

# INSTACART PRODUCT NETWORK ANALYSIS

## GROUP MEMBERS:

*Vijetha Shenoy*

*Gulrej Hamirani*

*Guided by Dr. Sridhar Nerur*

## INTRODUCTION

### Objective

Product network analysis on products purchased from Instacart to gain insights and answer questions regarding products most frequently purchased by customers, products that create cross department sales, finding correlation between products, understanding the departments and its products that bring high value to the customers and to Instacart.

### Dataset

Inspired from Kaggle dataset which had about 4 files providing information on orders, products, departments, and aisles. After framing the questions and areas of interest from the files, we prepared and cleaned data through Python and Excel making it ready for further analysis.

**Who benefits?** Beneficial to Instacart, a grocery ordering and delivery app.

### Solution

The insights gained through such network analysis can basically help the retailers improve inventory management and maximize their return on product investments. Correlation between products can help in providing promotional offers to least selling products.

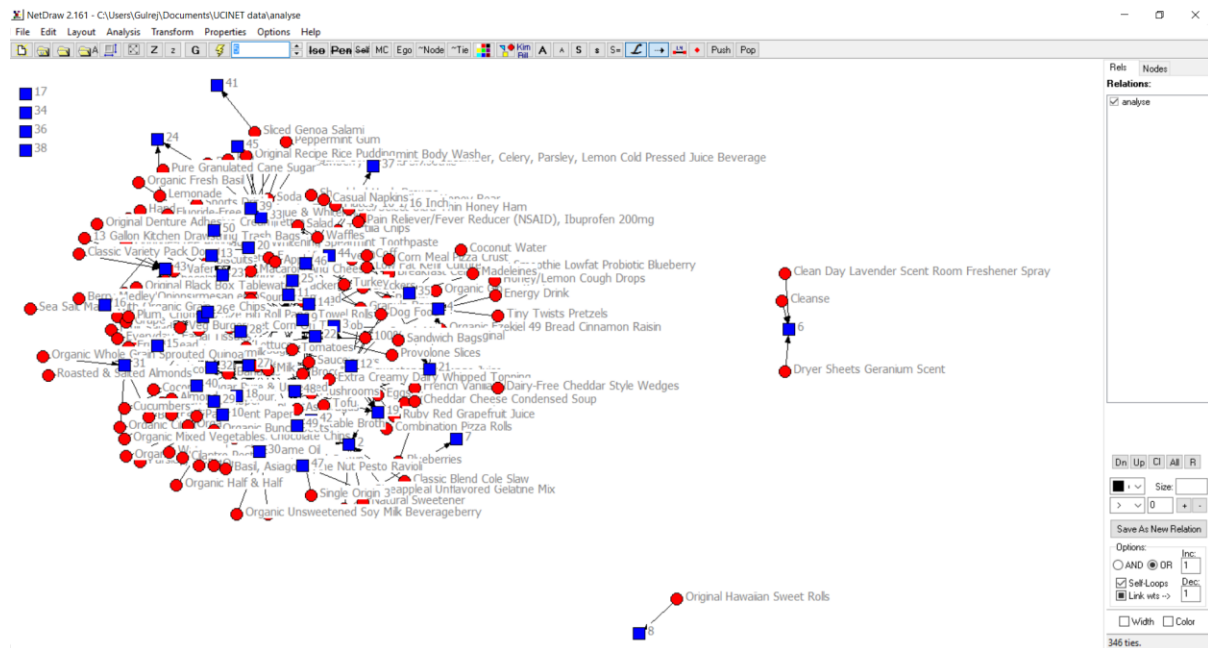
### Approach

We considered 160 products and 50 orders for analysis and created a 2-mode matrix converted it to a bipartite matrix followed by dichotomizing it to 1-mode matrix in UCINET, found multiple measures like Centrality, Betweenness, Beta, Eigenvector and Closeness centrality of the products and visualizing the network in NetDraw.

Additionally, converted the UCINET file to PAJEK file to view it in VOSviewer for network and density visualization. In VOSviewer, clustering and association techniques were applied to analyse the co-purchased products and frequently purchased respectively.

