmentation* ενέχει λάθη από τον χρήστη. Από την άλλη το να δώσει ο χρήστης ή ένα αυτόματο υποσύστημα μια αβεβαιότητα για την ισχύ κάθε περιορισμού είναι επίσης χρήσιμο και δίνει ευελιξία στην διαδικασία της κατάτμησης. Επίσης, είναι πληροφορία που λαμβάνεται πιά εύκολα στην πράξη.

#### 4.2.2 Τοποθέτηση ενός κόμβου σε ένα συγκεκριμένο cluster (*Unary Constraints*)

Αυτή η εφαρμογή είναι αντίστοιχη του transductive learning κατά το οποίο γνωρίζοντας τα labels ενός συνόλου εκπαίδευσης και έχοντας το δικαίωμα να δούμε και το σύνολο των unlabeled δεδομένων, σχηματίζουμε υποθέσεις εκμεταλευόμενοι την δομή των τελευταίων βελτιώνοντας έτσι την τελική κατηγοριοποίηση τους.

Για εφαρμογές εικόνων μπορεί να γνωρίζουμε π.χ. ότι τα πολύ φωτεινά pixel ανήκουν στο foreground ενώ τα σκοτεινά στο background και θα θέλαμε να έχουμε την δυνατότητα να τοποθετήσουμε ένα ή περισσότερα pixel σε ένα συγκεκριμένο cluster. Άλλες εφαρμογές παραδείγματος χάρη θα ήταν ότι τα συγκεκριμένα cluster έχουν κάποια σημασία για μας και θα θέλαμε να τοποθετήσουμε έναν εκπρόσωπο σε αυτά αναγκάζοντας ταυτόχρονα τους κόμβους-εκπροσώπους να μην πέσουν στο ίδιο cluster (π.χ. ομαδοποίηση συγκεκριμένων information hubs σε ένα κοινωνικό δίκτυο).

#### Μοντελοποίηση Περιορισμού:

$$\text{κόμβος } u_i \in K_j \Leftrightarrow (X_j(i) = \alpha_j \neq 0 \wedge X_k(i) = 0, \forall k \neq j) \Leftrightarrow e_i^T \cdot \mathbf{X} = \alpha_j \cdot e_j^T \quad (1),$$

όπου τα  $e_i, e_j$  έχουν τις κατάλληλες διαστάσεις.

*Σημείωση:* Θα μπορούσαμε να γράψουμε και ότι  $\mathbf{X}_{ij} = \alpha_j \neq 0$  γιατί ξέρουμε ότι ο πίνακας

$\mathbf{X}$  έχει σε κάθε γραμμή μόνο ένα στοιχείο μη μηδενικό. Όμως όταν θα κάναμε το relaxation του αρχικού προβλήματος όπως είδαμε δεν θα ίσχυε πια αυτός ο περιορισμός για τον πίνακα  $\mathbf{X}$  και επομένως ούτε και ο νέος περιορισμός θα ήταν αρκετός.

Επειδή όμως δεν ξέρουμε την τιμή των  $\alpha_j$  (είναι μεταβλητές) και θέλουμε να διατηρήσουμε το scale-invariance των περιορισμών για να απαλύσουμε τους παρονομαστές από τα Rayleigh πηλίκα θα πρέπει να βρούμε ένα τρόπο να γράψουμε αλλιώς την εξίσωση. Ένας τέτοιος τρόπος θα ήταν π.χ.

$$\mathbf{X}^T \cdot e_i \cdot \underbrace{e_j^T \cdot \mathbf{X}^\dagger \cdot \mathbf{1}_N}_{\frac{1}{\alpha_j}} = e_j$$

Τελικά, αν είχαμε ένα σύνολο  $S$  από κόμβους περιορισμένους να είναι προσκολλημένοι σε συγκεκριμένα cluster τότε ο χώρος λύσεων του προβλήματος θα περιοριζόταν από ένα σύνολο εξισώσεων της μορφής:

$$\begin{bmatrix} \mathbf{X}^T & 0 & \dots & 0 \\ 0 & \mathbf{X}^T & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \mathbf{X}^T \end{bmatrix} \cdot \begin{bmatrix} \mathbf{e}_{i_1 j_1} & 0 & \dots & 0 \\ 0 & \mathbf{e}_{i_2 j_2} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \mathbf{e}_{i_{|S|} j_{|S|}} \end{bmatrix} \cdot \begin{bmatrix} \mathbf{X}^\dagger \cdot \mathbf{1}_N \\ \mathbf{X}^\dagger \cdot \mathbf{1}_N \\ \vdots \\ \mathbf{X}^\dagger \cdot \mathbf{1}_N \end{bmatrix} = \begin{bmatrix} \mathbf{e}_{j_1} \\ \mathbf{e}_{j_2} \\ \vdots \\ \mathbf{e}_{j_N} \end{bmatrix}$$

όπου το  $\mathbf{e}_{ij} = e_i \cdot e_j^T$ .

Επειδή ο περιορισμός είναι scale-invariant μπορεί να εφαρμοστεί στο αρχικό ή στο ισοδύναμο πρόβλημα. Δυστυχώς, παρότι στο relaxed πρόβλημα μπορεί να περάσει αυτούσιος γιατί ορίζεται ο ψευδοαντίστροφος  $\mathbf{X}^\dagger$ , δεν θα σήμαινε ακριβώς το ίδιο πράγμα και επίσης δεν είναι rotation invariant. Η δυσκολία του περιορισμού ξεκινάει από το γεγονός ότι πρέπει να απαλύσουμε τα άγνωστα σε εμάς  $\alpha_j$ . Θα προσπαθήσουμε να επιβάλλουμε την συνθήκη αυτή στο ισοδύναμο πρόβλημα, όπου επειδή δουλεύουμε σε ένα και μόνο projective plane δεν μας ενδιαφέρει το scale invariance.

Με την απαίτηση:  $(X^i)^T D X^i = 1$  ουσιαστικά περιορίζουμε τις τιμές των  $\alpha_i = \frac{1}{\sqrt{\sum_{i \in K_i} d_i}} = \frac{1}{\sqrt{\text{vol}(K_i)}}$ . Τώρα η εξίσωση (1) είναι πιό χρήσιμη γιατί τα  $\alpha_j$  δεν είναι μεταβλητές. Παρόλαυτά επειδή εμπλέκουν τον βαθμό κάθε cluster τελικά δεν μπορούμε να υπολογίσουμε τις τιμές τους πριν κάνουμε την κατάτμηση.

Από τα παραπάνω συμπεραίνουμε ότι το καλύτερο που θα μπορούσαμε να κάνουμε είναι να επιβάλλουμε ένα περιορισμό στο σχήμα του πίνακα  $\mathbf{Q}$  (ορίστηκε στο προηγούμενο κεφάλαιο ως  $sh: R^{N \times K} \rightarrow R^{N \times K}$  με τύπο:  $sh(\mathbf{X})_{ij} = (X_{ij} \neq 0)$ ). Επομένως, αν έχουμε ένα σύνολο εξισώσεων από κόμβους που δεσμεύουμε να ανήκουν σε συγκεκριμένα clusters, ο περιορισμός που θα επιβάλλαμε στο αρχικό πρόβλημα είναι:

$$\mathbf{E}_N^T * sh(\mathbf{X}) = \mathbf{E}_K^T$$

όπου  $\mathbf{E}_N = [e_{c1}, \dots, e_c]$  είναι ένας πίνακας με  $N$  γραμμές και στήλες όσοι οι περιορισμοί.

### Επίλυση Περιορισμού :

Επειδή  $sh(\mathbf{X}\mathbf{\Lambda}^{-1}) = sh(\mathbf{X})$  και  $sh(\mathbf{D}^{-1/2}\mathbf{Y}) = sh(\mathbf{Y})$  ο περιορισμός μένει αυτο-ύσιος και στο ισοδύναμο πρόβλημα και κατά το relaxation. Μπορούμε να επιβάλλουμε τον περιορισμό αλγοριθμικά κάνοντάς τον drop κατά την διαδικασία του relaxation και επιβάλλοντάς τον κατά την διαδικασία του rounding. Η φάση που αλλάζουμε τον πίνακα  $\mathbf{X}$  όπως είδαμε στο προηγούμενο κεφάλαιο είναι κρατώντας το πίνακα  $\mathbf{Q}$  σταθερό και κάνοντας *non-maximal suppression* στις τιμές του  $\mathbf{ZQ}$ . Θα μπορούσαμε λοιπόν να επιβάλλουμε την συνθήκη σε αυτή την φάση του αλγορίθμου, φροντίζοντας να ικανοποιείται στους επίμαχους κόμβους αντί να στρογγυλοποιούμε τυφλά στο μηδέν όλα όσα δεν είναι μέγιστα. Το ίδιο θα μπορούσαμε να κάνουμε και με ένα *rounding scheme* που θα χρησιμοποιούσε τον  $\kappa$ -μεανς. Κατά την διαδικασία της ανάθεσης θα εξαιρούσαμε τους κόμβους που είναι ήδη *labeled* ενώ κατά την διαδικασία της εύρεσης εκπροσώπων οι κόμβοι αυτοί θα συμμετείχαν. Επιπλέον, εξαιτίας του *partial labeling* έχουμε και ένα σημείο αρχικοποίησης του  $k$  - means.

### Ερμηνεία του περιορισμού:

Διαισθητικά περιμένουμε η ελάχιστη σταθμισμένη τομή να είναι περίπου εκείνη που προέρχεται από τον επαγόμενο υπογράφο (χωρίς κόμβους που είναι ήδη ταιριασμένοι με cluster) όμως με ένα bias για τους κόμβους εκείνους του επαγόμενου υπογράφου που συνδέονται πιο ισχυρά με ήδη ταιριασμένους κόμβους, να παραμείνουν και αυτοί στο ίδιο cluster.

Πράγματι αν  $\Pi$  είναι το σύνολο των  $C$  κόμβων που εμπλέκονται στον περιορισμό, τότε θεωρώντας χωρίς βλάβη της γενικότητας ότι αυτοί είναι οι πρώτοι  $C$  κόμβοι του γράφου μπορούμε να γράψουμε τους πίνακες ως:

$$\mathbf{X} = \begin{bmatrix} \mathbf{X}_k \\ \mathbf{X}_u \end{bmatrix} \text{ και } \mathbf{W} = \begin{bmatrix} \mathbf{W}_{11} & \mathbf{W}_{12} \\ \mathbf{W}_{12}^T & \mathbf{W}_{22} \end{bmatrix} \text{ όπου ο } \mathbf{X}_k \text{ είναι ο πίνακας που αποτελείται από}$$

τις πρώτες  $C$  γραμμές του  $\mathbf{X}$ . Αντί να δουλέψουμε με την ελάχιστη σταθμισμένη τομή, θα δουλέψουμε ισοδύναμα με το μέγιστο σταθμισμένο association που είχαμε δείξει σε προηγούμενο κεφάλαιο ότι είναι συμπληρωματικά. Ο λόγος είναι ότι υπάρχει πιο κατανοητή ερμηνεία με αυτό τον τρόπο επειδή κάθε cluster σπάει σε 2 μέρη (ένα με "προσδεμένους" κόμβους και ένα με άγνωστους) και οι νέες τομές που θα δημιουργούνταν θα ήταν πάρα πολλές. Άρα, το ισοδύναμο πρόβλημα βελτιστοποίησης γράφεται ως:

