



ΠΡΟΓΡΑΜΜΑ ΜΕΤΑΠΤΥΧΙΑΚΩΝ ΣΠΟΥΔΩΝ
ΨΗΦΙΑΚΟ ΜΑΡΚΕΤΙΝΓΚ (DIGITAL MARKETING)

ΕΡΓΑΣΙΑ

PROJECT (ARM P2)

Δίκτυα Chord & Ανάλυση Δεδομένων
χρησιμοποιώντας τη γλώσσα R

ΤΩΝ ΜΕΤΑΠΤΥΧΙΑΚΩΝ ΦΟΙΤΗΤΩΝ

Ζωή-Όλγα Καραγιάννη

Δήμητρα Περδίκη

Καβάλα 2025

Περιεχόμενα

ΘΕΜΑ 1 ^ο	3
ΘΕΜΑ 2 ^ο	6

Πίνακας εικόνων

ΕΙΚΟΝΑ 1	3
ΕΙΚΟΝΑ 2	7
ΕΙΚΟΝΑ 3	7
ΕΙΚΟΝΑ 4	8
ΕΙΚΟΝΑ 5	9
ΕΙΚΟΝΑ 6	9
ΕΙΚΟΝΑ 7	10
ΕΙΚΟΝΑ 8	11
ΕΙΚΟΝΑ 9	12
ΕΙΚΟΝΑ 10	13
ΕΙΚΟΝΑ 11	14

ΘΕΜΑ 1^ο

Σκοπός της παρούσας εργασίας είναι ο σχεδιασμός και η ανάλυση ενός καταναμεμημένου δικτύου τύπου CHORD, το οποίο αποτελεί ένα από τα πιο γνωστά πρωτόκολλα καταναμεμημένων hash πινάκων. Το δίκτυο θα βασίζεται σε 64 identifiers (2^6) και θα περιλαμβάνει έναν αριθμό κόμβων που υπολογίζεται από τον τύπο:

$$\text{Αριθμός κόμβων} = ((AM1 + AM2) \bmod 8) + 8$$

Οι θέσεις των κόμβων επιλέγονται ομοιόμορφα από τον κυκλικό χώρο των 64 θέσεων, σύμφωνα με το συνοδευτικό αρχείο Κόμβοι.xlsx. Για τους αριθμούς μητρώου που χρησιμοποιούνται στην παρούσα εργασία ($AM1 = 236$ και $AM2 = 238$), προκύπτουν 10 κόμβοι: N0, N5, N11, N16, N22, N27, N33, N38, N44, N49.

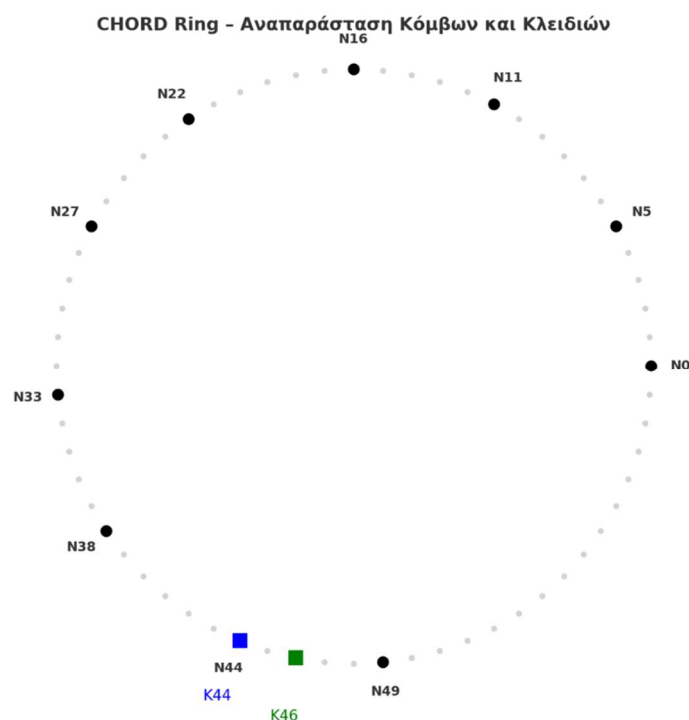
Στη συνέχεια, υπολογίζονται οι πίνακες finger table για κάθε κόμβο του δικτύου με $m = 6$, δηλαδή 6 γραμμές ανά πίνακα. Τέλος, περιγράφεται αναλυτικά η διαδικασία αναζήτησης δύο κλειδιών, KX και KΨ, όπου:

- $X = AM1 \bmod 64 = 44$
- $\Psi = AM2 \bmod 64 = 46$

Η αναζήτηση ξεκινά και στις δύο περιπτώσεις από τον κόμβο N0, και περιγράφεται βήμα-βήμα ο τρόπος με τον οποίο το δίκτυο εντοπίζει τον υπεύθυνο κόμβο για κάθε κλειδί.

Για $AM1 = 236$, $AM2 = 238$ θα έχουμε: $((236 + 238) \bmod 8) + 8 = (474 \bmod 8) + 8 = 2 + 8 = 10$ κόμβοι

Οι θέσεις των κόμβων είναι: N0, N5, N11, N16, N22, N27, N33, N38, N44, N49



ΕΙΚΟΝΑ 1

Στην συνέχεια, υπολογίζουμε το finger table για κάθε κόμβο για $m=6$ fingers.

