

**Δημοκρίτειο Πανεπιστήμιο Θράκης**  
**Σχολή Οικονομίας και Διοίκησης**  
**Τμήμα Διοικητικής Επιστήμης & Τεχνολογίας**



**ΔΗΜΟΚΡΙΤΕΙΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΘΡΑΚΗΣ** | **DEMOCRITUS UNIVERSITY OF THRACE**

---

**Εργασία εξαμήνου**

**Δίκτυα Chord & Ανάλυση Δεδομένων χρησιμοποιώντας τη γλώσσα R**

**Μάθημα: Ανάλυση Δεδομένων Μεγάλου Όγκου και Μέθοδοι Έρευνας  
(Analytics and Research Methods)**

**Υπεύθυνοι καθηγητές: Δρ Ιωάννης Γεροντίδης Δρ. Κούγκα Γεωργία,  
Δρ. Κρηνίδης Στυλιανός**

---

**Ονοματεπώνυμα Φοιτητριών**

**Γεωργία Κουμαντζιά( Α.Μ 222 )**

**Σεβαστή Χατζηδανιήλ (Α.Μ. 242)**

## Περιεχόμενα

|                                                       |    |
|-------------------------------------------------------|----|
| ΘΕΜΑ 1 <sup>ο</sup> .....                             | 3  |
| Εισαγωγή .....                                        | 3  |
| ΠΡΩΤΟ ΣΚΕΛΟΣ: Σχεδίαση CHORD Δικτύου.....             | 3  |
| ΔΕΥΤΕΡΟ ΣΚΕΛΟΣ: Finger Table για κάθε κόμβο .....     | 4  |
| ΤΡΙΤΟ ΣΚΕΛΟΣ: Αναζητήσεις Κλειδιών .....              | 6  |
| Περίληψη Διαδρομών:.....                              | 7  |
| ΤΕΤΑΡΤΟ ΣΚΕΛΟΣ: Οπτική Αναπαράσταση του Δικτύου ..... | 7  |
| Θέμα 2 <sup>ο</sup> .....                             | 8  |
| Εισαγωγή.....                                         | 8  |
| Βασικά Στατιστικά και Ανάλυση .....                   | 8  |
| Γράφημα Σπάνιων Προϊόντων .....                       | 10 |
| Μοτίβο αγορών καταναλωτή: .....                       | 10 |
| Εξόρυξη Κανόνων με Apriori .....                      | 11 |
| Top 5 Κανόνες με Μεγαλύτερο Confidence .....          | 11 |
| Σύγκριση Συχνών και Σπάνιων Προϊόντων .....           | 12 |
| Κανόνες με Συνέπεια το Προϊόν 'whole milk' .....      | 13 |
| Γραφική Απεικόνιση των Συνδυασμών για Γάλα .....      | 14 |
| Συμπεράσματα.....                                     | 14 |

## ΘΕΜΑ 1<sup>ο</sup>

### Εισαγωγή

Η παρούσα εργασία αφορά τη μελέτη και υλοποίηση ενός κατακεντρωμένου συστήματος δρομολόγησης τύπου CHORD. Το CHORD είναι ένας αλγόριθμος που επιτρέπει την αποδοτική αναζήτηση και εισαγωγή δεδομένων σε ένα αποκεντρωμένο σύστημα υπολογιστών (P2P), οργανώνοντας τους κόμβους σε έναν κυκλικό χώρο τιμών. Η εργασία βασίζεται σε χώρο 64 identifiers ( $2^6 = 64$ ), με το πλήθος των κόμβων να προκύπτει από τους αριθμούς μητρώου μας (AM. 242- 222). Το δίκτυο κατασκευάζεται βήμα-βήμα και περιλαμβάνει:

- Υπολογισμό και τοποθέτηση κόμβων στο κυκλικό δίκτυο
- Δημιουργία πίνακα δακτύλων (finger tables) με  $m = 6$
- Αναζητήσεις δύο κλειδιών ξεκινώντας από τον κόμβο N0
- Οπτική απεικόνιση του κύκλου και των δακτύλων του CHORD

