Grocery Association Rule Mining

2023-08-10

Let's take a look at the summary of the grocery carts

```
## transactions as itemMatrix in sparse format with
##  15296 rows (elements/itemsets/transactions) and
##  169 columns (items) and a density of 0.01677625
##
## 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
## 3485 2630 2102 7079
##
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   1.000   2.000   3.000   2.835   4.000   4.000
##
## includes extended item information - examples:
##             labels
## 1 abrasive cleaner
## 2 artif. sweetener
## 3   baby cosmetics
##
## includes extended transaction information - examples:
##   transactionID
## 1             1
## 2             2
## 3             3
```

Whole milk and Other Vegetables seems to be the most common items in people's carts.

Let's find some association rules, since the dataset is small, let's be lax with our threshold.Let's take the thresholds at 0.5% support, 10% confidence, and a max length of 4

We found 118 rules, let's take a look at the strongest ones with lift atleast at 3

```
##      lhs                rhs                support     confidence coverage
## [1]  {onions}        => {root vegetables}  0.005295502 0.2655738  0.01993985
## [2]  {onions}        => {other vegetables} 0.007452929 0.3737705  0.01993985
## [3]  {beef}          => {root vegetables}  0.008695084 0.2577519  0.03373431
```

```
## [4]  {root vegetables} => {beef}             0.008695084 0.1240672  0.07008368
## [5]  {pip fruit}       => {citrus fruit}     0.008172071 0.1680108  0.04864017
## [6]  {citrus fruit}    => {pip fruit}        0.008172071 0.1535627  0.05321653
## [7]  {pip fruit}       => {tropical fruit}   0.012683054 0.2607527  0.04864017
## [8]  {tropical fruit}  => {pip fruit}        0.012683054 0.1879845  0.06746862
## [9]  {citrus fruit}    => {tropical fruit}   0.012486925 0.2346437  0.05321653
## [10] {tropical fruit}  => {citrus fruit}     0.012486925 0.1850775  0.06746862
##      lift     count
## [1]  3.789381  81
## [2]  3.004306 114
## [3]  3.677774 133
## [4]  3.677774 133
## [5]  3.157116 125
## [6]  3.157116 125
## [7]  3.864800 194
## [8]  3.864800 194
## [9]  3.477820 191
## [10] 3.477820 191
```

Inference

The first set of people (rule 1-4) seem to be buying Onion, Veggies and Beef for dinner

The next set seems to be buying Fruits in bulk

Interestingly Whole milk does not appear in these strongest of rules.

Let's try to plot our network now

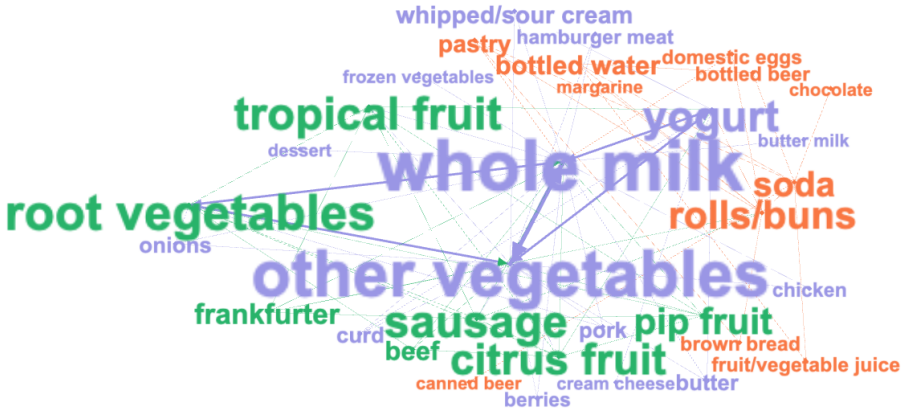

Figure 1: Rules Network

Whole Milk appears in a lot of rules, more than any other item, but the rules do not seem to be that strong.