Για τον κόμβο N0

$$N0+2^0 = 1 \rightarrow N5$$

$$N0+2^1 = 2 \rightarrow N5$$

$$N0+2^2 = 4 \rightarrow N5$$

$$N0+2^3 = 8 \rightarrow N11$$

$$N0+2^4 = 16 \rightarrow N16$$

$$N0+2^5 = 32 \rightarrow N33$$

Για τον κόμβο N5

$$N5+2^0 = 6 \rightarrow N11$$

$$N5+2^1 = 7 \rightarrow N11$$

$$N5+2^2 = 9 \rightarrow N11$$

$$N5+2^3 = 13 \rightarrow N16$$

$$N5+2^4 = 21 \rightarrow N22$$

$$N5+2^5 = 37 \rightarrow N38$$

Για τον κόμβο N11

$$N11+2^0 = 12 \rightarrow N16$$

$$N11+2^1 = 13 \rightarrow N16$$

$$N11+2^2 = 15 \rightarrow N16$$

$$N11+2^3 = 19 \rightarrow N22$$

$$N11+2^4 = 27 \rightarrow N27$$

$$N11+2^5 = 43 \rightarrow N44$$

Για τον κόμβο N16

$$N16+2^0 = 17 \rightarrow N22$$

$$N16+2^1 = 18 \rightarrow N22$$

$$N16+2^2 = 20 \rightarrow N22$$

$$N16+2^3 = 24 \rightarrow N27$$

$$N16+2^4 = 32 \rightarrow N33$$

$$N16+2^5 = 48 \rightarrow N49$$

Για τον κόμβο N22

$$N22+2^0 = 23 \rightarrow N27$$

$$N22+2^1 = 24 \rightarrow N27$$

$$N22+2^2 = 26 \rightarrow N27$$

$$N22+2^3 = 30 \rightarrow N33$$

$$N22+2^4 = 38 \rightarrow N38$$

$$N22+2^5 = 54 \rightarrow N0$$

Το N54 δεν υπάρχει, επόμενο είναι το N0.

Για τον κόμβο N27

$$N27+2^0 = 28 \rightarrow N33$$

$$N27+2^1 = 29 \rightarrow N33$$

$$N27+2^2 = 31 \rightarrow N33$$

$$N27+2^3 = 35 \rightarrow N38$$

$$N27+2^4 = 43 \rightarrow N44$$

$$N27+2^5 = 59 \rightarrow N0$$

Το N59 δεν υπάρχει, επόμενο είναι το N0.

Για τον κόμβο N33

$$N33+2^0 = 34 \rightarrow N38$$

$$N33+2^1 = 35 \rightarrow N38$$

$$N33+2^2 = 37 \rightarrow N38$$

$$N33+2^3 = 41 \rightarrow N44$$

$$N33+2^4 = 49 \rightarrow N49$$

$$N33+2^5 = 1 \rightarrow N5$$

Για τον κόμβο N38

$$N38+2^0 = 39 \rightarrow N44$$

$$N38+2^1 = 40 \rightarrow N44$$

$$N38+2^2 = 42 \rightarrow N44$$

$$N38+2^3 = 46 \rightarrow N49$$

$$N38+2^4 = 54 \rightarrow N0$$

$$N38+2^5 = 6 \rightarrow N11$$

Για τον κόμβο N44

$$N44+2^0 = 45 \rightarrow N49$$

$$N44+2^1 = 46 \rightarrow N49$$

$$N44+2^2 = 48 \rightarrow N49$$

$$N44+2^3 = 52 \rightarrow N0$$

$$N44+2^4 = 60 \rightarrow N0$$

$$N44+2^5 = 12 \rightarrow N16$$

Για τον κόμβο N49

$$N49+2^0 = 50 \rightarrow N0$$

$$N49+2^1 = 51 \rightarrow N0$$

$$N49+2^2 = 53 \rightarrow N0$$

$$N49+2^3 = 57 \rightarrow N0$$

$$N49+2^4 = 65 \rightarrow (65 \bmod 64 = 1) \rightarrow N5$$

$$N49+2^5 = 81 \rightarrow (81 \bmod 64 = 17) \rightarrow N22$$

$N65 > N64$ επομένως για τον λόγο αυτό πήραμε το υπόλοιπο της διαίρεσης $65 \bmod 64$.

Για την αναζήτηση των Κλειδιών KX και KΨ οι αριθμοί μητρώου είναι:

- $AM1 = 236 \rightarrow X = 236 \bmod 64 = 44$
- $AM2 = 238 \rightarrow \Psi = 238 \bmod 64 = 46$

Άρα πρέπει να εντοπίσουμε πού αποθηκεύονται τα κλειδιά K44 και K46, ξεκινώντας από τον κόμβο N0. Για την αναζήτηση του Κλειδιού K44 ξεκινάμε από τον κόμβο N0. Το κλειδί K44 βρίσκεται μετά από το N33, άρα το finger table του N0 δεν το καλύπτει απευθείας. Από το finger table του N0 το μεγαλύτερο $\text{finger} \leq 44$ είναι το N33. Πηγαίνουμε στον N33, όπου από το finger table του N33 έχει $\text{start}[4] = 41 \rightarrow \text{successor}: N44$. Πηγαίνουμε στον N44 όπου το K44 βρίσκεται στο διάστημα $(38, 44]$, άρα ο N44 είναι υπεύθυνος. Άρα το κλειδί K44 αποθηκεύεται στον κόμβο N44.