### ΠΡΩΤΟ ΣΚΕΛΟΣ: Σχεδίαση CHORD Δικτύου

Στο πλαίσιο αυτής της άσκησης θα σχεδιάσουμε ένα CHORD δίκτυο με 64 identifiers (0 έως 63). Ο αριθμός των κόμβων στο δίκτυο καθορίζεται από τους αριθμούς μητρώου (AM) ως εξής:

$$AM1 = 242, AM2 = 222$$

$$\text{Αριθμός κόμβων} = ((AM1 + AM2) \bmod 8) + 8 = ((242 + 222) \bmod 8) + 8 = 0 + 8 = 8 \text{ κόμβοι.}$$

Οι 8 κόμβοι κατανέμονται ομοιόμορφα στον κύκλο των 64 identifiers με απόσταση 8 θέσεων μεταξύ τους. Οι κόμβοι είναι οι εξής:

| Κόμβος | Identifier |
|--------|------------|
| N0     | 0          |
| N1     | 8          |
| N2     | 16         |

|    |    |
|----|----|
| N3 | 24 |
| N4 | 32 |
| N5 | 40 |
| N6 | 48 |
| N7 | 56 |

### ΔΕΥΤΕΡΟ ΣΚΕΛΟΣ: Finger Table για κάθε κόμβο

Για κάθε κόμβο κατασκευάζεται πίνακας δαχτύλων (finger table) με  $m = 6$ . Για κάθε εγγραφή  $\text{finger}[i]$  υπολογίζουμε:  
 $\text{finger}[i] = (n + 2^{i-1}) \bmod 64$  και βρίσκουμε τον πρώτο κόμβο που είναι μεγαλύτερος ή ίσος του αποτελέσματος.

Αναλυτικά οι πίνακες finger tables φαίνονται παρακάτω:

N0:

| i | Start | Finger Node (Successor) |
|---|-------|-------------------------|
| 1 | 1     | N8                      |
| 2 | 2     | N8                      |
| 3 | 4     | N8                      |
| 4 | 8     | N8                      |
| 5 | 16    | N16                     |
| 6 | 32    | N32                     |

N16:

| i | Start | Finger Node (Successor) |
|---|-------|-------------------------|
| 1 | 17    | N24                     |
| 2 | 18    | N24                     |
| 3 | 20    | N24                     |
| 4 | 24    | N24                     |
| 5 | 32    | N32                     |

|   |    |     |
|---|----|-----|
| 6 | 48 | N48 |
|---|----|-----|

N24:

| i | Start | Finger Node (Successor) |
|---|-------|-------------------------|
| 1 | 25    | N32                     |
| 2 | 26    | N32                     |
| 3 | 28    | N32                     |
| 4 | 32    | N32                     |
| 5 | 40    | N40                     |
| 6 | 56    | N56                     |

N32:

| i | Start | Finger Node (Successor) |
|---|-------|-------------------------|
| 1 | 33    | N40                     |
| 2 | 34    | N40                     |
| 3 | 36    | N40                     |
| 4 | 40    | N40                     |
| 5 | 48    | N48                     |
| 6 | 0     | N0                      |

N40:

| i | Start | Finger Node (Successor) |
|---|-------|-------------------------|
| 1 | 41    | N48                     |
| 2 | 42    | N48                     |
| 3 | 44    | N48                     |
| 4 | 48    | N48                     |
| 5 | 56    | N56                     |
| 6 | 8     | N8                      |

N48:

| i | Start | Finger Node (Successor) |
|---|-------|-------------------------|
|---|-------|-------------------------|

|   |    |     |
|---|----|-----|
| 1 | 49 | N56 |
| 2 | 50 | N56 |
| 3 | 52 | N56 |
| 4 | 56 | N56 |
| 5 | 0  | N0  |
| 6 | 16 | N16 |