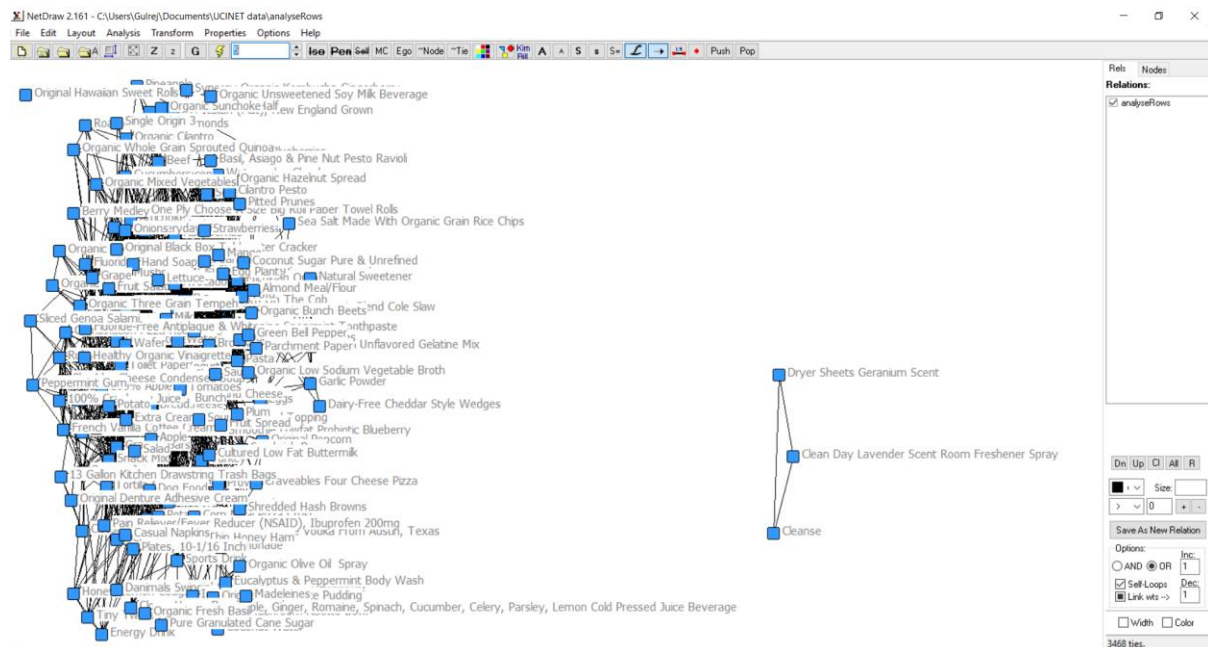
Further analysed on products, department, and aisle level to gain insights.

**Initial Graph of 2-mode network (products, order):**



The initial 2-mode network includes 160 nodes, 346 edges, 4 components and 4 isolates.

**Initial Graph of 1- mode network (products, products):**





## INSY 5380- Social Network Analysis 2017

- ❖ Below screenshot of log obtained using NETWORK-> CENTRALITY AND POWER-> MULTIPLE MEASURES in UCINET displays the centrality measures of total 160 products. Through Excel, we sorted the degree centrality to find the top 10 products with highest degree centrality.