Προχωράμε στην αναζήτηση του Κλειδιού K46 και ξεκινάμε από N0. Από το finger table του N0, πάλι το πλησιέστερο $\text{finger} \leq 46$ είναι ο N33. Πηγαίνουμε στον N33, με το K46 βρίσκεται μεταξύ 44 και 49. Από το finger table του N33 $\text{start}[4] = 41 \rightarrow \text{successor}: \text{N44}$. Πηγαίνουμε στον N44 και ελέγχουμε: $K46 \in (44, 49]$ άρα το K46 ανήκει στον N49. Επομένως το κλειδί K46 αποθηκεύεται στον κόμβο N49.

Ανακεφαλαίωση

Κλειδί	Τιμή	Υπεύθυνος Κόμβος	Διαδρομή
K44	44	N44	N0 $\rightarrow$ N33 $\rightarrow$ N44
K46	46	N49	N0 $\rightarrow$ N33 $\rightarrow$ N44 $\rightarrow$ N49

ΘΕΜΑ 2^ο

Στο πλαίσιο της παρούσας ανάλυσης, χρησιμοποιήσαμε το ενσωματωμένο dataset Groceries, το οποίο περιλαμβάνεται στο πακέτο `arules` της γλώσσας προγραμματισμού R (<https://rdrr.io/cran/arules/>). Πρόκειται για ένα σύνολο δεδομένων που αποτελείται από καλάθια συναλλαγών καταναλωτών σε ένα σούπερ μάρκετ, με σκοπό την εξαγωγή κανόνων συσχέτισης μεταξύ των προϊόντων που αγοράζονται.

Οι εντολές που εκτελέστηκαν στο περιβάλλον RStudio ήταν οι εξής:

```
install.packages("arules")
```

```
install.packages("arulesViz")
```

και έπειτα

```
library(arules)
```

```
library(arulesViz)
```

ώστε να φορτώσουν τα πακέτα.

Συνεχίζοντας, με την `summary(Groceries)` ήμασταν σε θέση να δούμε όλες τις συναλλαγές, όπως φαίνεται στην παρακάτω εικόνα:

```

> summary(Groceries)
transactions as itemMatrix in sparse format with
9835 rows (elements/itemsets/transactions) and
169 columns (items) and a density of 0.02609146

most frequent items:
      whole milk other vegetables      rolls/buns      soda
      2513          1903          1809          1715
      yogurt          (Other)
      1372          34055

element (itemset/transaction) length distribution:
sizes
 1    2    3    4    5    6    7    8    9   10   11   12   13   14   15   16
2159 1643 1299 1005  855  645  545  438  350  246  182  117  78   77   55   46
 17   18   19   20   21   22   23   24   26   27   28   29   32
 29   14   14    9   11    4    6    1    1    1    1    3    1

      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
      1.000  2.000  3.000  4.409  6.000 32.000

includes extended item information - examples:
      labels level2      levell
1 frankfurter sausage meat and sausage
2   sausage sausage meat and sausage
3  liver loaf sausage meat and sausage
> |

```

EIKONA 2

Το αρχείο δεδομένων Groceries περιλαμβάνει στοιχεία για 9835 συναλλαγές καταναλωτών σε μορφή καλαθιών αγορών. Καταγράφηκαν 169 μοναδικά προϊόντα. Η πυκνότητα των δεδομένων είναι πολύ χαμηλή 0,026, κάτι που σημαίνει ότι οι περισσότερες συναλλαγές περιλαμβάνουν λίγα μόνο προϊόντα. Τα πιο συχνά προϊόντα είναι καθημερινά είδη, όπως whole milk, other vegetables, rolls/buns και soda, γεγονός που μας δίνει μία πρώτη εικόνα της αγοραστικής συμπεριφοράς.

Προχωρώντας, για να εντοπίσουμε το index των προϊόντων citrus fruit, semi-finished bread, margarine, ready soups χρησιμοποιήσαμε την εντολή `which(itemLabels(Groceries) == "<όνομα προϊόντος>")` και λάβαμε τα αποτελέσματα citrus fruit: 14, semi-finished bread: 61, margarine: 70 και ready soups: 79

```

> which(itemLabels(Groceries) == "citrus fruit")
[1] 14
> which(itemLabels(Groceries) == "semi-finished bread")
[1] 61
> which(itemLabels(Groceries) == "margarine")
[1] 70
> which(itemLabels(Groceries) == "ready soups")
[1] 79
> |

```

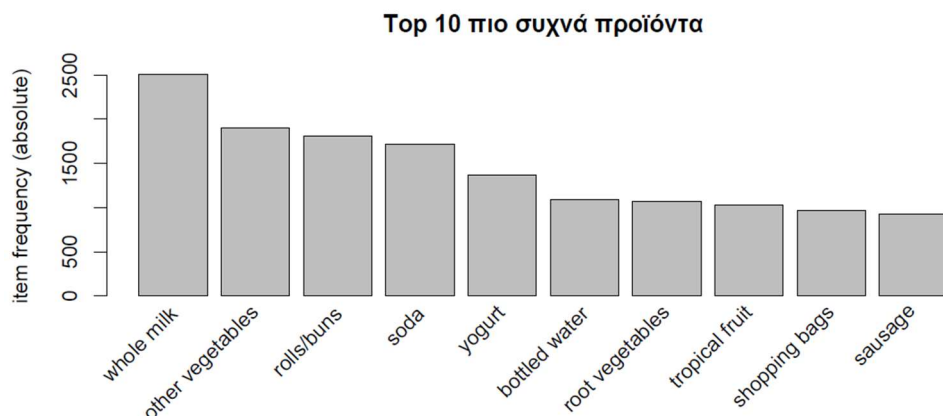
EIKONA 3

Οι αριθμοί αυτοί αντιπροσωπεύουν τις θέσεις των προϊόντων στο σύνολο των 169 διαθέσιμων ειδών.