(αντικειμενική συνάρτηση):

$$\max_{\mathbf{X}} \text{tr} \left( \begin{pmatrix} \mathbf{X}_k^T & \mathbf{X}_u^T \\ \mathbf{X}^T & \end{pmatrix} \begin{pmatrix} \mathbf{W}_{11} & \mathbf{W}_{12} \\ \mathbf{W}_{12}^T & \mathbf{W}_{22} \end{pmatrix} \begin{pmatrix} \mathbf{X}_k \\ \mathbf{X}_u \\ \mathbf{X} \end{pmatrix} \right) =$$

$$\max \text{tr}(\mathbf{X}_k^T \mathbf{W}_{11} \mathbf{X}_k + \mathbf{X}_u^T \mathbf{W}_{12}^T \mathbf{X}_k + \mathbf{X}_k^T \mathbf{W}_{12} \mathbf{X}_u + \mathbf{X}_u^T \mathbf{W}_{22} \mathbf{X}_u)$$

Επίσης,

- $\text{tr}(\mathbf{X}_u^T \mathbf{W}_{12}^T \mathbf{X}_k + \mathbf{X}_k^T \mathbf{W}_{12} \mathbf{X}_u) = \text{tr}(\mathbf{X}_u^T \mathbf{W}_{12}^T \mathbf{X}_k) + \text{tr}(\mathbf{X}_k^T \mathbf{W}_{12} \mathbf{X}_u) = \text{tr}((\mathbf{X}_k^T \mathbf{W}_{12} \mathbf{X}_u)^T) + \text{tr}(\mathbf{X}_k^T \mathbf{W}_{12} \mathbf{X}_u) = 2 \cdot \text{tr}(\mathbf{X}_k^T \mathbf{W}_{12} \mathbf{X}_u)$

- ο πίνακας  $\mathbf{X}_k$  είναι γνωστός ως προς το shape του αλλά οι τιμές του (τα  $\alpha_j = \frac{1}{\text{vol}(K_j)}$ ) είναι άγνωστες μέχρι το τέλος της κατάτμησης. Επομένως και ο όρος αυτός δεν είναι ακριβώς σταθερός

η αντικειμενική συνάρτηση γίνεται:  $\max \text{tr}(2 \cdot \mathbf{X}_k^T \mathbf{W}_{12} \mathbf{X}_u + \mathbf{X}_u^T \mathbf{W}_{22} \mathbf{X}_u) + \text{tr}(\mathbf{X}_k^T \mathbf{W}_{11} \mathbf{X}_k)$

Όπως βλέπουμε καλούμαστε να ελαχιστοποιήσουμε έναν τετραγωνικό όρο που μοιάζει με το σταθμισμένο association στον επαγόμενο υπογράφο καθώς και ένα γραμμικό όρο που αποτελεί ένα bias για κόμβους που συνδέονται πολύ ισχυρά με "γειωμένους" κόμβους. Αναλύοντας περισσότερο τους όρους βρίσκουμε μετά από πράξεις ότι:

- $\text{tr}(\mathbf{X}_k^T \mathbf{W}_{11} \mathbf{X}_k) = \sum_{l \in \{1, \dots, K\}} \sum_{i, j \in K_l \cap \Pi} w_{ij} \alpha_l^2$
- $\text{tr}(\mathbf{X}_u^T \mathbf{W}_{22} \mathbf{X}_u) = \sum_{l \in \{1, \dots, K\}} \sum_{i, j \in K_l - \Pi} w_{ij} \alpha_l^2$
- $2 \cdot \text{tr}(\mathbf{X}_k^T \mathbf{W}_{12} \mathbf{X}_u) = 2 \sum_{l \in \{1, \dots, K\}} \sum_{\substack{i \in K_l \cap \Pi \\ j \in K_l - \Pi}} w_{ij} \alpha_l^2$

Παρατηρούμε ότι ο πρώτος όρος συνδέεται με το  $\text{assoc}(\Pi)$  του επαγόμενου υπογράφου που αποτελείται από τους δεσμευμένους κόμβους, ο δεύτερος όρος με το  $\text{assoc}(\Pi')$  του επαγόμενου υπογράφου που αποτελείται από τους αδέσμευτους κόμβους, ενώ ο τρίτος όρος προσπαθεί να μεγιστοποιήσει το association των cluster του συνολικού γράφου, κατευθύνοντας αδέσμευτους κόμβους να ενσωματωθούν στο ίδιο cluster με εκείνους τους δεσμευμένους κόμβους με τους οποίους συνδέονται πιά ισχυρά.

Ο δεύτερος περιορισμός που αφορά τους παρονομαστές του Rayleigh κριτηρίου σπάει σε:

$$(X_u^i)^T \mathbf{D}_{22} X_u^j = 0, i \neq j$$

και

$$(X_u^i)^T \mathbf{D}_{22} X_u^i = 1 - (X_k^i)^T \mathbf{D}_{11} X_k^i = 1 - \sum_{i \in \{K_i \cap \Pi\}} d_i a_i^2 = 1 - \frac{\sum_{i \in \{K_i \cap \Pi\}} d_i}{\text{vol}(K_i)}$$

ο τρίτος περιορισμός που ήταν  $\mathbf{X}\mathbf{X}^\dagger\mathbf{1}_N = \mathbf{1}_N$  αφορούσε το disjointness των clusters. Επειδή γνωρίζουμε ότι οι κόμβοι του  $\mathbf{X}_k$  τον ικανοποιούν αφού ανήκουν σε ένα, γνωστό cluster, μπορούμε να θεωρήσουμε ότι αρκεί να ικανοποιείται για τους υπόλοιπους κόμβους. Δηλαδή ο τρίτος περιορισμός γράφεται:  $\mathbf{X}_u\mathbf{X}_u^\dagger\mathbf{1}_{N-C} = \mathbf{1}_{N-C}$

Παρότι μπορούμε να επιβάλλουμε αλγοριθμικά τον περιορισμό αυτόν κατά το τελικό στάδιο όπως είδαμε πριν, θα μπορούσαμε αντί να τον χαλαρώσουμε να προσπαθήσουμε να λύσουμε τον πρόβλημα επιβάλλοντάς τον εξαρχής. Ο λόγος που το πρόβλημα γίνεται πολύ πιο δύσκολο από την μη-επιβλεπόμενη εκδοχή του είναι ότι τα  $vol(K_i)$  είναι άγνωστα μέχρι την τελική κατάτμηση αλλά παρόλαυτά εμφανίζονται μέσα από τον, κατά τα άλλα γνωστό, πίνακα  $\mathbf{X}_k$ , και στην αντικειμενική συνάρτηση και στον δεύτερο περιορισμό. Ωστόσο, κατά την χαλάρωση του προβλήματος ο περιορισμός ότι  $a_j = vol(K_i)^{-1/2}$  χάνεται, όπως επίσης και το γεγονός ότι τα στοιχεία της ίδιας στήλης του αρχικού πίνακα μπορούν να πάρουν μόνο 2 τιμές. Σε αυτό το σημείο λοιπόν μπορούμε να θεωρήσουμε ότι και ο πίνακας  $\mathbf{X}_k$  έχει ως τιμές κάποια πολλαπλάσια μιας γνωστής δείκτριας συνάρτησης, άγνωστα σε μας αλλά ανεξάρτητα από την τελική κατάτμηση. Θα θέλαμε ιδανικά να δούμε αν οι τιμές των πολλαπλάσιων επηρεάζουν την κατάτμηση. Σίγουρα δεν μπορούμε να θέσουμε οποιαδήποτε τιμή γιατί πρέπει να εξασφαλίσουμε ότι ο δεύτερος περιορισμός δεν γίνεται αδύνατος. Η Lagrangian του προβλήματος είναι: όπου μπορούμε να θεωρήσουμε ότι ο πίνακας  $\mathbf{\Lambda}$  είναι διαγώνιος γιατί στο σημείο που μεγιστοποιείται η Lagrangian, τα μη διαγώνια στοιχεία του  $\mathbf{\Lambda}$  έτσι κι αλλιώς δεν συμμετέχουν λόγω του δεύτερου περιορισμού. Ωστόσο, δεν πρέπει να ξεχνάμε ότι τα μη-διαγώνια στοιχεία μας έδιναν κατά την παραγωγή τον περιορισμό ότι οι στήλες πρέπει να είναι  $\mathbf{D}$ -ορθογώνιες. Αυτό πρέπει να το επιβάλλουμε στη λύση. Με άλλα λόγια,

$$\mathbf{X}_u^* = \arg \max_{(\mathbf{X}_u, \mathbf{\Lambda})} (\mathcal{L}) = \arg \max_{\substack{\mathbf{X}_u \text{ D-orthogonal columns} \\ \mathbf{\Lambda} \text{ diagonal}}} (\mathcal{L}).$$

Επομένως,

$$\left\{ \begin{array}{l} \nabla_{\mathbf{X}_u} \mathcal{L} = 0 \\ \nabla_{\mathbf{\Lambda}} \mathcal{L} = 0 \end{array} \right\} \Rightarrow \left\{ \begin{array}{l} \mathbf{W}_{22}\mathbf{X}_u = \mathbf{D}_{22}\mathbf{X}_u\mathbf{\Lambda} - \mathbf{W}_{21}\mathbf{X}_k \\ (\mathbf{X}_u)^T\mathbf{D}_{22}\mathbf{X}_u = \mathbf{I} - (\mathbf{X}_k)^T\mathbf{D}_{11}\mathbf{X}_k \end{array} \right\}$$

Η ανά στήλη (cluster),

$$\left\{ \begin{array}{l} \mathbf{W}_{22}X_u^{(i)} = \lambda_i\mathbf{D}_{22}X_u^{(i)} - \mathbf{W}_{21}X_k^{(i)} \\ (X_u^{(i)})^T\mathbf{D}_{22}X_u^{(i)} = 1 - (X_k^{(i)})^T\mathbf{D}_{11}X_k^{(i)} \\ (X_u^{(i)})^T\mathbf{D}_{22}X_u^{(j)} = 0, \forall j \neq i \end{array} \right\}$$

Μοιάζει με πρόβλημα ιδιοτιμών. Στην πραγματικότητα το  $\lambda_i$  μπορεί να είναι ή όχι ιδιοτιμή του πίνακα  $\mathbf{D}_{22}^{-1}\mathbf{W}_{22}$ . Αν δεν είχαμε τον περιορισμό της καθετότητας αποδεικνύεται ότι είναι μεγαλύτερο ή ίσο από την μεγαλύτερη ιδιοτιμή του πίνακα αυτού. Η δυσκολία είναι να επιβάλλει κανείς την καθετότητα μεταξύ των διαφορετικών κλάσεων.

Σύμφωνα με την προηγούμενη ανάλυση που κάναμε για  $ML$  περιορισμούς, μπορούμε χωρίς βλάβη της γενικότητας να υποθέσουμε ότι υπάρχει το πολύ ένας κόμβος δεσμευμένος για κάθε κλάση. Αυτό γιατί περισσότεροι κόμβοι δεσμευμένοι στην ίδια κλάση μπορούν να συμπτυχθούν σε έναν με κατάλληλο βάρος ανακύκλωσης. Αυτό όμως σημαίνει ότι για κάθε κλάση θα έχουμε το πολύ 1 άσσο στο διάνυσμα  $X_k^{(i)}$  το οποίο τώρα έχει μέγεθος το πολύ  $K \times 1$ . Επίσης, έτσι το πρόβλημα μπορεί να το δει κανείς για  $K=2$  και σαν s-t normalized cut, ενώ για  $K > 2$  σαν την γενίκευση αυτού.