Finalcentrality\_measures1mode.xlsx - Notepad

File Edit Format View Help

		1	2	3	4	5
		Degree	BetaCent	Closeness	Eigenvect	Between
1	Eggs	18.000	3127.405	323.000	0.045	84.650
2	Michigan Organic Kale	46.000	9855.753	290.000	0.143	218.430
3	Garlic Powder	7.000	897.919	408.000	0.013	0.000
4	Butter	38.000	7725.687	296.000	0.112	218.165
5	Natural Sweetener	7.000	897.919	408.000	0.013	0.000
6	Carrots	43.000	7930.995	291.000	0.115	390.547
7	Original Unflavored Gelatine Mix	7.000	897.919	408.000	0.013	0.000
8	Classic Blend Cole Slaw	7.000	897.919	408.000	0.013	0.000
9	Yogurt	58.000	10879.273	273.000	0.158	338.767
10	Unsweetened Almondmilk	24.000	4797.826	310.000	0.070	66.570
11	Lemons	45.000	8884.456	289.000	0.129	147.484
12	Baby Spinach	31.000	4897.694	308.000	0.071	212.295
13	Organic Ginger Root	6.000	1219.348	371.000	0.018	0.000
14	Frozen Chicken	73.000	11994.814	257.000	0.174	691.872
15	Organic Ezekiel 49 Bread Cinnamon Raisin	6.000	1219.348	371.000	0.018	0.000
16	Bagels	25.000	3782.322	312.000	0.055	109.688
17	Honey/Lemon Cough Drops	8.000	767.181	416.000	0.011	0.000
18	Granula Bars	40.000	5996.688	296.000	0.087	203.194
19	Breakfast Cereal	23.000	3539.374	318.000	0.051	99.250
20	Tiny Twists Pretzels	8.000	767.181	416.000	0.011	0.000
21	Snack Mix	29.000	4276.855	307.000	0.062	111.370
22	Crackers	35.000	5206.434	299.000	0.075	261.829
23	Orange Juice	20.000	2364.327	343.000	0.034	59.136
24	Energy Drink	8.000	767.181	416.000	0.011	0.000
25	Bananas	51.000	11438.753	284.000	0.166	154.241
26	Just Crisp, Parmesan	23.000	6009.768	319.000	0.087	0.000
27	Fruit Salad	23.000	6009.768	319.000	0.087	0.000
28	Raspberries	46.000	10091.528	291.000	0.146	191.271
29	Toilet Paper	35.000	7782.666	300.000	0.113	46.710
30	Water	63.000	11875.432	268.000	0.172	521.832
31	Tea	30.000	6662.873	310.000	0.097	84.775
32	Cheese	109.000	16411.012	221.000	0.238	2715.047
33	Macaroni And Cheese	42.000	8572.042	289.000	0.124	119.517