Έχοντας ως σκοπό να εντοπίσουμε ποια προϊόντα εμφανίζονται συχνότερα στο σύνολο δεδομένων Groceries, χρησιμοποιήσαμε τη συνάρτηση `itemFrequency()` από το πακέτο `arules`, η οποία υπολογίζει τη συχνότητα εμφάνισης κάθε προϊόντος στο σύνολο των συναλλαγών. Στη συνέχεια, με τη συνάρτηση `itemFrequencyPlot()` δημιουργήσαμε τη σχετική γραφική παράσταση. Η εντολή που χρησιμοποιήσαμε ήταν:

```
itemFrequencyPlot(Groceries, topN = 10, type = "absolute", main = "Top 10 πιο συχνά προϊόντα")
```

Το παρακάτω γράφημα απεικονίζει τα 10 προϊόντα με την υψηλότερη συχνότητα εμφάνισης στο dataset. Παρατηρούμε ότι το “whole milk” είναι το προϊόν που εμφανίζεται συχνότερα, καθώς περιλαμβάνεται σε πάνω από 2500 συναλλαγές. Ακολουθούν τα προϊόντα “other vegetables”, “rolls/buns”, “soda”, “yogurt”, “bottled water”, “root vegetables”, “tropical fruit”, “shopping bags” και “sausage”. Η υψηλή συχνότητα εμφάνισης του πλήρους γάλακτος υποδηλώνει ότι αποτελεί βασικό στοιχείο στις αγοραστικές συνήθειες των καταναλωτών, καθώς εντοπίζεται συχνά σε συνδυασμό με άλλα καθημερινά προϊόντα.



ΕΙΚΟΝΑ 4

Προκειμένου να γίνει αναγνώριση των προϊόντων που αγοράζονται σπάνια από τους καταναλωτές, υπολογίσαμε τη συχνότητα εμφάνισης κάθε προϊόντος στο σύνολο συναλλαγών. Αρχικά, χρησιμοποιήσαμε την εντολή:

```
item_freq <- itemFrequency(Groceries, type = "absolute")
```

η οποία υπολογίζει την απόλυτη συχνότητα κάθε προϊόντος — δηλαδή πόσες φορές εμφανίζεται σε όλες τις συναλλαγές του dataset. Το αποτέλεσμα αποθηκεύεται στο αντικείμενο `item_freq`. Στη συνέχεια, ταξινομήσαμε τις συχνότητες κατά αύξουσα σειρά με την εντολή:

```
rare_items <- sort(item_freq, decreasing = FALSE)
```

έτσι ώστε τα προϊόντα με τη μικρότερη συχνότητα να βρίσκονται στην κορυφή. Τέλος, με την εντολή:

```
head(rare_items, 10)
```

εμφανίσαμε τα 10 προϊόντα με τις χαμηλότερες τιμές. Τα αποτελέσματα δείχνουν ότι τα προϊόντα που εμφανίζονται πιο σπάνια στις συναλλαγές είναι:

- baby food (1 εμφάνιση)
- sound storage medium (1 εμφάνιση)

- preservation products (2 εμφανίσεις)
- kitchen utensil, bags (4 εμφανίσεις)
- frozen chicken, baby cosmetics (6 εμφανίσεις)
- toilet cleaner (7 εμφανίσεις)
- salad dressing, whisky (8 εμφανίσεις)

Παρατηρούμε ότι τα παραπάνω προϊόντα δεν περιλαμβάνονται συχνά στα καλάθια των αγοραστών, γεγονός που τα κατατάσσει στις σπάνιες επιλογές. Τα περισσότερα από αυτά είναι είτε εξειδικευμένα είδη (π.χ. “sound storage medium”, “toilet cleaner”) είτε προϊόντα που δεν αγοράζονται σε καθημερινή βάση (όπως “whisky” ή “baby food”).

```
> item_freq <- itemFrequency(Groceries, type = "absolute")
> rare_items <- sort(item_freq, decreasing = FALSE)
> head(rare_items, 10)
```

baby food	1	sound storage medium	1	preservation products	2	kitchen utensil	4	bags	4
frozen chicken	6	baby cosmetics	6	toilet cleaner	7	salad dressing	8	whisky	8

EIKONA 5

Για την αξιολόγηση της δομής του συνόλου δεδομένων Groceries, υπολογίσαμε την πυκνότητα και την αραιότητα του πίνακα συναλλαγών. Αρχικά, προσδιορίσαμε τον συνολικό αριθμό στοιχείων του πίνακα με τη συνάρτηση `prod(dim(Groceries))` και στη συνέχεια υπολογίσαμε τον αριθμό των μη μηδενικών στοιχείων με τη συνάρτηση `length(Groceries)`, που επιστρέφει το πλήθος των προϊόντων που έχουν καταγραφεί σε συναλλαγές. Η πυκνότητα του πίνακα προκύπτει ως εξής:

```
density <- non_zero_elements / total_elements
```