N56:

| i | Start | Finger Node (Successor) |
|---|-------|-------------------------|
| 1 | 57    | N0                      |
| 2 | 58    | N0                      |
| 3 | 60    | N0                      |
| 4 | 0     | N0                      |
| 5 | 8     | N8                      |
| 6 | 24    | N24                     |

N8:

| i | Start | Finger Node (Successor) |
|---|-------|-------------------------|
| 1 | 9     | N16                     |
| 2 | 10    | N16                     |
| 3 | 12    | N16                     |
| 4 | 16    | N16                     |
| 5 | 24    | N24                     |
| 6 | 40    | N40                     |

### ΤΡΙΤΟ ΣΚΕΛΟΣ: Αναζητήσεις Κλειδιών

Αναζητούμε δύο κλειδιά, K50 και K30, ξεκινώντας από τον κόμβο N0. Οι τιμές τους προκύπτουν ως εξής:

- $K50 = 242 \bmod 64 = 50$
- $K30 = 222 \bmod 64 = 30$

Αναζήτηση K50 (από N0):

- Το κλειδί 50 βρίσκεται μεταξύ των κόμβων N48 και N56, άρα υπεύθυνος είναι ο κόμβος N56.
- Από τον πίνακα του N0:  
→ Finger[6] οδηγεί στο N32, άρα επόμενο βήμα: N32
- Από N32:  
→ Finger[5] οδηγεί στο N48
- Από N48:  
→ Finger[1] οδηγεί στο N56

Τελικός υπεύθυνος κόμβος: N56

Αναζήτηση K30 (από N0):

- Το κλειδί 30 βρίσκεται μεταξύ των κόμβων N24 και N32, άρα υπεύθυνος είναι ο κόμβος N32.
- Από τον πίνακα του N0:  
→ Finger[6] οδηγεί απευθείας στο N32

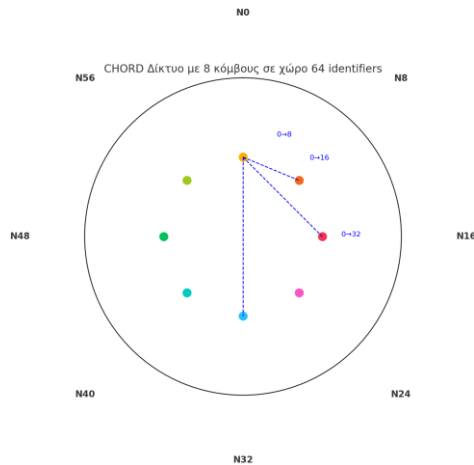
Τελικός υπεύθυνος κόμβος: N32

#### Περίληψη Διαδρομών:

| Κλειδί | Διαδρομή             |
|--------|----------------------|
| K50    | N0 → N32 → N48 → N56 |
| K30    | N0 → N32             |

## ΤΕΤΑΡΤΟ ΣΚΕΛΟΣ: Οπτική Αναπαράσταση του Δικτύου

Παρακάτω φαίνεται το κυκλικό διάγραμμα του CHORD δικτύου με 8 κόμβους σε χώρο 64 identifiers. Οι κόμβοι είναι τοποθετημένοι ισομερώς, και οι γραμμές δείχνουν μερικά παραδείγματα από τις διασυνδέσεις (fingers) του κόμβου N0.



## Θέμα 2°

### Εισαγωγή

Στην παρούσα εργασία μελετάμε αγοραστικά μοτίβα από ένα dataset που ονομάζεται "Groceries" και περιέχει περίπου 10.000 συναλλαγές από ένα σούπερ μάρκετ. Μέσω του αλγορίθμου Apriori που χρησιμοποιείται στην Εξόρυξη Δεδομένων, μπορούμε να βρούμε συνδυασμούς προϊόντων που αγοράζονται συχνά μαζί και να εξαγάγουμε χρήσιμους κανόνες.