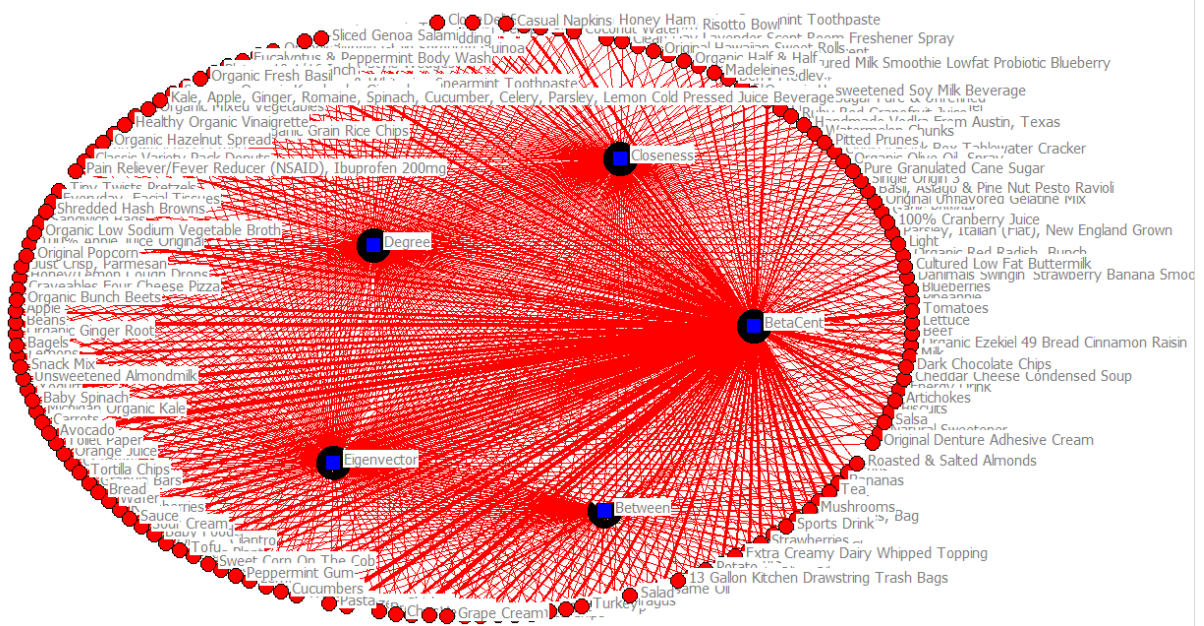
File Home Insert Draw Page Layout

R1C1 X ✓ fx Product nam

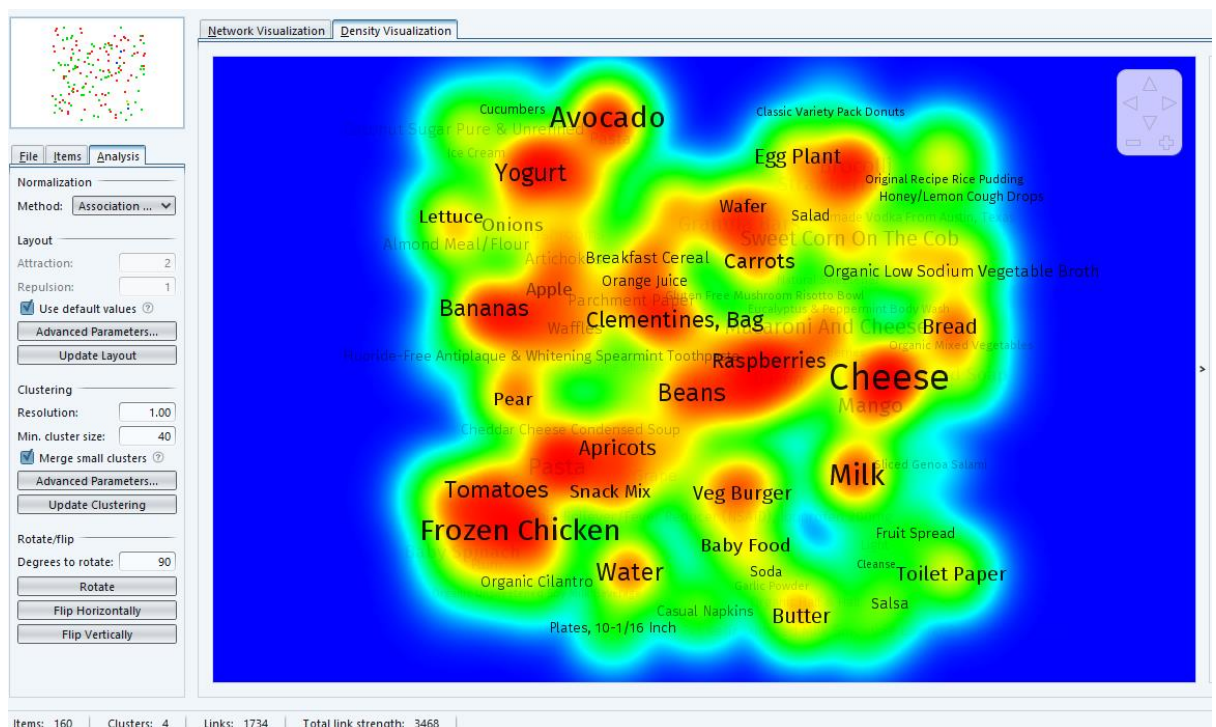
	1	2	3
1	Product names	Degree	
2	Cheese	109	
3	Avocado	78	
4	Milk	77	
5	Frozen Chicken	73	
6	Water	63	
7	Beans	63	
8	Yogurt	58	
9	Tomatoes	55	
10	Pasta	52	
11	Bananas	51	
12	Clementines, Bag	47	
13	Michigan Organic Kale	46	
14	Raspberries	46	
15	Lemons	45	
16	Sour Cream	44	
17	Carrots	43	
18	Mango	43	
19	Macaroni And Cheese	42	
20	Egg Plant	41	
21	Sweet Corn On The Cob	41	
22	Granula Bars	40	
23	Apricots	40	
24	Bread	40	
25	Broccoli	40	
26	Butter	38	

Finalcentrality\_measures1mode\_a

Below is the graphical representation in UCINET of the above centrality measures calculated.

