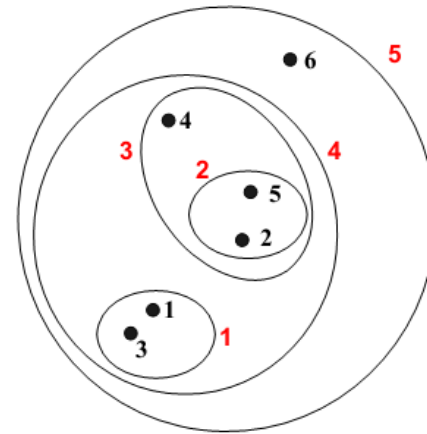
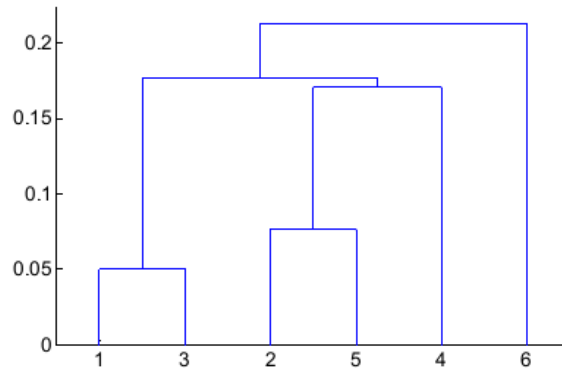


ΙΕΡΑΡΧΙΚΗ ΣΥΣΤΑΔΟΠΟΙΗΣΗ

Μπορεί να παρασταθεί με ένα **δένδρο-γραμμά**

Ένα διάγραμμα που μοιάζει με δένδρο και καταγράφει τις ακολουθίες από συγχωνεύσεις (merges) και διαχωρισμούς (splits)



Ιεραρχική Συσταδοποίηση: Πλεονεκτήματα

- Δε χρειάζεται να υποθέσουμε ένα συγκεκριμένο αριθμό από συστάδες

Οποιοσδήποτε επιθυμητός αριθμός από συστάδες μπορεί να επιτευχθεί κόβοντας το δένδρόγραμμα στο κατάλληλο επίπεδο

Δυο βασικοί τύποι ιεραρχικής συσταδοποίησης

- **Συσσωρευτικός (Agglomerative):**

- Αρχίζει με τα σημεία ως ξεχωριστές συστάδες
- Σε κάθε βήμα, συγχωνεύει το πιο κοντινό ζευγάρι συστάδων μέχρι να μείνει μόνο μία (ή k) συστάδες

- **Διαιρετικός (Divisive):**

- Αρχίζει με μία συστάδα που περιέχει όλα τα σημεία
- Σε κάθε βήμα, διαχωρίζει μία συστάδα, έως κάθε συστάδα να περιέχει μόνο ένα σημείο (ή να δημιουργηθούν k συστάδες)

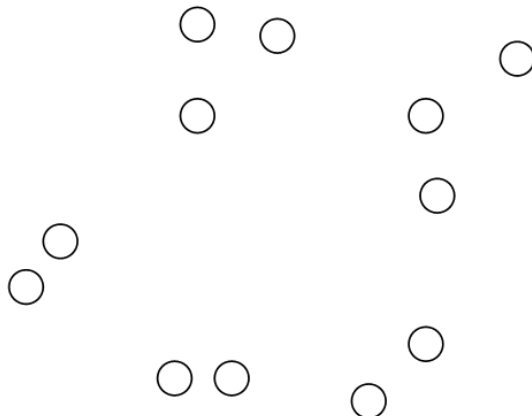
Βασικός Αλγόριθμος

- 1: Υπολογισμός του Πίνακα Γειτνίασης
- 2: Έστω κάθε σημείο αποτελεί και μια συστάδα
- 3: **Repeat**
- 4: Συγχώνευση των δύο κοντινότερων συστάδων
- 5: Ενημέρωση του Πίνακα Γειτνίασης
- 6: **Until** να μείνει μία μόνο συστάδα

Βασική λειτουργία είναι ο υπολογισμός της γειτνίασης δυο συστάδων

Διαφορετικοί αλγόριθμοι με βάση το πως ορίζεται η απόσταση ανάμεσα σε δύο συστάδες

Αρχικά: Κάθε σημείο και συστάδα και ένας Πίνακας Γειτνίασης (proximity matrix)