### Περίπτωση $K=2$

Μπορούμε να δουλέψουμε και με την *reduced* αναπαράσταση των διανυσμάτων. Εξαιτίας της  $ML$  ανάλυσης που κάναμε προηγουμένως, οι πιθανές περιπτώσεις είναι:  $X_k = [a]$ ,  $X_k = [b]$ ,  $X_k = [a, b]^T$  εκ των οποίων οι δύο πρώτες δεν αλλάζουν από την *unsupervised* εκδοχή του NCut γιατί για κάθε λύση  $X$ , η  $-X$  είναι επίσης αποδεκτή, και συνεπώς πάντα μπορούμε να τοποθετήσουμε ένα μοναδικό κόμβο στην κλάση που θέλουμε. Η τελευταία όμως, στην οποία θα δουλέψουμε, κωδικοποιεί και *Cannot - Link* περιορισμούς μεταξύ των δύο κόμβων.

Σπάζοντας και πάλι το διάνυσμα σε 2 κομμάτια, το γνωστό και το άγνωστο, καταλήγουμε μετά από πράξεις στο πρόβλημα βελτιστοποίησης:

$$\begin{aligned} \min_{X_u} \quad & \frac{X_u^T L_{22} X_u + 2(L_{12}^T X_k)^T X_u + X_k^T L_{11} X_k}{X_u^T D_{22} X_u + X_k^T D_{11} X_k} \\ \text{s.t.} \quad & X_u^T D_{22} 1_{N-2} = -X_k^T [d_1, d_2]^T, \quad X_u \in \mathcal{X}, \end{aligned} \quad (4.9)$$

Με γνωστά τα  $a, b$  ο πρώτος περιορισμός είναι ένας γραμμικός περιορισμός στο  $X_u$ . Γνωρίζουμε από το αρχικό πρόβλημα πως πάντα πρέπει να ισχύει:

$$a * \text{vol}(A) + b * \text{vol}(A') = 0.$$

Σε αυτό το σημείο μπορούμε να ακολουθήσουμε 2 δρόμους. Να λύσουμε το αρχικό ή να λύσουμε το ισοδύναμο. Να σημειωθεί ότι μόνο για το  $X_u$  το πρόβλημα δεν είναι scale invariant. Για το ζεύγος  $(X_k, X_u)$  μπορούμε να δουλέψουμε στο ισοδύναμο πρόβλημα, όμως θεωρώντας γνωστές τις τιμές του  $X_k$ , δεσμεύουμε και το άγνωστο μέρος του διανύσματος να λαμβάνει αυτές τις τιμές, διαφορετικά το πρόβλημα είναι *approximation* του αρχικού. Πάντως αν για το ισοδύναμο ισχύει ότι:  $X^T D X = \text{vol}(G)$  αυτό συνεπάγεται πως:  $a*b = -1$  (και πίο συγκεκριμένα πως  $a = \sqrt{\mu}$ , όπου  $\mu = \frac{\text{vol}(A')}{\text{vol}(A)}$ ).

Προφανώς δεν γνωρίζουμε τις πραγματικές τιμές των σταθερών, αφού εξαρτώνται από την κατάτμηση, αλλά μπορούμε να τις εκτιμήσουμε. Ένας απλός τρόπος θα ήταν:

- Αν  $d_1 + d_2 \simeq \text{vol}(G) \Rightarrow a = \sqrt{\frac{d_2}{d_1}}$ .
- Αν  $d_1 + d_2 \ll \text{vol}(G)$  και οι 2 κόμβοι τοποθετούνται, στο *unsupervised normalized cut*, σε διαφορετικές κλάσεις  $\Rightarrow a = a_U$ , όπου  $a_U$  είναι η τιμή από την *unsupervised κατάτμηση*.

- Αν  $d_1 + d_2 \ll \text{vol}(G)$  και οι 2 κόμβοι τοποθετούνται στο unsupervised normalized cut στην κλάση A ( $A'$ )  $\Rightarrow a = \sqrt{\frac{\text{vol}_U(A') + d_2}{\text{vol}_U(A) - d_2}}$  ( $a = \sqrt{\frac{\text{vol}_U(A') - d_1}{\text{vol}_U(A) + d_1}}$ ), όπου  $\text{vol}_U(A)$ ,  $\text{vol}_U(A')$  είναι οι τιμές από την unsupervised κατάτμηση.

Το πρόβλημα τώρα έχει πάρει την μορφή που λύνουν οι συγγραφείς του [10] αν κάνουμε *relaxation* αφήνοντας το διάνυσμα να παίρνει πραγματικές τιμές και αν αλλάξουμε τα  $L$ ,  $\min$  με  $W$ ,  $\max$ . Οι συγγραφείς χρησιμοποιούν το πρόβλημα αυτό σε εφαρμογές Balanced Graph Matching. Η προτεινόμενη λύση δημιουργεί ένα επαυξημένο διάνυσμα  $X_a = \begin{bmatrix} X_u \\ 1 \end{bmatrix}$ .

Για το πρόβλημά μας:

$$\bar{\mathbf{W}} = \begin{bmatrix} \mathbf{W}_{22} & \mathbf{W}_{12}^T X_k \\ (\mathbf{W}_{12}^T X_k)^T & X_k^T \mathbf{W}_{11} X_k \end{bmatrix}, \quad \bar{\mathbf{D}} = \begin{bmatrix} \mathbf{D}_{22} & \mathbf{0} \\ \mathbf{0} & X_k^T \mathbf{D}_{11} X_k \end{bmatrix},$$

$$\bar{\mathbf{C}} = \begin{bmatrix} (\mathbf{D}_{22} \mathbf{1}_{N-2})^T & X_k^T \mathbf{D}_{11} \mathbf{1}_2 \end{bmatrix}$$

Το πρόβλημα βελτιστοποίησης επόμενως γράφεται:

$$\begin{aligned} \max_{X_a} \quad & \frac{X_a^T \bar{\mathbf{W}} X_a}{X_a^T \bar{\mathbf{D}} X_a} \\ \text{s.t.} \quad & \bar{\mathbf{C}} X_a = 0 \end{aligned} \quad (4.10)$$

Επειδή το πρόβλημα είναι scale-invariant, αν η βέλτιστη λύση για το (4.10) είναι  $X_a = [X_u^* \ t]^T$  τότε και η λύση  $[X_u^*/t \ 1]^T$  είναι επίσης βέλτιστη για το (4.10) και άρα η  $X_u^*/t$  είναι βέλτιστη για την (4.9).

Αποδεικνύεται στο [10] με χρήση πολλαπλασιαστών *Lagrange* ότι η βέλτιστη λύση είναι το ιδιοδιάνυσμα που αντιστοιχεί στην μεγαλύτερη ιδιοτιμή του πίνακα:  $\mathbf{P}_C \mathbf{W}' \mathbf{P}_C$ , όπου:

- $\mathbf{W}' = \bar{\mathbf{D}}^{-1/2} \bar{\mathbf{W}} \bar{\mathbf{D}}^{-1/2}$
- $\mathbf{P}_C = \mathbf{I} - \mathbf{C}'^T (\mathbf{C}' \mathbf{C}'^T)^{-1} \mathbf{C}'$
- $\mathbf{C}' = \bar{\mathbf{C}} \bar{\mathbf{D}}^{-1/2}$

Προκειμένου να επιταχύνουμε τον αλγόριθμο για την ειδική περίπτωσή μας, μπορούμε να βρούμε από πριν την βάση του πυρήνα του  $\bar{\mathbf{C}}$ , αντί να χρησιμοποιήσουμε τον πίνακα προβολής. Αυτό γιατί ο πίνακας  $\mathbf{P}_C$  παίρνει κυβικό χρόνο για να υπολογιστεί και επίσης, μπορεί να είναι πολύ πυκνός και να επιβραδύνει πολύ την λύση. Χρησιμοποιώντας ως κίνητρο το προηγούμενο πρόβλημα στο οποίο συνδέσαμε τον πυρήνα με ένα γράφο μπορούμε να σκεφτούμε πως κάθε διάνυσμα του πυρήνα πρέπει να έχει τουλάχιστον δύο μη μηδενικά στοιχεία (αλλιώς θα ήταν το μηδενικό διάνυσμα) και να ψάξουμε να βρούμε την πιο αραιή δυνατή αναπαράσταση για τα διανύσματα της βάσης του πυρήνα, δηλ. διανύσματα με ακριβώς 2 στοιχεία. Αρχικά μπορούμε να επιλέξουμε ποιές θέσεις θα αφήσουμε μη-μηδενικές. Κατόπιν, αν αυτές οι θέσεις είναι οι  $i, j$ , οι τιμές αυτών θα είναι  $\bar{C}_j$  και  $-\bar{C}_i$  ή  $-\bar{C}_j$  και  $\bar{C}_i$  αντίστοιχα, όπου  $\bar{\mathbf{C}} = [\bar{C}_1, \dots, \bar{C}_n]$ . Αυτά τα ζευγάρια όμως είναι  $n * (n - 1)/2$  ενώ εμείς

θα θέλαμε  $n - 1$  μόνο εξάυτών τα οποία να είναι γραμμικά ανεξάρτητα. Αν πάρουμε ένα γράφημα  $n$  κόμβων και τοποθετήσουμε μια ακμή  $\{i, j\}$  για κάθε διάνυσμα που έχει τις  $i, j$  θέσεις του μη μηδενικές τότε το ζητούμενο είναι ισοδύναμο με το να βρούμε ένα συνεκτικό δέντρο σε αυτόν τον γράφο (όταν  $a * d_1 + b * d_2 \neq 0$ ).

Η απόδειξη είναι οφείλεται σε δύο παρατηρήσεις. Η πρώτη είναι ότι θέλουμε ακριβώς  $n - 1$  ακμές από τον γράφο για να βρούμε τα διανύσματα βάσης και η δεύτερη ότι τα διανύσματα που αντιστοιχούν σε ένα κύκλο στο γράφο είναι γραμμικά εξαρτημένα. Το τελευταίο ισχύει γιατί κάθε κόμβος στον κύκλο εμφανίζεται σε δύο ακμές. Αν κάνουμε τον κύκλο κατευθυνόμενο τότε εμφανίζεται σε μια εξερχόμενη και σε μια εισερχόμενη και αν θεωρήσουμε ότι η ακμή  $(i, j)$  θα έχει τιμές  $\bar{C}_j$  και  $-\bar{C}_i$  στις θέσεις  $i, j$  αντίστοιχα, τότε κάθε κόμβος θα εμφανίζεται στα διανύσματα με μη μηδενική τιμή ακριβώς δύο φορές, μια με μείον την τιμή του προηγούμενου γείτονά του και μια με συν την τιμή του επόμενου γείτονά του. Αν επιπλέον πολλαπλασιάσουμε κάθε διάνυσμα με το γινόμενο των τιμών του  $\bar{C}$  που ανήκουν στον κύκλο εκτός από τις δύο μη μηδενικές τιμές του διανύσματος αυτού, τότε ο κόμβος  $i$  εμφανίζεται σε δύο διανύσματα με μη μηδενικές τιμές και αυτές είναι αντίθετες γιατί είναι το γινόμενο όλων των τιμών του  $\bar{C}$  που ανήκουν στον κύκλο εκτός από την τιμή  $\bar{C}_i$ . Επομένως, προσθέτοντας όλα τα διανύσματα αυτά παίρνουμε το μηδενικό διάνυσμα. Επίσης, τουλάχιστον ένας συντελεστής σε αυτούς είναι μη μηδενικός γιατί όλα τα  $d_i > 0$  και μόνο το  $a * d_1 + b * d_2$  μπορεί να είναι μηδενικό, αλλά ακόμη και να είναι μηδενικό, είτε δεν θα ανήκει στον κύκλο αυτός ο κόμβος, είτε οι συντελεστές των δύο ακμών που έχουν τον κόμβο αυτόν δεν θα περιλαμβάνουν στο γινόμενο τον όρο αυτόν. Άρα, τα διανύσματα είναι γραμμικά εξαρτημένα. Σπάζοντας όλους τους κύκλους καταλήγουμε με ένα συνεκτικό δέντρο  $n - 1$  ακμών.