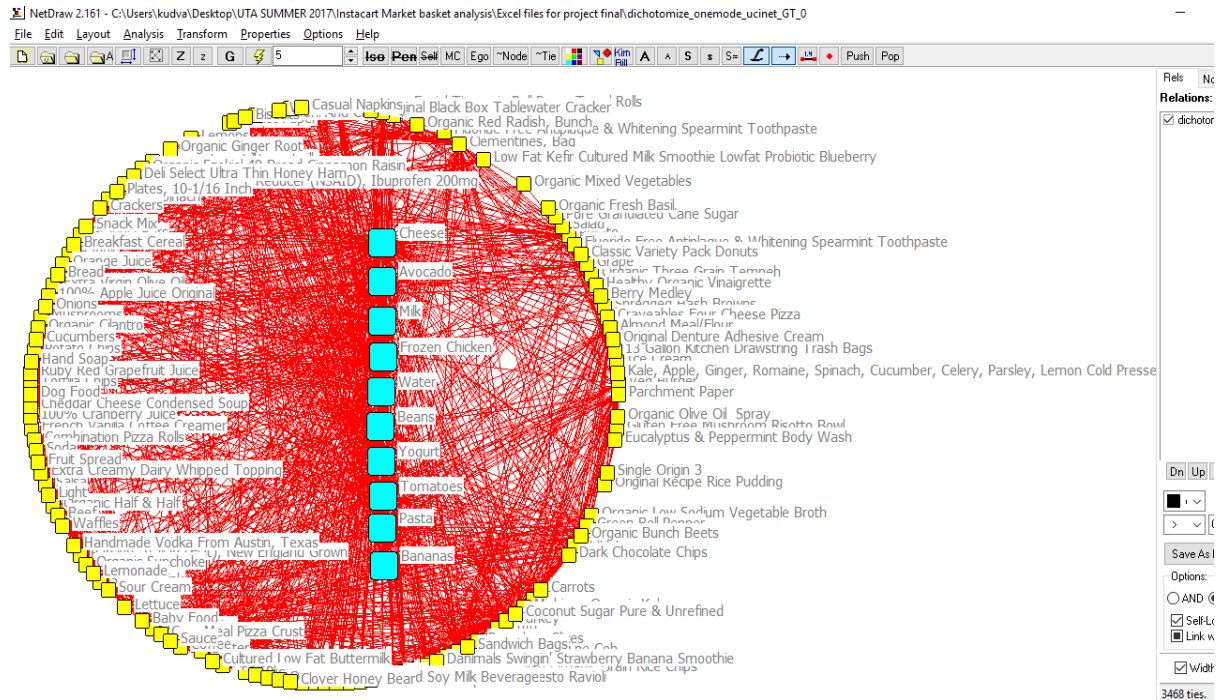


- ❖ Below is density visualization of products in VOSviewer.
- ❖ As observed, the darker red areas are highly central products which return maximum return on product investments.