Αρχικά εξετάσαμε βασικές πληροφορίες για το dataset, εντοπίσαμε ποια προϊόντα εμφανίζονται πιο συχνά, ποια είναι πιο σπάνια, και στη συνέχεια εφαρμόσαμε τον αλγόριθμο Apriori για να ανακαλύψουμε συσχετίσεις μεταξύ των προϊόντων, με ιδιαίτερη έμφαση στην αγορά γάλακτος ("whole milk").

### Βασικά Στατιστικά και Ανάλυση

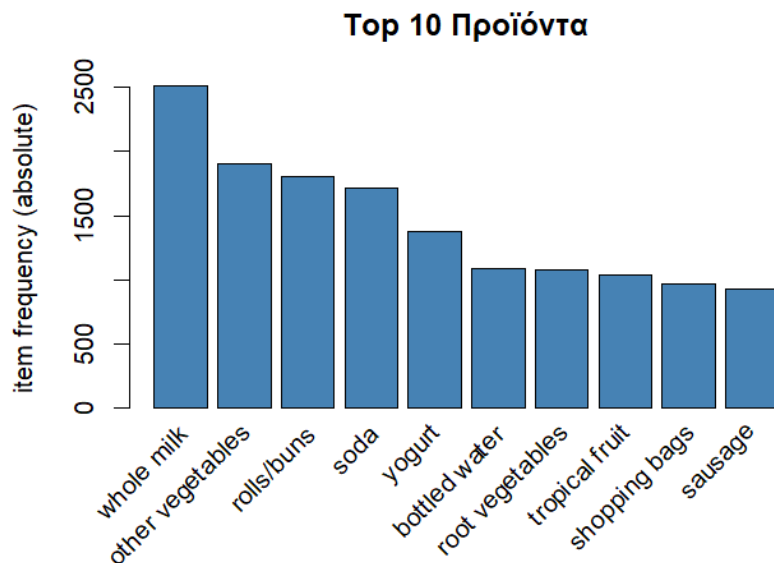
Το dataset αποτελείται από 9.835 συναλλαγές και περιλαμβάνει 169 διαφορετικά προϊόντα. Το πιο συχνό προϊόν είναι το "whole milk", το οποίο εμφανίζεται σε 2.513 συναλλαγές.

Επιπλέον, η πυκνότητα των δεδομένων είναι μόλις 2.6%, κάτι που επιβεβαιώνει ότι το σύνολο δεδομένων είναι αραιό (sparse). Αυτό είναι φυσιολογικό, καθώς κάθε συναλλαγή περιέχει μόνο ένα μικρό υποσύνολο των διαθέσιμων προϊόντων.

Χρησιμοποιώντας τις συναρτήσεις του R, εντοπίσαμε τα index τεσσάρων προϊόντων που ζητά η άσκηση:

- citrus fruit: θέση 14
- semi-finished bread: θέση 61
- margarine: θέση 70
- ready soups: θέση 79

Στη συνέχεια δημιουργήσαμε γράφημα με τα 10 πιο συχνά προϊόντα στο dataset, στο οποίο φαίνεται καθαρά ότι το γάλα, τα λαχανικά και τα ψωμάκια είναι τα προϊόντα που εμφανίζονται πιο συχνά.



Εικόνα 1: Τα 10 πιο συχνά αγορασμένα προϊόντα.

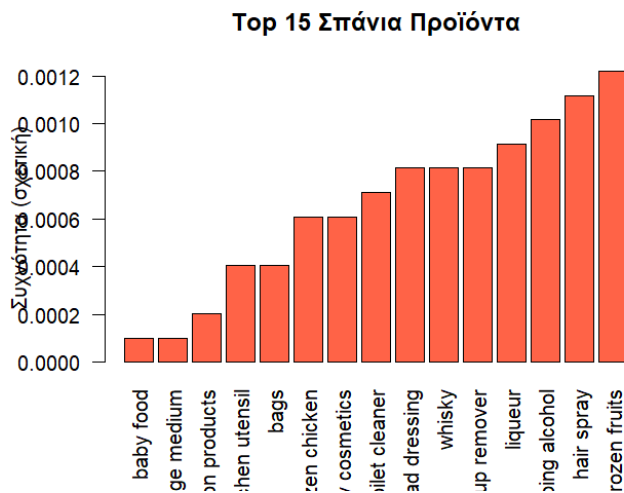
Έπειτα, υπολογίσαμε πόσα από τα προϊόντα αγοράζονται σπάνια, δηλαδή με συχνότητα μικρότερη του 1%. Βρήκαμε ότι 81 προϊόντα από τα 169 (σχεδόν τα μισά) είναι σπάνια.