Αυτό που λείπει από την απόδειξη είναι ότι και κάθε δέντρο είναι αποδεκτή βάση του πυρήνα (ώστε να σπάμε τυχαία τους κύκλους). Αυτό δυστυχώς δεν ισχύει όταν  $a * d_1 + b * d_2 = 0$ . Π.χ. αν πάρει κανείς το δέντρο από τον τελευταίο κόμβο προς όλους τους άλλους προκύπτει ο πίνακας  $Q = [(a * d_1 + b * d_2) \mathbf{I}_{N-2} \quad d_{3:N}]^T$ , που δεν είναι αποδεκτή βάση γιατί έχει βαθμό 1. Πάντα όμως ο πίνακας που προκύπτει από τον κόμβο 1 προς όλους τους άλλους ήτοι

$$Q = \begin{bmatrix} d_4 & & \\ \vdots & & -d_3 \mathbf{I}_{N-2} \\ a * d_1 + b * d_2 & & \end{bmatrix}^T,$$

είναι βάση του πυρήνα.

Τελικά, και η βάση κατασκευάζεται σε γραμμικό χρόνο και (επειδή είναι πολύ αραιή) οι μετέπειτα πολλαπλασιασμοί παίρνουν τετραγωνικό χρόνο. Το μοναδικό πρόβλημα είναι ότι μόλις θέσουμε  $\mathbf{X} = \mathbf{Q}z$  θα κληθούμε να λύσουμε το γενικευμένο πρόβλημα ιδιοτιμών  $(\mathbf{Q}^T \mathbf{W} \mathbf{Q}, \mathbf{Q}^T \mathbf{D} \mathbf{Q})$  και οι τελικοί πίνακες δεν θα είναι πια αραιοί. Στο μέλλον θα δοκιμάσουμε εναλλακτικές τεχνικές για περαιτέρω επιτάχυνση της λύσης.



### 4.2.3 Τοποθέτηση 2 κόμβων σε 2 διαφορετικά clusters. (*Cannot Link Constraint*)

Σε εφαρμογές που είναι χρήσιμο να διαφοροποιήσουμε κόμβους όσο και αν μοιάζουν μεταξύ τους (π.χ. τους θέλουμε να είναι εκπρόσωποι των κλάσεών τους) ή επειδή θέλουμε να δώσουμε μεγαλύτερη ευελιξία στο interactive segmentation είναι πολύ χρήσιμο να ενσωματώσουμε στο πρόβλημα *cannot-link* περιορισμούς. Αυτοί οι περιορισμοί έχουν αποδειχθεί πολύ χρήσιμοι σε εφαρμογές που χρησιμοποιούν άλλους αλγορίθμους κατάτμησης όπως ο *k-means*. Στο πρόβλημα του *Normalized Cut* αυτοί οι περιορισμοί δύσκολα επιβάλλονται επακριβώς. Οι κύριοι λόγοι είναι:

- Η μη-μεταβατικότητα της σχέσης *cannot – link*.
- Η τετραγωνική φύση κάθε *CL* περιορισμού. Για κάθε κλάση που μπορεί να ανήκει ο ένας κόμβος υπάρχουν  $N-1$  πιθανές κλάσεις για τον δεύτερο κόμβο.

Μπορούμε όπως είπαμε και πριν να πάρουμε δύο δρόμους. Είτε να επέμβουμε κατευθείαν στον γράφο, να ενσωματώσουμε τους περιορισμούς και να λύσουμε το μη-επιβλεπόμενο πρόβλημα, είτε να θέσουμε τους περιορισμούς στο πρόβλημα βετιστοποίησης συρρικνώνοντας έτσι τον εφικτό χώρο λύσεων. Η δεύτερη μέθοδος είναι πιο καλά θεωρητικά θεμελιωμένη αλλά στην βιβλιογραφία οι *CL* περιορισμοί πάντα καταλήγουν να λύνονται προσεγγιστικά. Γιαυτό το λόγο αξίζει κανείς να κοιτάξει και τις πρώτες ώστε να βρει έναν τρόπο να ερμηνεύσει τις αλλαγές στον γράφο σαν αλλαγές στο πρόβλημα βελτιστοποίησης.

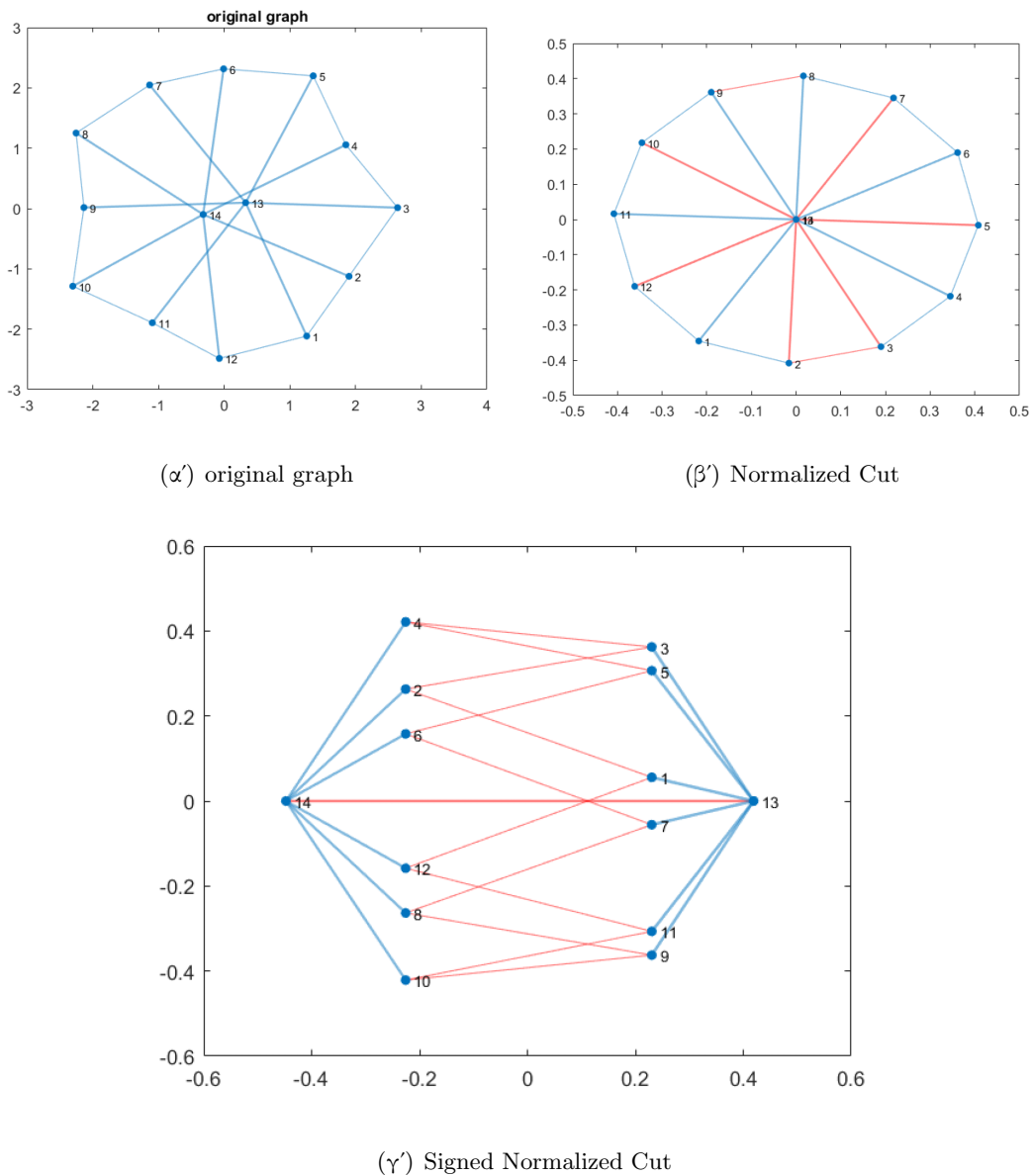
#### Μια πρώτη λύση: *Signed Normalized Cuts*

Ένα από τα πρώτα papers που αναφέρθηκαν σε semi-supervised spectral clustering είναι αυτό των D.Kamvar et al. ([21]) στο οποίο οι συγγραφείς επεμβαίνουν κατευθείαν στον σταθμισμένο πίνακα γειτνίασης θέτοντας  $A_{ij} = 1$  για κάθε *ML* περιορισμό μεταξύ των κόμβων  $i, j$  και  $A_{ij} = 0$  για κάθε *CL* περιορισμό μεταξύ αυτών των κόμβων. Παρότι με αυτό το τρόπο αποδυναμώνουν ισχυρές ακμές μεταξύ κόμβων που επιβάλλουμε να μην ανήκουν στο ίδιο cluster, δεν εξασφαλίζουν την τήρηση των περιορισμών. Μάλιστα, επειδή συχνά οι γράφοι κατασκευάζονται χρησιμοποιώντας τοπική πληροφορία μεταξύ των κόμβων, είναι πολύ συχνό να υπάρχουν πολλά ισχυρά μονοπάτια που συνδέουν δύο ισχυρά συνδεδεμένους κόμβους. Επομένως, το να κοπεί η απευθείας σύνδεση τους δεν αρκεί για να μην τοποθετηθούν οι κόμβοι στην ίδια κλάση. Αυτός είναι και ο λόγος που η μέθοδος αυτή δεν έχει καλά αποτελέσματα σε κατάτμηση εικόνων. Μια βελτιωμένη εκδοχή είναι αυτή των Z.Lu et al. ([24]) που μέσα από μια *Gaussian* διαδικασία ερμηνεύουν έναν νέο πίνακα γειτνίασης ο οποίος μάλιστα, έχει και αρνητικά βάρη στους *CL* περιορισμούς, και διαδίδουν αυτά τα βάρη και έξω από την ακμή  $\{i, j\}$  επηρεάζοντας και τα περαιτέρω μονοπάτια μεταξύ αυτών των κόμβων. Βέβαια, τελικά και εδώ οι συγγραφείς μηδενίζουν τα αρνητικά βάρη γιατί δεν μπορεί να τα διαχειριστεί ο αλγόριθμος του *Normalized Cut* αλλά τα αποτελέσματα είναι πάλι καλύτερα γιατί έχει γίνει το *affinity propagation*.