και ισούται με 0.0059, δηλαδή μόλις το 0.59% των κελιών περιέχει προϊόντα. Η υπόλοιπη έκταση του πίνακα είναι κενή, κάτι που φαίνεται και από την αραιότητα:

```
sparsity <- 1 - density
```

η οποία ισούται με 0.9940 (99.4%). Αυτό καταδεικνύει ότι πρόκειται για έναν εξαιρετικά αραιό πίνακα, όπως είναι αναμενόμενο σε datasets που περιγράφουν αγορές καταναλωτών, καθώς κάθε συναλλαγή περιλαμβάνει μικρό αριθμό προϊόντων σε σχέση με το συνολικό πλήθος των διαθέσιμων ειδών.

```
> total_elements <- prod(dim(Groceries))
> non_zero_elements <- length(Groceries)
> density <- non_zero_elements / total_elements
> sparsity <- 1 - density
> cat("Η πυκνότητα των δεδομένων είναι:", density, "\n")
Η πυκνότητα των δεδομένων είναι: 0.00591716
> cat("Η αραιότητα των δεδομένων είναι:", sparsity, "\n")
Η αραιότητα των δεδομένων είναι: 0.9940828
> |
```

EIKONA 6

Η συνάρτηση `dim()` επιστρέφει τις διαστάσεις του dataset, ενώ η `prod()` πολλαπλασιάζει τις διαστάσεις για να υπολογίσει το συνολικό πλήθος κελιών του πίνακα. Η `length()` μετρά τον συνολικό αριθμό αγορασμένων προϊόντων. Η πυκνότητα προκύπτει από τον λόγο των

μη μηδενικών στοιχείων προς το συνολικό πλήθος, ενώ η αραιότητα είναι το συμπληρωματικό ποσοστό. Οι τιμές εμφανίζονται με τη συνάρτηση cat().

Στο επόμενο βήμα, εφαρμόσαμε τον αλγόριθμο Apriori προκειμένου να εντοπίσουμε κανόνες συσχέτισης μεταξύ προϊόντων στο σύνολο δεδομένων Groceries. Ο Apriori είναι μια ευρέως χρησιμοποιούμενη μέθοδος εξόρυξης συχνών προτύπων, ιδανική για αναλύσεις τύπου "αν αγοράζεται το προϊόν A, τότε πιθανότατα αγοράζεται και το B".

Χρησιμοποιήσαμε τις εξής παραμέτρους:

- Ελάχιστη υποστήριξη (support) = 0.006
- Ελάχιστη εμπιστοσύνη (confidence) = 0.25
- Ελάχιστο μήκος κανόνα (minlen) = 2

Ο κώδικας που εφαρμόστηκε ήταν:

```
rules <- apriori(Groceries, parameter = list(supp = 0.006, conf = 0.25, minlen = 2))
```

Το αποτέλεσμα ήταν η δημιουργία ενός συνόλου κανόνων που αποτυπώνουν αγοραστικά μοτίβα, τα οποία μπορούν να αξιοποιηθούν για στοχευμένες στρατηγικές προώθησης και διαχείρισης αποθέματος.

```
> # Εφαρμογή του αλγορίθμου Apriori στο dataset Groceries
> rules <- apriori(Groceries,
+               parameter = list(supp = 0.006, conf = 0.25, minlen = 2))
Apriori

Parameter specification:
 confidence minval smax arem aval originalSupport maxtime support minlen maxlen target ext
      0.25    0.1    1 none FALSE              TRUE        5   0.006     2    10 rules TRUE

Algorithmic control:
 filter tree heap memopt load sort verbose
  0.1 TRUE TRUE  FALSE TRUE    2    TRUE

Absolute minimum support count: 59

set item appearances ...[0 item(s)] done [0.00s].
set transactions ...[169 item(s), 9835 transaction(s)] done [0.00s].
sorting and recoding items ... [109 item(s)] done [0.00s].
creating transaction tree ... done [0.00s].
checking subsets of size 1 2 3 4 done [0.00s].
writing ... [463 rule(s)] done [0.00s].
creating S4 object ... done [0.00s].
> |
```

EIKONA 7

Στη συνέχεια, εντοπίστηκαν οι πέντε (5) κανόνες με το υψηλότερο επίπεδο εμπιστοσύνης (confidence) από το σύνολο των 463 κανόνων που εξήχθησαν με τον αλγόριθμο Apriori. Η ταξινόμηση έγινε με βάση το confidence σε φθίνουσα σειρά, με την εντολή:

```
top_conf_rules <- sort(rules, by = "confidence", decreasing = TRUE)
```

```
inspect(head(top_conf_rules, 5))
```

Οι κανόνες αυτοί δείχνουν ποια προϊόντα οδηγούν σταθερά στην αγορά πλήρους γάλακτος (whole milk). Συγκεκριμένα:

- Ο πρώτος κανόνας δείχνει ότι όταν ένας πελάτης αγοράζει βούτυρο (butter) και κρέμα (whipped/sour cream), υπάρχει 66% πιθανότητα να αγοράσει και πλήρες

γάλα. Το lift του κανόνα είναι 2.58, που σημαίνει ότι η πιθανότητα είναι 2,58 φορές μεγαλύτερη από την τυχαία συνύπαρξη.