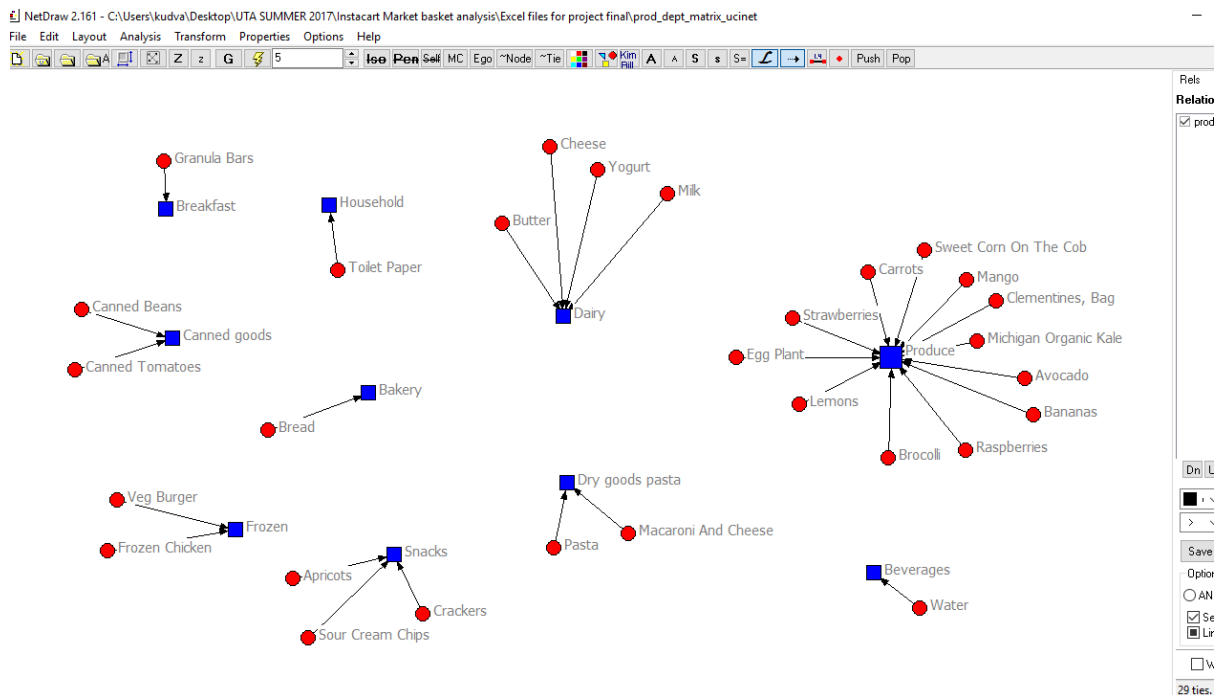


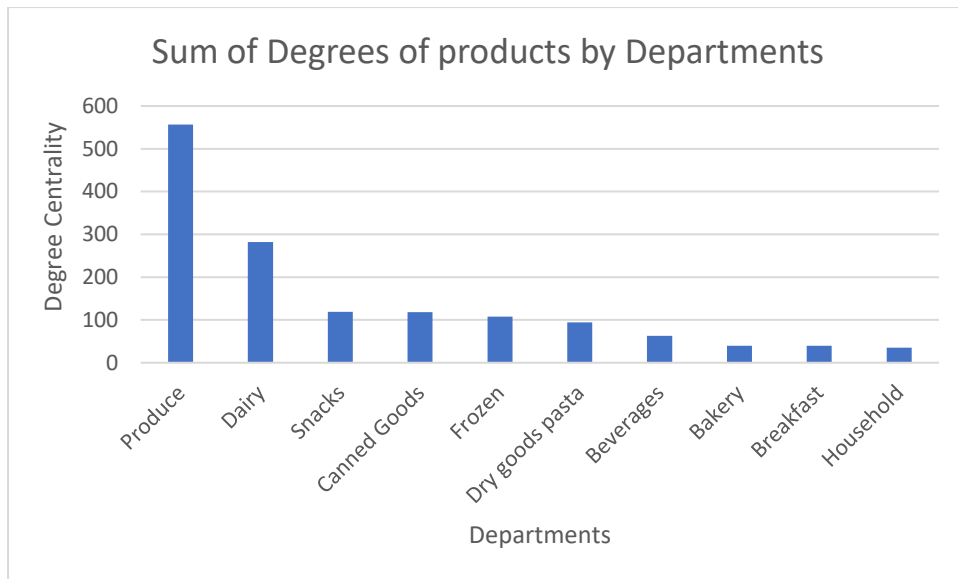
In UCINET, as per the centrality measures obtained, below is an **ego -alter network of only the top 10 products frequently purchased**. The blue colour indicates ego and yellow colour indicates it's respective alters. Since there are about 3468 ties, we have chosen to display the network using Circle for better visualization.



### Insights from products and departments data

As observed in the below network, the departments like Produce, Dairy, Household, Breakfast, Canned Goods, Bakery, Frozen, Snacks, Dry goods pasta and Beverages hold top 30 highly central products making these departments most profitable.