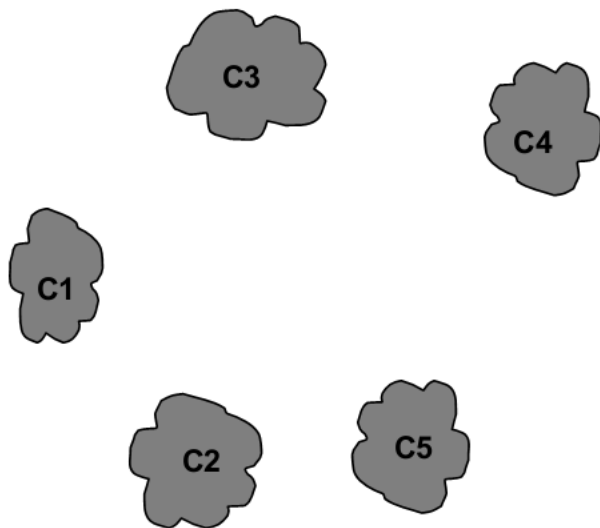


	p1	p2	p3	p4	p5	...
p1						
p2						
p3						
p4						
p5						
.						
.						
.						

Πίνακας Γειτνίασης

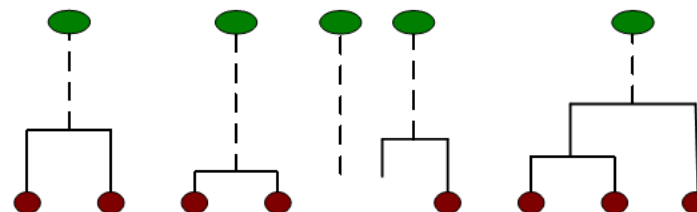


Μετά από κάποιες συγχωνεύσεις,
έχουμε κάποιες συστάδες

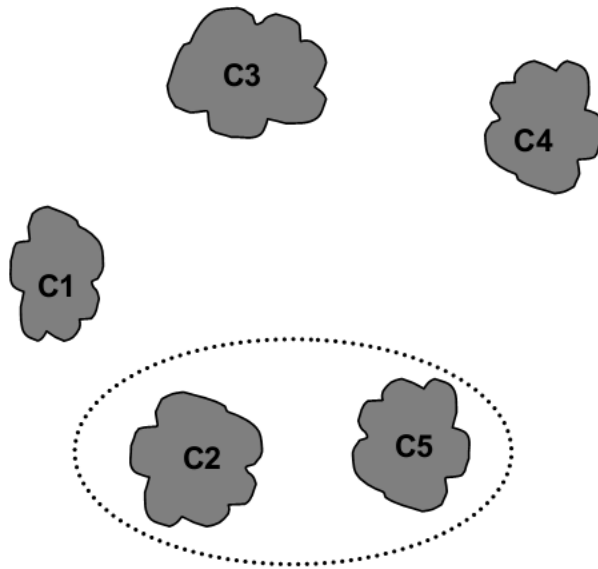


	C1	C2	C3	C4	C5
C1					
C2					
C3					
C4					
C5					

Πίνακας Γειτνίασης

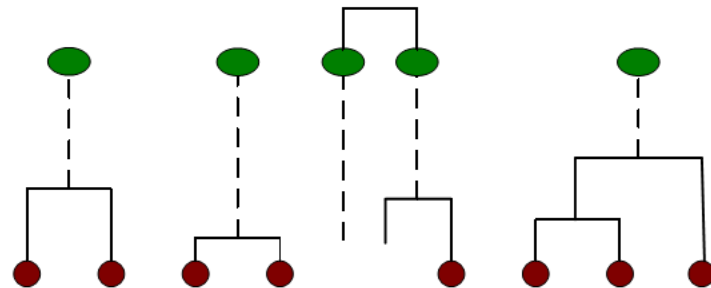


Θέλουμε να συγχωνεύσουμε τις δύο κοντινότερες συστάδες (C2 και C5) και να ενημερώσουμε τον πίνακα γειτνίασης.