- Στον δεύτερο κανόνα, η αγορά βουτύρου και γιαουρτιού οδηγεί στην αγορά γάλακτος με 63.88% πιθανότητα (confidence = 0.638889), και lift = 2.50.
- Ο τρίτος κανόνας περιλαμβάνει ριζώδη λαχανικά (root vegetables) και βούτυρο, και οδηγεί σε αγορά πλήρους γάλακτος με confidence 0.637953 και lift 2.49. Αυτό δείχνει ότι πρόκειται για έναν πολύ κοινό συνδυασμό στην καθημερινή αγορά.
- Ο τέταρτος κανόνας δείχνει ότι η αγορά τροπικών φρούτων (tropical fruit) και τυριού τύπου cottage (curd) οδηγεί σε αγορά πλήρους γάλακτος με πιθανότητα 63.36%, και lift 2.47.
- Τέλος, ο πέμπτος κανόνας συνδέει την αγορά τροπικών φρούτων και βουτύρου με το πλήρες γάλα, με 62.24% πιθανότητα και lift 2.43.

Παρατηρούμε ότι σε όλους τους κανόνες, η τιμή του lift είναι σημαντικά υψηλότερη από το 1, γεγονός που επιβεβαιώνει την ύπαρξη ισχυρής συσχέτισης ανάμεσα στα προϊόντα. Το πλήρες γάλα φαίνεται να αποτελεί κεντρικό προϊόν στις καταναλωτικές συνήθειες, καθώς σχετίζεται σταθερά με βασικά προϊόντα όπως το βούτυρο, τα γαλακτοκομικά και συγκεκριμένα φρούτα και λαχανικά.

Για την καλύτερη κατανόηση των συσχετίσεων μεταξύ προϊόντων που αγοράζονται συχνά και αυτών που αγοράζονται σπάνια, δημιουργήθηκε μία γραφική παράσταση τύπου “association graph” μέσω του πακέτου arulesViz. Χρησιμοποιήθηκαν οι πέντε κανόνες με το υψηλότερο confidence.

Ο κώδικας που χρησιμοποιήθηκε ήταν:

```
library(arulesViz)
```

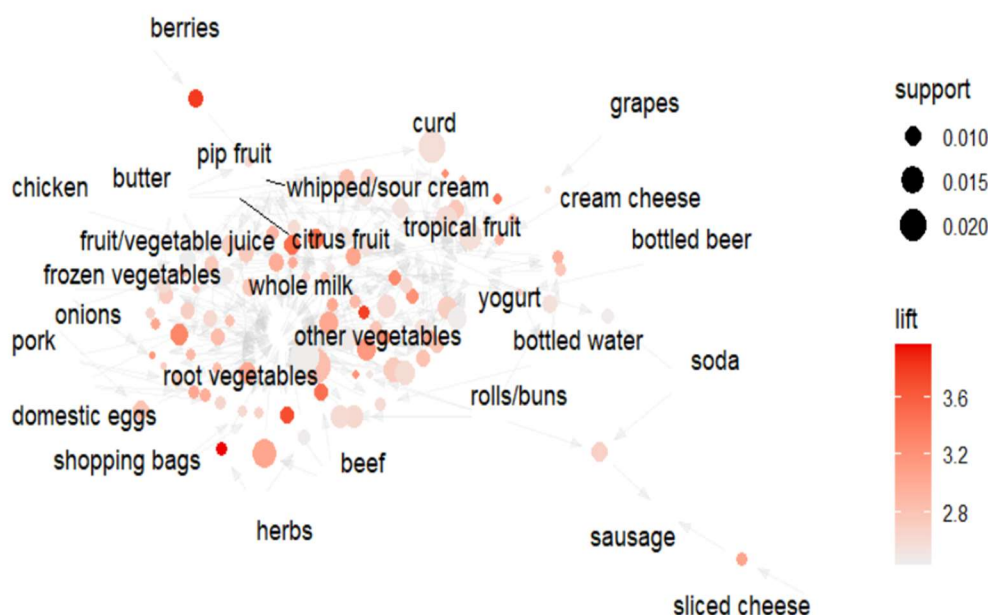
```
plot(top_conf_rules, method = "graph", control = list(type = "items"))
```

```
> library(arulesviz)
> plot(top_conf_rules, method = "graph", control = list(type = "items"))
Warning: Unknown control parameters: type
Available control parameters (with default values):
layout      = stress
circular    = FALSE
ggraphdots  = NULL
edges       = <environment>
nodes       = <environment>
nodetext    = <environment>
colors      = c("#EE0000FF", "#EEEEEEFF")
engine      = ggplot2
max         = 100
verbose     = FALSE
Warning message:
Too many rules supplied. Only plotting the best 100 using 'lift' (change control parameter max if needed).
> |
```

EIKONA 8

Η παρακάτω γραφική παράσταση παρουσιάζει ένα γράφημα, όπου κάθε κόμβος αντιπροσωπεύει ένα προϊόν και κάθε γραμμή μία σχέση συσχέτισης που εξήχθη με τον αλγόριθμο Apriori. Το μέγεθος των κόμβων αντιστοιχεί στη συχνότητα εμφάνισης των προϊόντων (support), ενώ το χρώμα των κόμβων εκφράζει την ένταση της συσχέτισης μέσω του δείκτη lift. Παρατηρούμε ότι προϊόντα όπως το whole milk, τα other vegetables, τα root vegetables και το yogurt βρίσκονται στο επίκεντρο του διαγράμματος, συνδεδεμένα με μεγάλο αριθμό άλλων προϊόντων. Αυτό υποδεικνύει ότι τα προϊόντα αυτά αγοράζονται συχνά και συμμετέχουν σε πολλούς ισχυρούς κανόνες συσχέτισης. Αντίθετα,

προϊόντα όπως τα herbs, berries, domestic eggs και sliced cheese έχουν μικρότερο μέγεθος κόμβου και λιγότερες συνδέσεις, γεγονός που καταδεικνύει ότι αγοράζονται σπανιότερα και έχουν πιο περιορισμένες συσχετίσεις. Συμπερασματικά, το γράφημα αναδεικνύει τη δυναμική σχέση μεταξύ συχνών και σπάνιων προϊόντων, επιβεβαιώνοντας τη σημασία του γάλακτος, των λαχανικών και των γαλακτοκομικών ως κεντρικά σημεία των αγοραστικών μοτίβων.