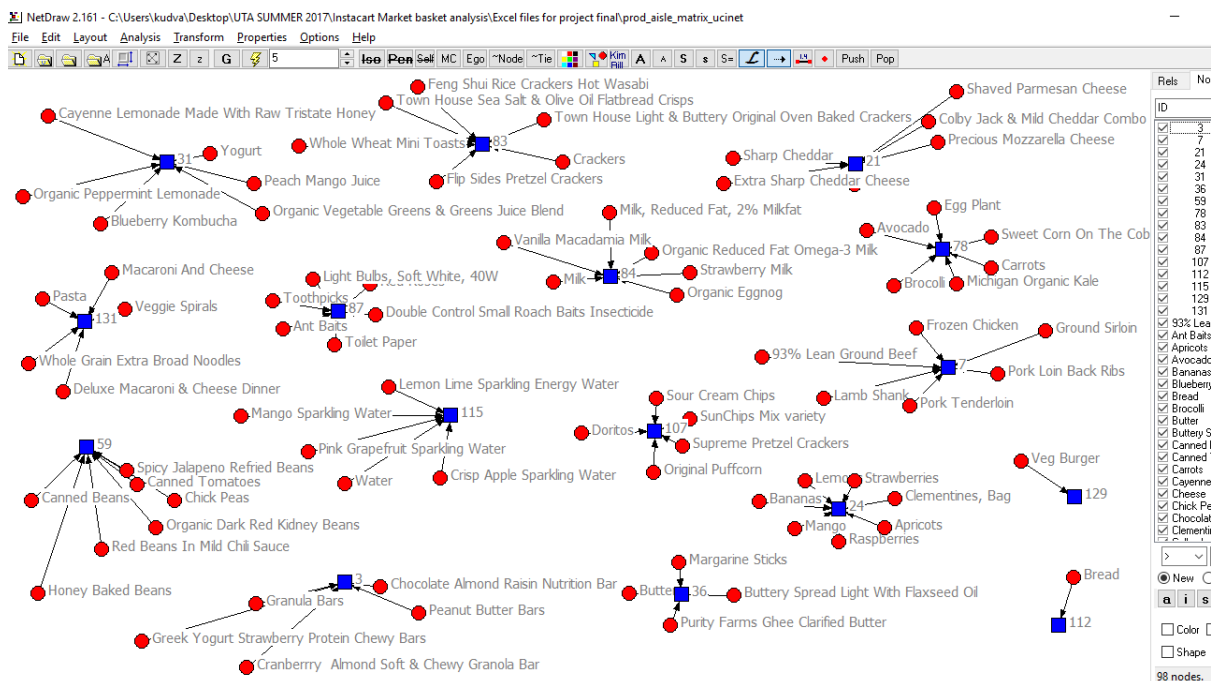




Above distribution chart displays the departments with highest centrality products which can help to identify the top selling categories. As per the chart, produce i.e. vegetables and fruits section contains frequently purchased products. Based on this chart, promotions and sales in these departments can be offered to promote and sell the least selling products with high selling ones.

### Insights from products and aisle data

Considering 50 products and their Aisle ID, analysis of top 5 products with highest degree in respective aisles can help to place those products in the aisle and shelves accordingly for better profits and apply marketing strategy using this insight.





### **Conclusion and further research scope:**

As we understand, product purchase network analysis opens a new range of insights which can maximize return on category investments and improve inventory management. This analysis can further help retain the loyal customers through loyalty programs.

If we have more information on customer's purchases history over the year and personal information like age, location, and gender, we can predict products likely to be purchased by the customer and even prepare a recommendation system. If any isolates among products or upon setting a threshold to see which products generate very low sales, they could be discontinued.

In conclusion, we analysed most frequently purchased products, association of co-purchased products, department level profits, aisle level marketing strategy for promotions and triggering sales within and among other departments.

### **References:**

<https://www.kaggle.com/c/instacart-market-basket-analysis>

<https://forteconsultancy.wordpress.com/2013/02/19/product-network-analysis-the-next-big-thing-in-retail-data-mining/>

<http://www.sciencedirect.com/science/article/pii/S0957417412000796#b0060>

[https://www.egsh.eur.nl/fileadmin/egs3h\\_content/Documents/Doctoral\\_education/PhD\\_information\\_skills/VOSviewer\\_-\\_create\\_and\\_visualize\\_bibliometric\\_networks.pdf](https://www.egsh.eur.nl/fileadmin/egs3h_content/Documents/Doctoral_education/PhD_information_skills/VOSviewer_-_create_and_visualize_bibliometric_networks.pdf)

<http://www.vosviewer.com>