Για να καταλάβουμε καλύτερα πόσα προϊόντα αγοράζει κάθε πελάτης, δημιουργήσαμε ένα ιστόγραμμα. Φάνηκε πως οι περισσότεροι πελάτες αγοράζουν 3 έως 5 προϊόντα ανά συναλλαγή.

## Γράφημα Σπάνιων Προϊόντων

Για να αναδείξουμε ποια προϊόντα αγοράζονται σπάνια, δημιουργήθηκε το παρακάτω γράφημα, στο οποίο απεικονίζονται τα 15 προϊόντα με τη χαμηλότερη σχετική συχνότητα εμφάνισης στο dataset (κάτω του 1%).

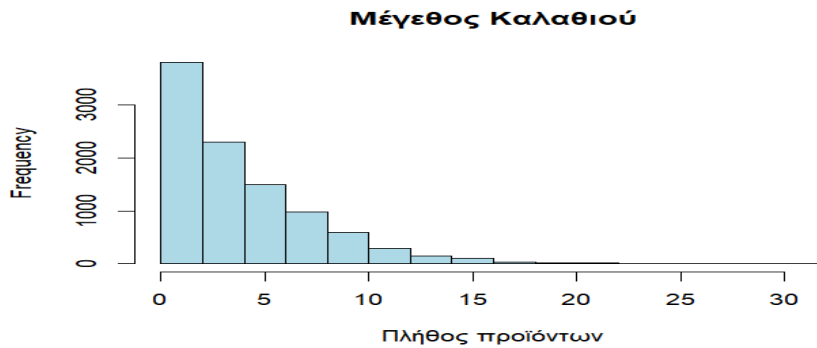
Αυτά τα προϊόντα εμφανίζονται ελάχιστες φορές στις συναλλαγές των πελατών, με πιο σπάνια τα "baby food", "package medium", "preservation products", "make up remover" κ.ά.



Εικόνα: Top 15 Σπάνια Προϊόντα στο dataset 'Groceries' (συχνότητα < 1%).

## Μοτίβο αγорών καταναλωτή:

Από την ανάλυση προκύπτει ότι τα περισσότερα καλάθια περιέχουν 1 έως 5 προϊόντα, με μέσο όρο περίπου 4,4 προϊόντα. Τα πιο συχνά προϊόντα είναι: whole milk, other vegetables, rolls/buns και soda. Οι καταναλωτές προτιμούν βασικά είδη πρώτης ανάγκης σε συνδυασμούς που επαναλαμβάνονται συχνά.



Εικόνα 2: Ιστόγραμμα μεγέθους καλαθιών (πλήθος προϊόντων ανά συναλλαγή).

### Εξόρυξη Κανόνων με Apriori

Εφαρμόσαμε τον αλγόριθμο Apriori με τις παραμέτρους που ζητούνται στην εκφώνηση:

- Support  $\geq 0.006$
- Confidence  $\geq 0.25$
- Ελάχιστο μήκος κανόνα: 2

Με αυτές τις ρυθμίσεις, εξήχθησαν 463 κανόνες. Στη συνέχεια εντοπίσαμε τους 5 ισχυρότερους (με το μεγαλύτερο confidence). Όλοι σχετίζονταν με το προϊόν "whole milk".

#### Top 5 Κανόνες με Μεγαλύτερο Confidence

Οι πέντε κανόνες με τη μεγαλύτερη εμπιστοσύνη εμφανίστηκαν με confidence έως 66%.