EIKONA 9

Από την ανάλυση του συνόλου δεδομένων Groceries προκύπτει ένα ξεκάθαρο μοτίβο αγοραστικής συμπεριφοράς των καταναλωτών. Παρατηρείται ότι ο μέσος καταναλωτής αγοράζει κατά κύριο λόγο καθημερινά βασικά είδη, με πιο χαρακτηριστικό το πλήρες γάλα, το οποίο εμφανίζεται στις περισσότερες συναλλαγές και λειτουργεί ως κεντρικός κόμβος στο γράφημα συσχετίσεων. Οι κανόνες συσχέτισης με το υψηλότερο confidence επιβεβαιώνουν ότι προϊόντα όπως το βούτυρο, το γιαούρτι, τα ριζώδη λαχανικά και η κρέμα γάλακτος τείνουν να αγοράζονται μαζί με το γάλα, κάτι που δείχνει έναν σταθερό συνδυασμό ειδών πρωινού και καθημερινής μαγειρικής. Επίσης, παρατηρείται ότι προϊόντα που αγοράζονται σπανιότερα, όπως το curd ή τα tropical fruits, παρότι εμφανίζονται λιγότερο συχνά, σχετίζονται στενά με πιο δημοφιλή προϊόντα, όταν συμπεριλαμβάνονται σε συναλλαγές. Αυτό υποδεικνύει ότι οι αγοραστές ενσωματώνουν ορισμένα σπάνια είδη σε πιο συγκεκριμένες ή προγραμματισμένες αγορές, όπως για γλυκά, συνταγές ή ειδικές περιστάσεις. Συμπερασματικά, το μοτίβο του καταναλωτή χαρακτηρίζεται από συχνή αγορά λίγων βασικών προϊόντων, επαναλαμβανόμενους συνδυασμούς (κυρίως dairy + λαχανικά/φρούτα) και επιλεκτική ενσωμάτωση εξειδικευμένων προϊόντων.

Για την εξαγωγή των συνδυασμών προϊόντων που οδηγούν στην αγορά πλήρους γάλακτος (whole milk), επιλέχθηκαν όλοι οι κανόνες συσχέτισης που έχουν στη δεξιά πλευρά (rhs) το προϊόν "whole milk". Οι κανόνες ταξινομήθηκαν κατά φθίνουσα τιμή εμπιστοσύνης (confidence), με τις εντολές:

```
milk_rules <- subset(rules, rhs %in% "whole milk")
```

```
top_milk_rules <- sort(milk_rules, by = "confidence", decreasing = TRUE)
```

```
inspect(head(top_milk_rules, 10))
```

```
> # Επιλογή των κανόνων που οδηγούν στην αγορά γάλακτος
> milk_rules <- subset(rules, rhs %in% "whole milk")
>
> # Ταξινόμηση με βάση το confidence (πιθανότητα αγοράς γάλακτος)
> top_milk_rules <- sort(milk_rules, by = "confidence", decreasing = TRUE)
>
> # Εμφάνιση των 10 κανόνων με τη μεγαλύτερη εμπιστοσύνη
> inspect(head(top_milk_rules, 10))
```

	lhs	rhs	support	confidence	coverage	lift	count
[1]	{butter, whipped/sour cream}	=> {whole milk}	0.006710727	0.6600000	0.010167768	2.583008	66
[2]	{butter, yogurt}	=> {whole milk}	0.009354347	0.6388889	0.014641586	2.500387	92
[3]	{root vegetables, butter}	=> {whole milk}	0.008235892	0.6377953	0.012913066	2.496107	81
[4]	{tropical fruit, curd}	=> {whole milk}	0.006507372	0.6336634	0.010269446	2.479936	64
[5]	{tropical fruit, butter}	=> {whole milk}	0.006202339	0.6224490	0.009964413	2.436047	61
[6]	{tropical fruit, other vegetables, yogurt}	=> {whole milk}	0.007625826	0.6198347	0.012302999	2.425816	75
[7]	{tropical fruit, domestic eggs}	=> {whole milk}	0.006914082	0.6071429	0.011387900	2.376144	68
[8]	{root vegetables, other vegetables, yogurt}	=> {whole milk}	0.007829181	0.6062992	0.012913066	2.372842	77
[9]	{root vegetables, domestic eggs}	=> {whole milk}	0.008540925	0.5957447	0.014336553	2.331536	84
[10]	{other vegetables, sugar}	=> {whole milk}	0.006304016	0.5849057	0.010777834	2.289115	62

EIKONA 10

Από την ανάλυση των 10 κορυφαίων κανόνων, παρατηρούμε ότι ο πιο ισχυρός κανόνας είναι: $\{butter, whipped/sour cream\} \Rightarrow \{whole milk\}$ με 66% confidence και lift = 2.58.