Τα αρνητικά βάρη που προέκυψαν μέσα από την *Gaussian* ερμηνεία των περιορισμών δίνουν μια διαίσθηση του πως πρέπει να αλλάξει ο γράφος όταν έχουμε τέτοιους περιορισμούς. Δείχνουν μια ανομοιότητα μεταξύ των κόμβων πολύ πιο ισχυρή από την μη ύπαρξη απευθείας μονοπατιού μεταξύ τους και ως τέτοια πρέπει να την εκμεταλλευτούμε. Για να τις ενσωματώσουμε όμως στο πρόβλημα πρέπει να αλλάξουμε την μετρική. Μια λογική μετρική που έχει χρησιμοποιηθεί στην βιβλιογραφία ([36], [13], [23]) είναι το Signed Normalized Cut γιατί ικανοποιεί την απαίτηση κόμβοι με αρνητικές ακμές μεταξύ τους να ανήκουν σε διαφορετικές κλάσεις αλλά η συνολική τομή να είναι ελάχιστη. Επιπλέον, αν δεν υπάρχουν αρνητικά βάρη καταλήγει στο γνωστό *Normalized Cut*. Τέλος, διαχειρίζεται και κόμβους με αρνητικό βαθμό γιατί αθροίζει σαν θετικές τις ακμές στην τομή αλλά τιμωρεί τους κόμβους αν καταλήξουν στην ίδια κλάση με θετικό πρόστιμο. Αντίθετα άλλες μετρικές μπορεί να έχουν ένα bias στο να απομονώσουν τον κόμβο με τον αρνητικό βαθμό. Θυμίζουμε ότι η μετρική που θέλουμε να ελαχιστοποιήσουμε είναι:

$$sNCut(A_1, \dots, A_K) = \sum_{j=1}^K \frac{cut(A_j, \bar{A}_j)}{vol(A_j)} + 2 \sum_{j=1}^K \frac{links^-(A_j, A_j)}{vol(A_j)}$$

Ο τρόπος να ελαχιστοποιήσουμε αυτή την τομή είναι να βρούμε τα  $K$  πρώτα ιδιοδιανύσματα του γενικευμένου προβλήματος ιδιοτιμών των  $(\bar{L}, \bar{D})$  όπου  $\bar{d}_i = \sum_{j=1}^m |w_{ij}|$ . Περισσότερη ανάλυση για το *signed normalized cuts* θα δοθεί στο επόμενο κεφάλαιο. Στην παρακάτω εικόνα 4.13 παρουσιάζουμε την επίδραση που έχουν οι *CL* περιορισμοί εφαρμοζόμενοι ως αρνητικές ακμές επί του γράφου, στην κατάτμηση. Στο πρώτο σχήμα φαίνεται ο γράφος που αποτελείται από δύο *interlaced stars*, που έχουν συνδέσεις στα φύλλα τους. Οι ακμές των αστεριών είναι βάρους 2 ενώ των φύλλων βάρους 1. Στο δεύτερο σχήμα φαίνεται η τομή που θα πραγματοποιούσε ο αλγόριθμος του *Normalized Cuts*, στον γράφο αυτό, όπου ο γράφος απεικονίζεται τώρα στο χώρο των ιδιοδιανυσμάτων για εποπτικούς λόγους. Η τομή βρίσκεται από τα πρόσημα των κορυφών στα άκρα κάθε ακμής. Παρατηρούμε ότι η τελική κατάτμηση δεν είναι βέλτιστη με την έννοια του *Normalized Cuts* κριτηρίου και ούτε διαχωρίζει σωστά τις κλάσεις, αφού ο αλγόριθμος προτιμάει να κόψει τις μισές ακμές κάθε αστεριού και δύο συνδέσεις από τα φύλλα. Ο λόγος είναι ότι οι δύο κορυφές {13, 14} των αστεριών, λόγω των ελκτικών δυνάμεων από κάθε κορυφή με την οποία είναι συνδεδεμένες καταλήγουν μαζί στην αρχή των αξόνων λόγω συμμετρίας. Παρότι λοιπόν δεν υπάρχει άμεση σχέση ισχυρής ομοιότητας μεταξύ αυτών των κόμβων, καταλήγουν πολύ κοντά στο χώρο των ιδιοδιανυσμάτων, artifact που οφείλεται στο συγκεκριμένο embedding των κορυφών. Χρειάζεται λοιπόν, να απεικονίσουμε ρητά την ανομοιότητα των δύο κόμβων θέτοντας ένα *cannot link* περιορισμό μέσα από την δημιουργία μιας ακμής αρνητικού βάρους μεταξύ τους. Παρατηρούμε λοιπόν στο τελευταίο σχήμα πως αλλάζει με την προσθήκη της ακμής αυτής η αναπαράσταση του γράφου στον χώρο ιδιοδιανυσμάτων και πως ανακαλύπτει ο αλγόριθμος την βέλτιστη τομή.



Σχήμα 4.13: Αποτελεσματικότητα των αρνητικών βαρών στο πρόβλημα της κατάτμησης

### Παρένθεση: Έλεγχος συνέπειας *CL* περιορισμών

Στην περίπτωση *soft* περιορισμών κατά την οποία κάθε *CL* περιορισμός έχει μια αβεβαιότητα ως προς την ισχύ του ή ακόμη και όταν όλοι οι περιορισμοί αυτοί ανθροίζονται στην συνάρτηση βελτιστοποίησης, δεν υπάρχει πρόβλημα συνέπειας αλλά μπορούμε να ελέγξουμε μετά την κατάτμηση σε ποιό βαθμό ικανοποιήθηκαν όλοι οι περιορισμοί μαζί και κατά πόσο αυτό είναι αποδεκτό.

Στην περίπτωση των *hard* περιορισμών όμως θέλουμε να ξέρουμε αν το σύνολο που δόθηκε είναι συνεπές και αν μπορούμε να εξάγουμε κάποια πληροφορία από την δομή του. Δεδομένου ότι θέλουμε οι κόμβοι να μπορούν να χωριστούν σε  $K$  κλάσεις, αν φτιάξουμε

τον γράφο των  $CL$  περιορισμών με μια ακμή μεταξύ κόμβων που ανήκουν στο σύνολο που μας δώθηκε, τότε το πρόβλημα ελέγχου συνέπειας είναι ισοδύναμο με το πρόβλημα του  $K$ -χρωματισμού ενός γράφου. Το πρόβλημα αυτό έχει αποδειχθεί NP-complete για  $K \geq 3$ . Για  $K=2$  όμως υπάρχει ένα ενδιαφέρον στην δομή του γράφου, καθώς αυτός πρέπει να είναι διμερής (ή να αποτελείται από διμερείς συνιστώσες αν δεν είναι συνεκτικός). Ωστόσο, κάθε δύο κόμβοι που έχουν ένα κοινό γείτονα πρέπει να ανήκουν στην ίδια κλάση. Άρα κάποιοι περιορισμοί  $CL$  μπορούν να εξαφανιστούν και να αντικατασταθούν με  $ML$  περιορισμούς που έχουμε ήδη λύσει. Τελικά, λόγω της μεταβατικότητας των  $ML$  περιορισμών και της λύσης που κάναμε με την συμπίεση των κόμβων, κάθε διμερής συνιστώσα αντικαθίσταται από μια και μόνο ακμή και κάθε κόμβος μπορεί να συμμετάσχει το πολύ σε ένα  $CL$  περιορισμό.

## Εναλλακτική λύση: Περιορισμός του εφικτού χώρου λύσεων

### Μοντελοποίηση Περιορισμών

$$K_{u_i} \neq K_{u_j} \Leftrightarrow \mathbf{X}^T e_i \neq \mathbf{X}^T e_j,$$

Στον διακριτό χώρο η συνθήκη μπορεί να γραφεί με πολλούς εναλλακτικούς τρόπους. Το ζητούμενο είναι όταν γίνει το relaxation του προβλήματος οι περιορισμοί να διατηρούν την σημασία τους, τουλάχιστον προσεγγιστικά. Παραδείγματος χάρη στα ([47],[48],[50]) οι συγγραφείς μοντελοποιούν κάθε  $CL$  περιορισμό (σε 2 κλάσεις) σαν :  $|x_i - x_j| = 1$ . Ωστόσο, επειδή στο συνεχή χώρο δεν μπορούμε να ελέγξουμε την απόσταση των τιμών του διανύσματος κατάτμησης οι συγγραφείς αντί να προσπαθήσουν να τον επιβάλλουν τον απαλύφουν εντελώς και επικεντρώνονται μόνο στους  $ML$  περιορισμούς.

Από την άλλη κάποια papers ([22]) θέτουν (πάλι για δύο κλάσεις)  $x_i = -x_j$  για αυτούς τους περιορισμούς. Και αυτός ο τρόπος όμως έχει προβλήματα μοντελοποίησης γιατί παρότι επιβάλλει το αντίθετο πρόσημο μεταξύ των κόμβων, επιβάλλει και αντίθετες τιμές στους κόμβους. Αυτό περιορίζει πολύ το χώρο εφικτών λύσεων και εισάγει μεγάλο bias προς κατατμήσεις με ίσα volumes, γιατί γνωρίζουμε ότι οι 2 τιμές του διανύσματος συνδέονται με τη σχέση  $a = -\frac{\text{vol}(A')}{\text{vol}(A)}b$ . Αυτό αποτυπώνεται και στα πειράματα αυτού και των επόμενων paper. Γενικά, η μοντελοποίηση του περιορισμού σαν γραμμικό με οποιοδήποτε τρόπο, θα περιορίσει περισσότερο από όσο πρέπει το χώρο εφικτών λύσεων και θα οδηγήσει σε κακή σταθμισμένη τομή.

Ίσως πιο χρήσιμη είναι η αντιμετώπιση των X.Wang, I.Davidson ([43], [44]) οι οποίοι αναγνωρίζοντας την δυσκολία να επιβάλλουν κάθε ένα περιορισμό ξεχωριστά, τους αντιμετωπίζουν όλους συνολικά μέσα από την τετραγωνική μορφή:

$$f(u) = u^T \mathbf{Q} u = \sum_{i,j=1}^N u_i Q_{ij} u_j$$

Θεωρώντας ότι σε hard (soft)  $ML$  περιορισμούς  $Q_{ij} = +1 (> 0)$  ενώ σε hard (soft)  $CL$  περιορισμούς  $Q_{ij} = -1 (< 0)$ , η απαίτηση να μεγιστοποιείται η  $f(u)$  ενθαρρύνει έμμεσα τα  $u_i$  και  $u_j$  να παίρνουν το ίδιο πρόσημο όταν συμμετάσχουν σε  $ML$  περιορισμούς και

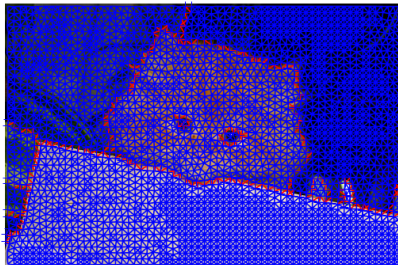
αντίθετο όταν συμμετάσχουν σε  $CL$  περιορισμούς. Αντί για μεγιστοποίηση της συνάρτησης οι συγγραφείς θέωρούν ένα εξωτερικό κατώφλι δοσμένο από την χρήστη και απαιτούν να ικανοποιείται ο περιορισμός:  $f(u) \geq \alpha$ . Κάνοντας χρήση των συνθηκών  $KKT$  καταλήγουν σε ένα πρόβλημα γενικευμένων ιδιοτιμών για το ζεύγος  $(\bar{\mathbf{L}}, \bar{\mathbf{Q}} - \frac{\beta}{\text{vol}(G)} \mathbf{I})$  (κρατώντας μόνο θετικές ιδιοτιμές) όπου το  $\beta$  είναι μια συνάρτηση του κατωφλίου του χρήστη. Αυτή δεν πρέπει να ξεπερνάει το  $\lambda_{\max}(\bar{\mathbf{Q}})\text{vol}(G)$  ώστε να έχει λύση το πρόβλημα ιδιοτιμών. Ο παραπάνω τρόπος έχει το πλεονέκτημα της ευελιξίας στην ικανοποίηση των περιορισμών (μοντελοποιεί κατά κάποιον τρόπο το *purity* ενός *cluster*) και επίσης είναι αρκετά εύρωστος σε λάθη του χρήστη. Τέλος ενοποιεί τους *hard* και *soft* περιορισμούς αλλά τελικά ακολουθεί την λογική των *soft* περιορισμών γιατί δεν εξασφαλίζει την ικανοποίησή τους αλλά πιο πολύ εφαρμόζει *regularization* μέσα από την *Lagrangian*. Από την άλλη, ο τρόπος αυτός έχει το αρνητικό ότι δεν δίνει την δυνατότητα να επιλέξουμε ποιούς μεμονωμένους περιορισμούς θα ικανοποιήσουμε. Πολλές φορές ακόμη και ισχυροί περιορισμοί μπορεί να μην ικανοποιούνται γιατί στην συνάρτηση παίζουν ρόλο και οι τιμές των  $u_i$  εκτός από τις τιμές των  $Q_{ij}$ . Αν οι πρώτες είναι μικρές κατεβάζουν την σημασία ισχυρών περιορισμών. Ένας ακόμη λόγος για το τελευταίο επιχείρημα είναι ότι τελικά ο χρήστης δίνει σαν κατώφλι το  $\beta$  και ο αλγόριθμος αποφασίζει για το  $\alpha$  που θα είναι πάντα μεγαλύτερο αλλά μπορεί όχι αρκετά ώστε να περιορίσει το πρόβλημα στο βαθμό που θα ικανοποιούνται αρκετοί περιορισμοί. Τέλος, η γενίκευση που κάνουν οι συγγραφείς για  $K > 2$  κλάσεων δεν είναι καλά θεμελιωμένη.