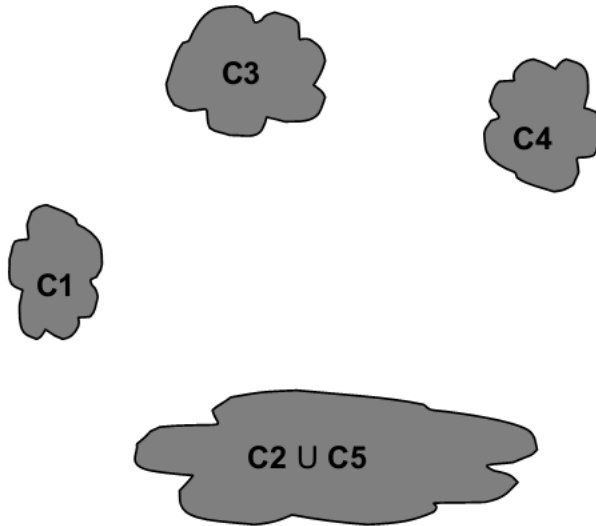


	C1	C2	C3	C4	C5
C1					
C2					
C3					
C4					
C5					

Πίνακας Γειτνίασης

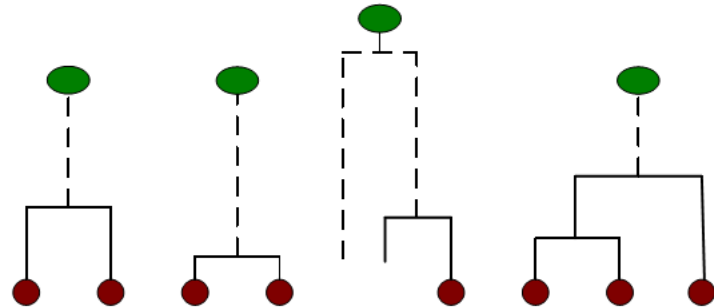


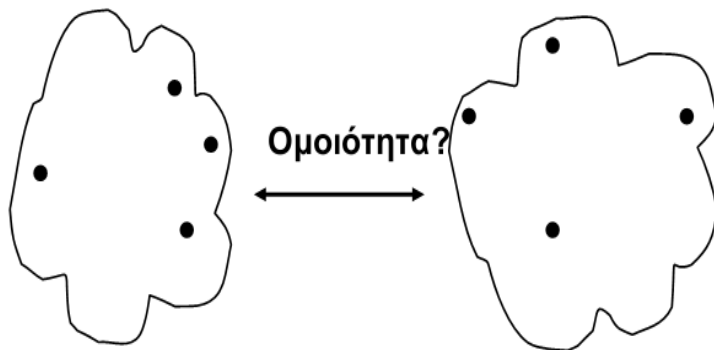
Μετά τη συγχώνευση η ερώτηση είναι: Πως ενημερώνουμε τον πίνακα γειτνίασης



	C1	C2 U C5	C3	C4
C1		?		
C2 U C5	?	?	?	?
C3		?		
C4		?		

Πίνακας Γειτνίασης

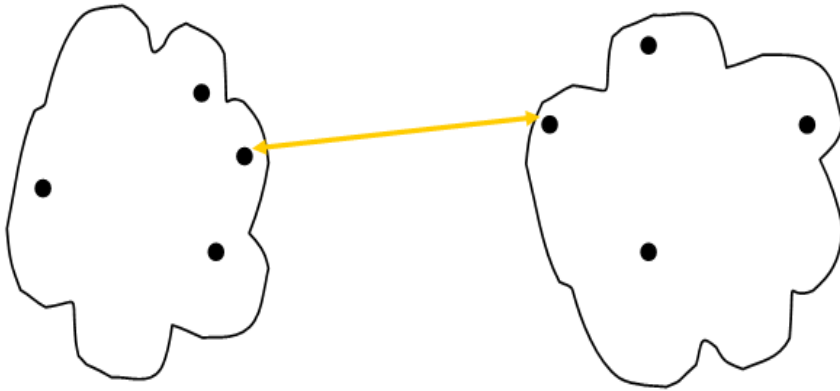




- MIN
- MAX
- Μέσος όρος της συστάδας
- Η απόσταση μεταξύ των κεντρικών σημείων
- Άλλες μέθοδοι βασισμένες σε μια αντικειμενική συνάρτηση
 - Η μέθοδος του Ward χρησιμοποιεί τετραγωνικά λάθη

	p1	p2	p3	p4	p5	...
p1						
p2						
p3						
p4						
p5						
.						
.						
.						

Πίνακας Γειτνίασης



- **MIN**
- MAX
- Μέσος όρος της ομάδας
- Η απόσταση μεταξύ των κεντρικών σημείων
- Άλλες μέθοδοι βασισμένες σε μια αντικειμενική συνάρτηση
 - Η μέθοδος του Ward χρησιμοποιεί τετραγωνικά λάθη

	p1	p2	p3	p4	p5	...
p1						
p2						
p3						
p4						
p5						
.						
.						
.						

Πίνακας Γειτνίασης

	I1	I2	I3	I4	I5
I1	1,00	0,90	0,10	0,65	0,20
I2	0,90	1,00	0,70	0,60	0,50
I3	0,10	0,70	1,00	0,40	0,30
I4	0,65	0,60	0,40	1,00	0,80
I5	0,20	0,50	0,30	0,80	1,00