Ακολουθούν συνδυασμοί όπως:

- ✓ $\{butter, yogurt\} \rightarrow 63.89\%$
- ✓ $\{root\ vegetables, butter\} \rightarrow 63.79\%$
- ✓ $\{tropical\ fruit, curd\} \rightarrow 63.36\%$
- ✓ $\{tropical\ fruit, butter\} \rightarrow 62.24\%$

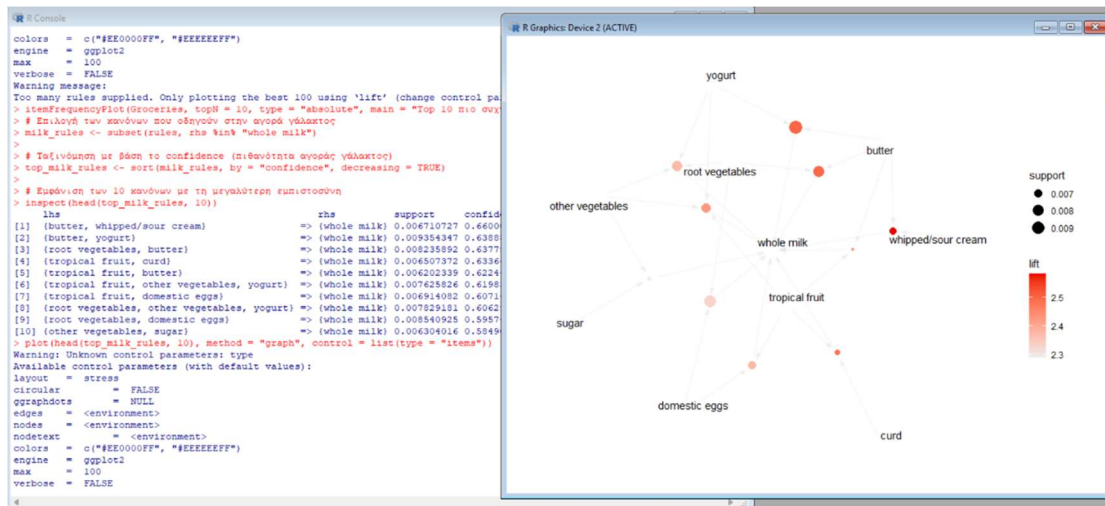
Εμφανίζονται και πιο σύνθετοι συνδυασμοί με 3 προϊόντα, όπως:

- ✓ $\{tropical\ fruit, other\ vegetables, yogurt\} \rightarrow 61.93\%$
- ✓ $\{root\ vegetables, other\ vegetables, yogurt\} \rightarrow 60.62\%$
- ✓ $\{root\ vegetables, domestic\ eggs\} \rightarrow 59.57\%$

Οι παραπάνω συνδυασμοί αναδεικνύουν τη συχνή ταυτόχρονη αγορά γάλακτος μαζί με προϊόντα καθημερινής χρήσης, όπως γαλακτοκομικά, λαχανικά και αυγά. Η υψηλή τιμή του δείκτη lift (> 2.3 σε όλους τους κανόνες) δηλώνει πως οι συσχετίσεις δεν είναι τυχαίες, αλλά αντιπροσωπεύουν σταθερά αγοραστικά μοτίβα. Το πλήρες γάλα αποτελεί ένα βασικό προϊόν που λειτουργεί ως κοινός προορισμός σε πολλές διαφορετικές αγοραστικές διαδρομές.

Η παρακάτω γραφική παράσταση δημιουργήθηκε με τη χρήση της συνάρτησης plot() από το πακέτο arulesViz και απεικονίζει τους 10 κανόνες συσχέτισης με τη μεγαλύτερη εμπιστοσύνη που οδηγούν στην αγορά **πλήρους γάλακτος (whole milk)**. Ο κώδικας που χρησιμοποιήθηκε ήταν:

```
plot(head(top_milk_rules, 10), method = "graph", control = list(type = "items"))
```



ΕΙΚΟΝΑ 11

Στο γράφημα παρατηρείται πως κάθε κόμβος αντιπροσωπεύει ένα προϊόν, ενώ οι γραμμές δείχνουν σχέσεις του τύπου "αν A, τότε B". Το "whole milk" βρίσκεται στο κέντρο του διαγράμματος, καθώς είναι το κοινό προϊόν στην έξοδο (rhs) όλων των κανόνων. Το μέγεθος του κόμβου υποδηλώνει τη συχνότητα εμφάνισης (support) του προϊόντος και η απόχρωση του κόκκινου αντιστοιχεί στη τιμή του lift, δηλαδή στην ένταση της συσχέτισης. Παρατηρούμε ότι προϊόντα όπως το butter, η whipped/sour cream, τα root vegetables, τα tropical fruit, τα domestic eggs και το yogurt βρίσκονται κοντά στο γάλα και συνδέονται με βέλη προς αυτό. Αυτό σημαίνει ότι όταν οι καταναλωτές αγοράζουν αυτούς τους συνδυασμούς, υπάρχει μεγάλη πιθανότητα να αγοράσουν και γάλα. Οι σχέσεις αυτές παρουσιάζουν υψηλές τιμές εμπιστοσύνης (έως και 66%) και lift (άνω του 2.3), αποδεικνύοντας ότι το πλήρες γάλα αποτελεί βασικό προϊόν στα καλάθια των καταναλωτών και σχετίζεται στενά με συγκεκριμένα είδη πρωτεΐνης και μαγειρικής.