#### 4.2.4 Παράδειγμα επίδρασης ML και CL περιορισμών σε εικόνες

Εφαρμόσαμε τον MLCLNcut αλγόριθμό μας σε παραδείγματα εικόνων από το Berkeley Dataset και παρουσιάζουμε την επίδραση που έχουν στην τελική κατάτμηση. Ο γράφος των αρνητικών ακμών όταν υπάρχουν *Cannot link* περιορισμοί, κατασκευάζεται αντιστρέφοντας το πρόσημο του βάρους των ακμών αν αυτοί συνδέονται με ακμή. Αυτό έχει μεγαλύτερη επίδραση προφανώς σε κόμβους που αρχικά είχαν ισχυρή σύνδεση. Η λογική είναι ότι αν δύο κόμβοι έχουν πολύ ασθενή σύνδεση τότε πιθανότατα ο αλγόριθμος θα τους στρέψει σε άλλες κλάσεις και συνεπώς δεν έχει νόημα να αλλάξουμε πολύ τον γράφο. Πιο θεμελιωμένος διαισθητικά τρόπος θα ήταν να τους συνδέαμε με ακμή αντίθετη από το μεγαλύτερο βάρος του βαρύτερου μονοπατιού που τα συνδέει, γιατί ακόμη και αν έχουν ασθενή σύνδεση μεταξύ τους μπορεί να υπάρχει κάποιο μονοπάτι πολύ ισχυρής σύνδεσης που πρέπει να κοπεί σε κάποιο σημείο για να αποσυνδεθούν. Το να βρει κανείς το βαρύτερο μονοπάτι όμως είναι πρόβλημα πολύ δύσκολο στην πράξη (*NP - Complete*) και για αυτό το λόγο θα χρησιμοποιήσουμε αυτή την ευρετική. Επίσης, στην πράξη παρατηρήσαμε πως μεγάλα αρνητικά βάρη χαλάνε την αριθμητική ευστάθεια του προβλήματος.



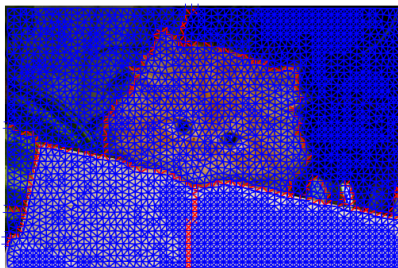
(α') original



(β') Ncut 11 clusters



(γ') Ncut 11 clusters eig

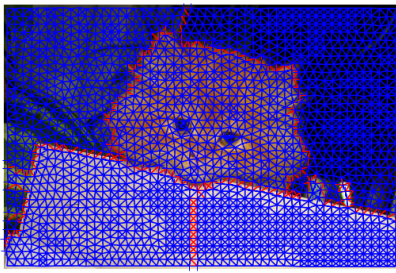


(δ') ML Ncut 9 clusters



(ε') ML Ncut 9 clusters eig

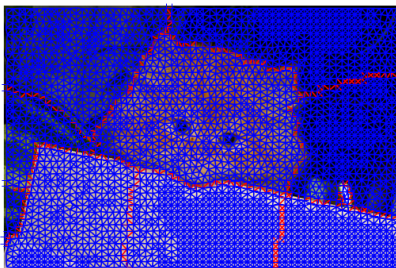




(ε') ML 2 Ncut 7 clusters



(ζ') ML 2 Ncut 7 clusters



(η') ML CL Ncut 9 clusters



(θ') ML CL Ncut 9 clusters eig

Σχήμα 4.14: Εφαρμογή του αλγορίθμου ML, CL Ncut σε εικόνες

Στην παραπάνω εικόνα, εφαρμόσαμε τον αλγόριθμο των σταθμισμένων τομών ενσωματώνοντας και *Must Link* και *Cannot Link* περιορισμούς. Στο πρώτο σχήμα φαίνεται η κατάτμηση της εικόνας σε 11 κλάσεις, όπου παρατηρούμε ότι ο αλγόριθμος απομονώνει τα μάτια της γάτας και κάνει υπερβολική κατάτμηση στο καρπούζι. Κατεβάζοντας τις κλάσεις σε 9 απλά ενσωματώθηκαν τα μικρά cluster στα δεξιά, στο μεγάλο cluster του κουτιού. Στην συνέχεια, εφαρμόσαμε *ML* περιορισμούς μεταξύ των ματιών της γάτας και μεταξύ των ματιών και του κεφαλιού. Κάνοντας ξανά κατάτμηση της εικόνας παρατηρούμε ότι οι περιορισμοί ενσωματώθηκαν στο πρόβλημα επιτυχώς. Στην συνέχεια κάναμε το ίδιο για το καρπούζι σε έναν αριθμό pixel εκατέρωθεν του πράσινου και του καφέ κομματιού που φαίνεται στο σχήμα (ε') καθώς και του μικρού γκρι cluster στα δεξιά της εικόνας με το πορτοκαλί background. Η κατάτμηση (απαιτώντας 2 λιγότερες κλάσεις) φαίνεται στο (ζ') όπου παρατηρούμε ότι αν και τα συγκεκριμένα pixel κατέληξαν μαζί, απλά τραβήξαν λίγο πιο χαμηλά την γραμμή διαχωρισμού. Το άλλο cluster ενσωματώθηκε πλήρως αλλά άλλαξε αρκετά την δομή της συνολικής κατάτμησης. Παρατηρούμε στο τελευταίο σχήμα πόσο μεγάλη χρησιμότητα και ευελιξία δίνει το να βάζουμε *Cannot Link* περιορισμούς. Αν αντί να βάλουμε *ML* περιορισμούς μεταξύ των διαχωρισμένων κομματιών του καρπουζιού, βάλουμε

$CL$  περιορισμούς μεταξύ του καρπούζιου και του background, παρατηρούμε στο (θ') πως καταφέρνει να διαχωρίσει το καρπούζι από το background. Εν γένει δεν μας πειράζει να έχουμε καταταμήσει ένα αντικείμενο σε περισσότερα κομμάτια, αρκεί να έχουμε διατηρήσει το περίγραμμά του σαυτα. Εδώ φαίνεται η χρησιμότητα των  $CL$  περιορισμών, που μπορούν να ξεχωρίσουν αντικείμενα όταν αυτά έχουν ενσωματωθεί στο περιβάλλον τους, ακριβώς όπως οι  $ML$  περιορισμοί μπορούν να ενώσουν κομμάτια αντικειμένων (όχι απαραίτητα κλάσεων ολόκληρων όπως εδώ).

#### 4.2.5 Άλλοι Περιορισμοί

Μπορεί για την εφαρμογή μας να είναι απαραίτητη η ενσωμάτωση άλλων περιορισμών όπως τα μεγέθη των κλάσεων, κόμβοι που απαγορεύεται να μπουν σε συγκεκριμένη κλάση κλπ. Αφήνουμε τα προβλήματα αυτά για μελλοντική ενασχόληση και αναφέρουμε μόνο λύσεις που έχουν προταθεί βιβλιογραφικά.

#### Περιορισμός του μεγέθους ενός cluster

Έχουμε παρατηρήσει πειραματικά ότι γενικά το Normalized Cut δημιουργεί πολύ ισορροπημένα cluster. Ωστόσο, μπορεί για κάποιο λόγο να ξέρουμε a-priori ότι το αντικείμενο που θέλουμε να εστιάσουμε έχει μικρότερο ή μεγαλύτερο μέγεθος από τα υπόλοιπα, οπότε κατευθύνουμε την τομή να συμπεριλάβει τουλάχιστον ένα τέτοιο αντικείμενο. Επιπλέον, λόγω του πολλαπλασιαστικού όρου με τον οποίο επιβάλλεται η κανονικοποίηση των κλάσεων αρκετές φορές η απόδοση του αλγορίθμου επηρεάζεται από την ύπαρξη *outliers* με αποτέλεσμα να μην δημιουργούνται ισορροπημένες κλάσεις. Τέλος, μπορεί ο περιορισμός να επεκταθεί και σε ανισότητα αν δεν είναι γνωστό a-priori το ακριβές μέγεθος των κλάσεων.

#### Μοντελοποίηση Περιορισμού

$$size(K_j) = c \Leftrightarrow \#nz(X_j) = c \Leftrightarrow \mathbf{1}_N^T \mathbf{X} \mathbf{e}_j = \alpha_j \cdot c$$

Θα μπορούσαμε να απαλύσουμε το  $\alpha_j$  το οποίο είναι άγνωστο πριν την κατάτμηση αλλά οι εξισώσεις γίνονται όπως είδαμε και παραπάνω πολύπλοκες. Εξάλλου όταν θα γίνει relax το πρόβλημα και οι τιμές του  $X$  πραγματικές ο περιορισμός δεν θα έχει ακριβώς το ίδιο νόημα. Για αυτό το λόγο, θα μπορούσαμε ακόμη και να θεωρήσουμε το δεύτερο μέλος σαν σταθερά. Ακόμη πιό χρήσιμο στην πράξη θα ήταν να επεκτείνουμε τον περιορισμό σε ανισότητα.

Στο [8] προτείνεται για δύο κλάσεις η εισαγωγή ενός προσθετικού όρου μαζί με τον όρο της ελάχιστης τομής στην αντικειμενική συνάρτηση, αντί για τον πολλαπλασιαστικό τρόπο που επιβάλλεται στο Normalized Cut. Το πρόβλημα τελικά ανάγεται στην εύρεση του μεγαλύτερου ιδιοδιανύσματος του πίνακα  $\mathbf{W} - \alpha\beta\beta^T$ , όπου  $\beta$  το διάνυσμα βαρών των κόμβων ενώ  $\alpha$  μια παράμετρος ομαλοποίησης που σχετίζεται με το μέγεθος των κλάσεων και για την οποία γίνεται δυαδική αναζήτηση για την εύρεση της τιμής που θα δώσει τα ζητούμενα μεγέθη κλάσεων.