Ένας ενδεικτικός κανόνας:

{butter, bread}  $\rightarrow$  {whole milk}

Οι πέντε κανόνες με τη μεγαλύτερη εμπιστοσύνη είναι οι εξής:

1. {butter, whipped/sour cream}  $\rightarrow$  {whole milk}

- Confidence: 66%
- Lift: 2.58
- Count: 66 συναλλαγές

2. {butter, yogurt}  $\rightarrow$  {whole milk}

- Confidence: 63.89%
- Lift: 2.50
- Count: 92

3. {root vegetables, butter} → {whole milk}

- Confidence: 63.78%
- Lift: 2.50
- Count: 81

4. {tropical fruit, curd} → {whole milk}

- Confidence: 63.36%
- Lift: 2.48
- Count: 64

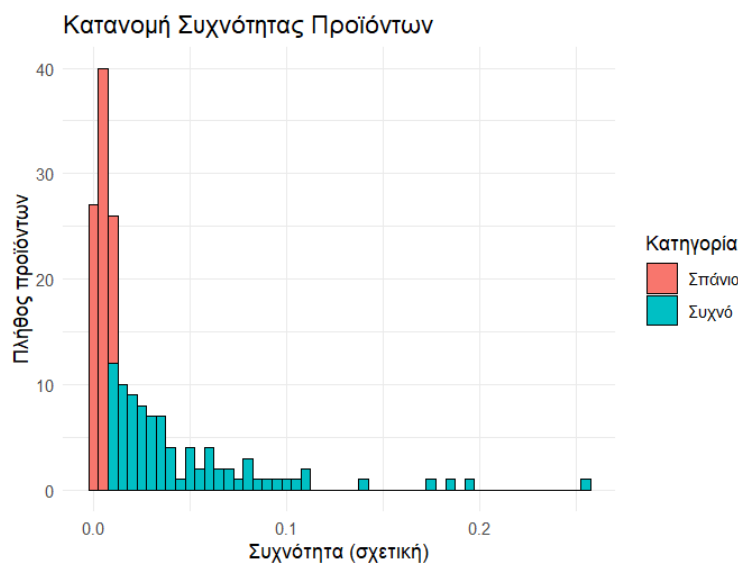
5. {tropical fruit, butter} → {whole milk}

- Confidence: 62.24%
- Lift: 2.43
- Count: 61

Όλοι οι παραπάνω κανόνες οδηγούν στην αγορά "whole milk", γεγονός που υποδεικνύει ότι το προϊόν αυτό είναι κεντρικό στις αγοραστικές συνήθειες.

### Σύγκριση Συχνών και Σπάνιων Προϊόντων

Για να δούμε πόσα συχνά και σπάνια προϊόντα υπάρχουν στο dataset, δημιουργήσαμε ένα γράφημα σύγκρισης. Χρησιμοποιήσαμε διαφορετικά χρώματα για να δείξουμε ποια προϊόντα εμφανίζονται πάνω ή κάτω από το όριο του 1%. Η πλειοψηφία των προϊόντων (81 από τα 169) έχουν σχετική συχνότητα κάτω από 1% και χαρακτηρίζονται ως «σπάνια», ενώ τα υπόλοιπα είναι πιο συχνά. Αυτό δείχνει πως οι καταναλωτές προτιμούν λίγα, δημοφιλή προϊόντα σε πολλές συναλλαγές.



Εικόνα 4: Σύγκριση προϊόντων ανάλογα με τη συχνότητα εμφάνισής τους.

### Κανόνες με Συνέπεια το Προϊόν 'whole milk'

Εστιάσαμε αποκλειστικά στους κανόνες που έχουν ως αποτέλεσμα την αγορά του γάλακτος ("whole milk"). Οι 5 σημαντικότεροι κανόνες δείχνουν ότι όταν ένας πελάτης αγοράζει butter, yogurt ή root vegetables, τότε είναι πολύ πιθανό να αγοράσει και γάλα.