### Απαγόρευση ενός κόμβου να ανατεθεί σε ένα cluster

Έστω ότι, ανάποδα με πριν, θέλουμε ο κόμβος  $i$  να μην μπορεί να ανατεθεί στο cluster  $j$ . Αυτό είναι ισοδύναμο με τη συνθήκη:

$$X_j(i) = 0 \Leftrightarrow e_i^T X e_j = 0$$

Αν ξέρουμε ότι κάποιος κόμβος δεν μπορεί να ανήκει σε περισσότερα του ενός cluster τότε προσθέτουμε και άλλα διανύσματα στήλης στο δεξί διάνυσμα ( $e_j$ ) και η μορφή της εξίσωσης γίνεται:

$$e_i^T \cdot \mathbf{X} \cdot [e_j | e_{j'} | \dots | e_{j''}] = 0.$$

Αν ξέρουμε ότι παραπάνω από ένας κόμβοι έχουν αυτό τον περιορισμό τότε θα επεκτείνουμε το αριστερό διάνυσμα ( $e_i$ ) προσθέτοντας γραμμές οπότε η νέα μορφή της εξίσωσης θα ήταν:

$$[e_i | e_{i'} | \dots | e_{i''}]^T \cdot \mathbf{X} \cdot e_j = \mathbf{0}$$

Και προφανώς αν οι πολλοί αυτοί κόμβοι απαγορευόταν να μπουν στο ίδιο σύνολο από cluster τότε οι εξισώσεις τους θα ομαδοποιούνταν ως:

$$[e_i | e_{i'} | \dots | e_{i''}]^T \cdot \mathbf{X} \cdot [e_j | e_{j'} | \dots | e_{j''}] = \mathbf{0}$$

Όλα τα παραπάνω είναι ειδικές περιπτώσεις του περιορισμού:  $\mathbf{A}^T \cdot \mathbf{X} \cdot \mathbf{B} = \mathbf{0}$ ,

όπου οι πίνακες  $\mathbf{A}, \mathbf{B}$  έχουν ειδική μορφή με τα μοναδιαία διανύσματα σαν στήλες (μπορούμε να θεωρήσουμε ότι ενός από τους 2 είναι διατεταγμένες) και επίσης ο  $\mathbf{A}$  δεν μπορεί να έχει όλους του κόμβους (κενό cluster) ενώ ο  $\mathbf{B}$  δεν μπορεί να έχει όλα τα cluster (κόμβος χωρίς ανάθεση).

Στην γενικότερη περίπτωση κάποιος μπορεί να έχει ένα αριθμό από τέτοιες εξισώσεις από ομάδες κόμβων που δεν μπορούν να ανήκουν σε άλλες ομάδες cluster. (Οι πίνακες συνολικά πρέπει να ικανοποιούν κάποιες συνθήκες που μπορούν να βγουν αν πάρουμε τις στήλες τους ξεχωριστά)

Οι εξισώσεις, στο relaxed πρόβλημα έχουν ακριβώς την ίδια ερμηνεία (εκτός από την πολύ σπάνια περίπτωση που ένα cluster μένει χωρίς στοιχεία και επιλέγεται να μεταφερθεί ένα στοιχείο από το μεγαλύτερο cluster σε ένα άλλο που απαγορεύεται να πάει. Τότε πρέπει να φροντίσουμε απλά να μην πάρουμε το συγκεκριμένο στοιχείο).



## Κεφάλαιο 5

# Επίλογος

### 5.1 Συνεισφορές της Διπλωματικής

1. Η κύρια συνεισφορά της εργασίας αυτής είναι η ανάπτυξη ενός ημι-επιβλεπόμενου framework για την επίλυση προβλημάτων κατάτμησης με φασματικές γραφοθεωρητικές μεθόδους. Συγκεκριμένα, μετά τον συμπαγή φορμαλισμό του προβλήματος κατάτμησης στο κεφάλαιο 3 προσπαθήσαμε να ενσωματώσουμε τρία είδη εξωγενών περιορισμών στο πρόβλημα των σταθμισμένων γραφοτομών.

(α') Ο πρώτος είναι οι Must Link περιορισμοί μεταξύ κόμβων του γράφου. Μοντελοποιήσαμε και λύσαμε αναλυτικά το πρόβλημα αυτό βρίσκοντας έναν αλγόριθμο για την ελαχιστοποίηση των περιορισμών και κατόπιν, επεκτείναμε τον αλγόριθμο για την πλήρη επίλυση του προβλήματος σε γραμμικό χρόνο. Οι αποδείξεις ορθότητας του αλγορίθμου και τα παρεμφερή θεωρήματα παρουσιάστηκαν αναλυτικά στο κεφάλαιο 4. Τα αποτελέσματα αυτά ενοποιούν μεθόδους που επέλυαν το πρόβλημα της βελτιστοποίησης με μεθόδους (λιγότερο θεμελιωμένες) που επενέβαιναν απευθείας στους γράφους ομοιότητας. Κατόπιν, ο αλγόριθμος αυτός εφαρμόστηκε σε point cloud δεδομένα για να επιβεβαιωθεί και πειραματικά η ανωτερότητά του ως προς τον αλγόριθμο μη-επιβλεπόμενης κατάτμησης και κατέπεχτα η αναγκαιότητα ενσωμάτωσης τέτοιων περιορισμών. Θεμελιώσαμε κυρίως θεωρητικά την ανωτερότητά του αλγορίθμου ως προς άλλους αλγορίθμους ημι-επιβλεπόμενης μάθησης που επιλύουν το ίδιο πρόβλημα. Στο τέλος του κεφαλαίου παρουσιάστηκε και η εφαρμογή του σε δεδομένα εικόνων.

(β') Ωστόσο η περιορισμένη εκφραστικότητα του παραπάνω περιορισμού, που δείξαμε μέσα από εφαρμογές σε δεδομένα, μας οδήγησε στην ενσωμάτωση Cannot Link περιορισμών στο πρόβλημα. Η τετραγωνική φύση του περιορισμού δεν μας επέτρεψε την αναλυτική επίλυσή του, οπότε καταφύγαμε στην επίλυσή του με μετασχηματισμό του γράφου. Αντίθετα με την τάση στην βιβλιογραφία, χρησιμοποιήσαμε αρνητικές ακμές για να περιγράψουμε την απώθηση μεταξύ κόμβων και κατόπιν εφαρμόσαμε την τομή με χρήση προσημασμένων σταθμισμένων το-

μών. Ένα αποτέλεσμα του περιορισμού αυτού σε εικόνες φαίνεται στο τέλος του κεφαλαίου 4.

(γ') Τέλος, μοντελοποιήσαμε unary περιορισμούς για το πρόβλημα των σταθμισμένων τομών και επιλύσαμε το πρόβλημα προσεγγιστικά για δύο κλάσεις. Στο τέλος του κεφαλαίου προτείνουμε κάποιους ακόμη περιορισμούς που θα πρέπει να λυθούν, ίσως με αναγωγή στους προηγούμενους.

2. Η επόμενη συνεισφορά της διπλωματικής είναι πειραματική και έχει να κάνει με την σύγκριση μεθόδων γράφων για κατάτμηση δεδομένων αλλά και με την χρήση του *RAG – SLIC graph* για τον μετασχηματισμό της εικόνας από την κανονική δομή της με εικονοστοιχεία, στην δομή του γράφου από *superpixel*. Πολλές εφαρμογές σε εικόνες δείχνουν την αποτελεσματικότητά της σε σχέση με μεθόδους που χρησιμοποιούσαν *Watershed superpixels* αλλά και την επιτάχυνση και βελτίωση της ποιότητας της κατάτμησης που αποφέρουν γενικότερα στο πρόβλημα των σταθμισμένων τομών.
3. Τέλος, στην διπλωματική αυτή θέσαμε την ανάλυση του αλγορίθμου των σταθμισμένων γραφοτομών σε πιο γερά θεμέλια μέσα από τις ακριβείς αποδείξεις κάποιων θεωρημάτων και μέσα από την σύνδεσή του με άλλους καλά μελετημένους αλγορίθμους όπως οι αραιές τομές. Ακομη, επιμείναμε αρκετά να χτίσουμε την διαίσθηση πίσω από το γιατί λειτουργεί η μέθοδος, αντιπαραβάλλοντας ή αντιστοιχίζοντας την με πληθώρα άλλων γνωστών αλγορίθμων κατάτμησης, αλλά και πού δεν λειτουργεί ικανοποιητικά, καταδεικνύοντας κάποια από τα προβλήματα της μεθόδου σε πρακτικό κυρίως επίπεδο.

## 5.2 Μελλοντικές επεκτάσεις της διπλωματικής

1. **Multiscale Normalized Cuts.** Παρατηρήσαμε, ότι ο αλγόριθμος των σταθμισμένων γραφοτομών δημιουργεί αρκετά σταθμισμένα *clusters* και για αυτό παρουσιάζει κακά αποτελέσματα όταν τα δεδομένα (της εικόνας ή γενικότερα) είναι σε πολλές διαφορετικές κλίμακες. Ένα θέμα που συσχετίζεται με το παραπάνω, είναι η δυσκολία προσέγγισης του αριθμού των κλάσεων. Τέλος, αρκετές φορές η κακή επίδοση της κατάτμησης οφείλεται στην ευαισθησία του αλγορίθμου σε *outliers*. Όλα αυτά τα προβλήματα μπορούν να επιλυθούν αν επεκτείνουμε τον αλγόριθμο να ενεργεί σε πολλές διαφορετικές κλίμακες. Η ιδέα του *Multiscale Normalized Cuts* έχει σαν αφορμή το ημι επιβλεπόμενο *framework* που υλοποιήσαμε προηγουμένως. Συγκεκριμένα, ο τρόπος που σκοπεύουμε να προχωρήσουμε είναι ο εξής: Στην αρχή θα δημιουργείται μια πυραμίδα γράφων από την δειγματοληψία σημαντικών κόμβων (έχουμε δοκιμάσει τη δειγματοληψία με τον αλγόριθμο του *page rank* και με τον βαθμό των κόμβων). Κατόπιν, το τελευταίο στάδιο της πυραμίδας θα κατατμείται με χρήση του κλασικού αλγορίθμου *Normalized Cuts*. Από αυτήν την κατάτμηση σχηματίζονται κλάσεις από κόμβους, οι οποίες κωδικοποιούν την πληροφορία του ποιοί κόμβοι συνδέονται μεταξύ

τους και ποιοί όχι. Κρατώντας λοιπόν τους Must Link και Cannot Link περιορισμούς από αυτήν την κατάτμηση, προχωράμε στο επόμενο επίπεδο της πυραμίδας λύνοντας πλέον ένα ημι-επιβλεπόμενο πρόβλημα ML - CL Normalized Cuts, όπου οι περιορισμοί ML και CL μπορούν να είναι και soft, ώστε να μην να παραμένει η συσχέτιση μεταξύ των επιπέδων, αλλά να μην λαμβάνεται η απόλυτη απόφαση για την ταξινόμηση ενός κόμβου σε ένα επίπεδο της πυραμίδας, αλλά σε όλα. Έχουμε κατασκευάσει τον αλγόριθμο αυτόν με κάποια καλά αποτελέσματα σε εικόνες που η μιας κλίμακας κατάτμηση αποτυγχάνει. Επίσης, αξίζει να παρατηρηθεί ότι παρότι έχουμε χρησιμοποιήσει το ημι-επιβλεπόμενο framework ο αλγόριθμος δεν χρειάζεται καθόλου την επέμβαση του χρήστη.

2. Το επόμενο βήμα για την βελτίωση της ποιότητας της κατάτμησης και της γενίκευσης της καλής συμπεριφοράς του αλγορίθμου *RAG-SLIC NCut* σε όλα τα είδη εικόνων είναι η ενσωμάτωση καλύτερων χαρακτηριστικών. Πέρα από την χρωματική πληροφορία και την πληροφορία ακμών, είναι απαραίτητη η ενσωμάτωση χαρακτηριστικών υφής σε εικόνες που αυτό είναι το εξέχον χαρακτηριστικό. Επιπλέον, υπάρχουν 3 παράμετροι που έχουν ρυθμιστεί πειραματικά και αυτές είναι: το μέγεθος των *superpixel*, το  $\sigma$  της Γκαουσιανής στην απόσταση των χαρακτηριστικών και ο αριθμός των κλάσεων. Οι παράμετροι αυτές είναι αλληλοεξαρτώμενες και πρέπει να βρεθεί ένας αυτόματος τρόπος να ρυθμίζονται στο πρόβλημα.
3. Τέλος, χρήσιμη θα ήταν η επέκταση του παρόντος ημι-επιβλεπόμενου framework σε περιορισμούς ποικίλης φύσεως, όπως αυτοί που περιγράφουμε στο τέλος του κεφαλαίου ή των *soft* εκδοχών τους, που δεν εστιάσαμε ιδιαίτερα στην διπλωματική. Δεδομένου ότι όλοι είναι κυρτοί θα ήταν χρήσιμη η ανάπτυξη μιας ενιαίας θεωρίας, που θα εχμεταλλεύεται μεθόδους βελτιστοποίησης για *Semi-Definite* προβλήματα βελτιστοποίησης.



# Bibliography

- [1] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Susstrunk. «SLIC Superpixels Compared to State-of-the-Art Superpixel Methods». In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 34.11 (2012), pp. 2274–2282.
- [2] C. Alzate and J. A. K. Suykens. «Image Segmentation using a Weighted Kernel PCA Approach to Spectral Clustering». In: *Symposium on Computational Intelligence in Image and Signal Processing*. IEEE. 2007, pp. 208–213.
- [3] S. Arora, S. Rao, and U. Vazirani. «Expander Flows, Geometric Embeddings and Graph Partitioning». In: *Proceedings of the Thirty-sixth Annual Symposium on Theory of Computing*. ACM, 2004, pp. 222–231.
- [4] C. G. Bampis, P. Maragos, and A. C. Bovik. «Graph-Driven Diffusion and Random Walk Schemes for Image Segmentation». In: *IEEE Transactions on Image Processing* 26.1 (2017), pp. 35–50.
- [5] M. Belkin and P. Niyogi. «Laplacian Eigenmaps for Dimensionality Reduction and Data Representation». In: *Neural Computation* 15.6 (2003), pp. 1373–1396.
- [6] M. Van den Bergh, X. Boix, G. Roig, and L. Van Gool. «SEEDS: Superpixels Extracted via Energy-Driven Sampling». In: *International Journal of Computer Vision* 111.3 (2015), pp. 298–314.
- [7] A. Challa, S. Danda, B. S. Daya Sagar, and L. Najman. «An Introduction to Gamma-Convergence for Spectral Clustering». In: *Discrete Geometry for Computer Imagery*. Vol. 10502. Lecture Note In Computer Sciences. Springer, 2017, pp. 185–196.
- [8] Y. Chen, Y. Zhang, and X. Ji. «Size Regularized Cut for Data Clustering». In: *Proceedings of the Eighteenth International Conference on Advances in Neural Information Processing Systems*. NIPS Foundation. 2006, pp. 211–218.
- [9] F. R. K. Chung. *Spectral Graph Theory*. American Mathematical Society, 1997.
- [10] T. Cour and J. Shi. «Solving Markov Random Fields with Spectral Relaxation». In: *Proceedings of the Eleventh International Conference on Artificial Intelligence and Statistics*. PMLR. 2007, pp. 75–82.

- [11] I. S. Dhillon, Y. Guan, and B. Kulis. «Weighted Graph Cuts without Eigenvectors A Multilevel Approach». In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 29.11 (2007), pp. 1944–1957.
- [12] P. G. Doyle and J. L. Snell. *Random Walks and Electric Networks*. Mathematical Association of America, 1984.
- [13] J. Gallier. «Spectral Theory of Unsigned and Signed Graphs. Applications to Graph Clustering: a Survey». In: *Computing Research Repository* (2016). arXiv: [1601.04692](https://arxiv.org/abs/1601.04692).
- [14] G. H. Golub. «Some Modified Matrix Eigenvalue Problems». In: *SIAM Review* 15.2 (1973), pp. 318–334.
- [15] L. Grady. «Random Walks for Image Segmentation». In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 28 (2006), pp. 1768–1783.
- [16] X. He, S. Zhang, and Y. Liu. «An Adaptive Spectral Clustering Algorithm Based on the Importance of Shared Nearest Neighbors». In: *Algorithms* 8 (2015), pp. 177–189.
- [17] H. Hu, J. Feng, C. Yu, and J. Zhou. «Multi-Class Constrained Normalized Cut With Hard, Soft, Unary and Pairwise Priors and its Applications to Object Segmentation». In: *IEEE Transactions on Image Processing* 22.11 (2013), pp. 4328–4340.
- [18] T. Joachims. «Transductive Learning via Spectral Graph Partitioning». In: *Proceedings of the Twentieth International Conference on Machine Learning*. ACM, 2003, pp. 290–297.
- [19] P. D. K. He G. Gkioxari and R. Girshick. «Mask R-CNN». In: *Computing Research Repository* (2017). arXiv: [1703.06870v2](https://arxiv.org/abs/1703.06870v2).
- [20] A. K. Jain. «Data Clustering: 50 Years Beyond K-Means». In: *Pattern Recognition Letters* 31 (2010), pp. 651–666.
- [21] S. D. Kamvar, D. Klein, and C. D. Manning. «Spectral Learning». In: *Proceedings of the Eighteenth International Joint Conference on Artificial Intelligence*. ACM, 2003, pp. 561–566.
- [22] J. Kawale and D. Boley. «Constrained spectral clustering using L1 regularization». In: *Proceedings of the 2013 SIAM International Conference on Data Mining*. SIAM Society, 2013, pp. 103–111.
- [23] J. Kunegis, S. Schmidt, A. Lommatzsch, J. Lerner, E. W. De Luca, and S. Albayrak. «Spectral Analysis of Signed Graphs for Clustering, Prediction and Visualization». In: *Proceedings of the 2010 SIAM International Conference on Data Mining*. SIAM Society, 2010, pp. 559–570.



- [24] Z. Lu and M. A. Carreira-Perpinan. «Constrained spectral clustering through affinity propagation». In: *Proceedings of the International Conference on Computer Vision and Pattern Recognition*. 2008, pp. 1–8.
- [25] U. Luxburg. «A Tutorial on Spectral Clustering». In: *Statistics and Computing* 17.4 (2007), pp. 395–416.
- [26] U. V. Luxburg, O. Bousquet, and M. Belkin. «Limits of Spectral Clustering». In: *Proceedings of the Seventeenth International Conference on Advances in Neural Information Processing Systems*. NIPS Foundation. 2004, pp. 857–864.
- [27] M. W. Mahoney. «Lecture Notes on Spectral Graph Methods». In: *Computing Research Repository* (2016). arXiv: [1608.04845](https://arxiv.org/abs/1608.04845).
- [28] M. Meila and J. Shi. «Learning Segmentation by Random Walks». In: *Proceedings of the Thirteenth International Conference on Advances in Neural Information Processing Systems*. NIPS Foundation. 2001, pp. 873–879.
- [29] A. Y. Ng, M. I. Jordan, and Y. Weiss. «On Spectral Clustering: Analysis and an Algorithm». In: *Proceedings of the Fourteenth International Conference on Advances in Neural Information Processing Systems*. NIPS Foundation. 2001, pp. 849–856.
- [30] L. Page, S. Brin, R. Motwani, and T. Winograd. «The PageRank citation ranking: Bringing order to the Web». In: *Proceedings of the Seventh International Conference on World Wide Web*. IW3C2. 1998, pp. 161–172.
- [31] A. Rahimi and B. Recht. «Clustering with Normalized Cuts is Clustering with a Hyperplane». In: *Proceedings of the Fifty-sixth International Workshop on Statistical Learning in Computer Vision*. 2004.
- [32] D. Raviv, R. Kimmel, and A. M. Bruckstein. «Graph Isomorphisms and Automorphisms via Spectral Signatures». In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 35 (2012), pp. 1985–1993.
- [33] X. Ren and J. Malik. «Learning a Classification Model for Segmentation». In: *Proceedings of the Ninth International Conference on Computer Vision*. IEEE. 2003, pp. 10–17.
- [34] I. S. Dhillon, Y. Guan, and B. Kulis. «Kernel k-means, Spectral Clustering and Normalized Cuts». In: *Proceedings of the Tenth SIG International Conference on Knowledge Discovery and Data Mining*. ACM. 2004, pp. 551–556.
- [35] A. Sandryhaila and J. M. F. Moura. «Discrete Signal Processing on Graphs». In: *IEEE Transactions on Signal Processing* 61 (2013), pp. 1644–1656.
- [36] J. Sedoc, J. Gallier, L. H. Ungar, and D. P. Foster. «Semantic Word Clusters Using Signed Normalized Graph Cuts». In: *Computing Research Repository* (2016). arXiv: [1601.05403](https://arxiv.org/abs/1601.05403).
- [37] J. Shi and J. Malik. «Normalized Cuts and Image Segmentation». In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 22.8 (2000), pp. 888–905.

- [38] D. I. Shuman, S. K. Narang, P. Frossard, A. Ortega, and P. Vandergheynst. «The emerging field of signal processing on graphs: Extending high-dimensional data analysis to networks and other irregular domains». In: *IEEE Signal Processing Magazine* 30 (2013), pp. 83–98.
- [39] D. Spielman and S. Teng. In: *SIAM Journal on Matrix Analysis and Applications* 35.3 (2014), pp. 835–885.
- [40] D. Stutz, A. Hermans, and B. Leibe. «Superpixels: An Evaluation of the State-of-the-Art». In: *Computer Vision and Image Understanding* 166 (2018), pp. 1–27.
- [41] A. Tharwat, T. Gaber, A. Ibrahim, and A. E. Hassanien. «Linear discriminant analysis: A detailed tutorial». In: 30 (2017), pp. 169–190.
- [42] L. Trevisan. «Lecture Notes on Expansion, Sparsest Cut and Spectral Graph Theory». In: (2013). URL: <http://www.eecs.berkeley.edu/~luca/books/expanders.pdf>.
- [43] X. Wang and I. Davidson. «Flexible Constrained Spectral Clustering». In: *Proceedings of the Sixteenth SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM. 2010, pp. 563–572.
- [44] X. Wang, B. Qian, and I. Davidson. «On Constrained Spectral Clustering and Its Applications». In: *Data Mining and Knowledge Discovery* 28.1 (2014), pp. 1–30.
- [45] M. Weirtheimer. «Laws of Organization in Perceptual Forms». In: *Psychologische Forschung* 4 (1923), pp. 301–350.
- [46] K. Wu and H. Simon. «Thick-Restart Lanczos Method for Large Symmetric Eigenvalue Problems». In: *SIAM Journal on Matrix Analysis and Applications* 22.2 (2000), pp. 602–616.
- [47] S. X. Yu and J. Shi. «Grouping with Bias». In: *Technical Report* (2001).
- [48] S. X. Yu and J. Shi. «Grouping with Bias». In: *Proceedings of the Fourteenth International Conference on Advances in Neural Information Processing Systems*. NIPS Foundation. 2001, pp. 3–8.
- [49] S. X. Yu and J. Shi. «Multiclass Spectral Clustering». In: *Proceedings of the Ninth International Conference on Computer Vision*. IEEE. 2003, pp. 313–320.
- [50] S. X. Yu and J. Shi. «Segmentation given partial grouping constraints». In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 26 (2004), pp. 173–183.
- [51] X. Zhu, Z. Ghahramani, and J. Lafferty. «Semi-Supervised Learning Using Gaussian Fields and Harmonic Functions». In: *Proceedings of the Twentieth International Conference on Machine Learning*. ACM. 2003, pp. 912–919.