Συνολικά, οι συνδυασμοί που οδηγούν στην αγορά γάλακτος έχουν υψηλό confidence (πάνω από 62%) και lift μεγαλύτερο του 2.4, γεγονός που δείχνει δυνατούς και σημαντικούς κανόνες.

Οι πέντε κανόνες με το μεγαλύτερο confidence είναι:

1. {butter, whipped/sour cream} → {whole milk}

- Confidence: 66%
- Lift: 2.58
- Count: 66

2. {butter, yogurt} → {whole milk}

- Confidence: 63.89%
- Lift: 2.50
- Count: 92

3. {root vegetables, butter} → {whole milk}

- Confidence: 63.78%
- Lift: 2.50
- Count: 81

4. {tropical fruit, curd} → {whole milk}

- Confidence: 63.36%
- Lift: 2.48
- Count: 64

5. {tropical fruit, butter} → {whole milk}

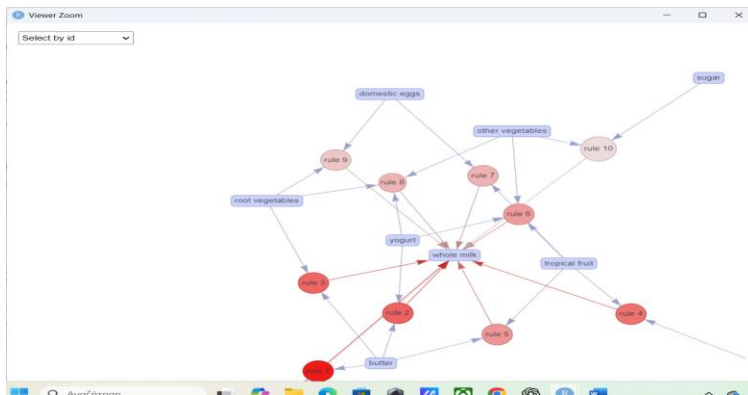
- Confidence: 62.24%
- Lift: 2.43
- Count: 61

## Γραφική Απεικόνιση των Συνδυασμών για Γάλα

Χρησιμοποιώντας την εντολή plot από το πακέτο arulesViz δημιουργήσαμε ένα γράφημα-δίκτυο, όπου εμφανίζονται οι 10 ισχυρότεροι συνδυασμοί προϊόντων που οδηγούν στην αγορά του "whole milk".

Ο κάθε κόμβος αναπαριστά ένα προϊόν ή έναν κανόνα, και οι κατευθυνόμενες γραμμές (ακμές) δείχνουν τη ροή προς το γάλα.

Προϊόντα όπως butter, yogurt, tropical fruit και root vegetables εμφανίζονται συχνά ως πρόδρομοι της αγοράς γάλακτος.



Εικόνα 3: Γραφική απεικόνιση των κανόνων που οδηγούν στην αγορά του 'whole milk'.

## Συμπεράσματα

Από την ανάλυση φάνηκε ότι το γάλα ("whole milk") είναι το προϊόν που αγοράζεται περισσότερο. Η πλειοψηφία των πελατών αγοράζει λίγα και βασικά προϊόντα, ενώ σχεδόν τα μισά προϊόντα είναι σπάνια.

Ο αλγόριθμος Apriori μας βοήθησε να βρούμε ποιοι συνδυασμοί προϊόντων οδηγούν συχνά στην αγορά γάλακτος. Οι σημαντικότεροι συνδυασμοί περιλαμβάνουν το butter, το yogurt και άλλα βασικά είδη. Οι κανόνες που προέκυψαν έχουν υψηλό confidence, κάτι που τους καθιστά χρήσιμους για εμπορικές στρατηγικές.

Γενικά, η μεθοδολογία αυτή μπορεί να χρησιμοποιηθεί σε πραγματικά δεδομένα σούπερ μάρκετ ώστε να βελτιωθεί η τοποθέτηση προϊόντων ή να γίνουν πιο στοχευμένες προωθήσεις.