



# Food For Thought:

*A Network Analysis of Supermarket Ingredients,  
a visualization of our food system, and  
a guide to understanding what we are consuming*

Presented by Samuel Oliveira



# Today's Agenda

- Introduction
- Historical Context
- Methodology
- Construction & Statistics
- Network Measures
- Conclusion

Please feel free to ask any questions during the presentation



# Introduction & Motivation

**“In the age of industrial food, eating is an act of trust.”**

*~ Eric Schlosser*

**“Tell me what you eat, and I will tell you what you are.”**

*~ Jean Anthelme Brillat-Savarin*

# The History of Processed Foods

## Rise of Processed Foods

The development of food processing machinery marks the beginning of a new dietary era

## Convenience is King

Social proclivities select foods for their convenience rather than their health and fresh sourcing

## Diet Obfuscation

We are left with a poor understanding of how our foods are made and what they contain



## Risk of food-borne pathogen infection and malnutrition

### Risk of micronutrient deficiency

### Risk of overeating and non-communicable diseases

### RELATED HEALTH

#### agriculture & live stock farming

hunters & gatherers

crop selection  
animal domestication

#### industrialisation

2,000,000 BP  
fire use  
(grilling, cooking,  
roasting, charring)

250,000 BP  
salting

drying

pounding or  
grinding

drying

fermentation  
(alcohol, cheese, bread)

oven use (baking,  
cooking)

1802 first sugar  
beet factory

1812 first canning  
factory

1867 invention of  
pasteurisation

1870 market launch  
of margarine

1874 first refrigerator

food shortage  
(world  
wars, depression)

canning, chilling, pasteurisation  
1876 debut of  
root beer\* &  
1886 Coca-Cola

hydrogenation of oils

spray drying, freeze drying  
mass production

1908 invention of  
stock cubes (Maggi)

1912 first breakfast  
cereals (Kellogg's)\*\*

trans-fatty acids

increasing industrial use of salt

economic growth  
automatisation, liberalisation & globalisation

obesity pandemic takes its course

improved product packaging & shelf life  
fast food restaurants

ready-to-eat meals

sweeteners, fat substitutes,  
calorie-reduced products

functional food, organic food,  
whole food, super food

1988 market launch of  
„Lunchables“ (Kraft)\*\*\*

1994 first GMO food  
„Flavr Savr“ tomato

SOCIOECONOMIC  
TRANSITION

EVOLUTION OF FOOD  
PROCESSING

12,000-4,000 BP  
1800  
1945  
1950  
2020

2,000,000 BP

250,000 BP

fire use  
(grilling, cooking,  
roasting, charring)

salting

drying

fermentation  
(alcohol, cheese, bread)

oven use (baking,  
cooking)

1802 first sugar  
beet factory

1812 first canning  
factory

1867 invention of  
pasteurisation

1870 market launch  
of margarine

1874 first refrigerator

12,000-4,000 BP

1800

1945

1950

2020





## Expectations and Potential Insights

- 1) Understanding of the most common active ingredients in an American Diet**
  - 2) Visualization of the 'hidden' ingredients in our dietary network**
  - 3) Ability to trace back our foods and diets to their source**
  - 4) Information on what are the major corporations that provide us with the food we eat**
-



# Data Collection & Methodology

## Data Collection

To collect the data for ingredients in foods and their connections, I used a database from USDA Food-Data Central. Data collection is changing with the emergence of artificial intelligence. With these new tools at our disposal, I wanted to experiment with the power of a Claude-Developed web-scraper. So that I could collect enough data to have some definitive results.

## Methodology

The **nodes** in this graph are **ingredients** in foods.

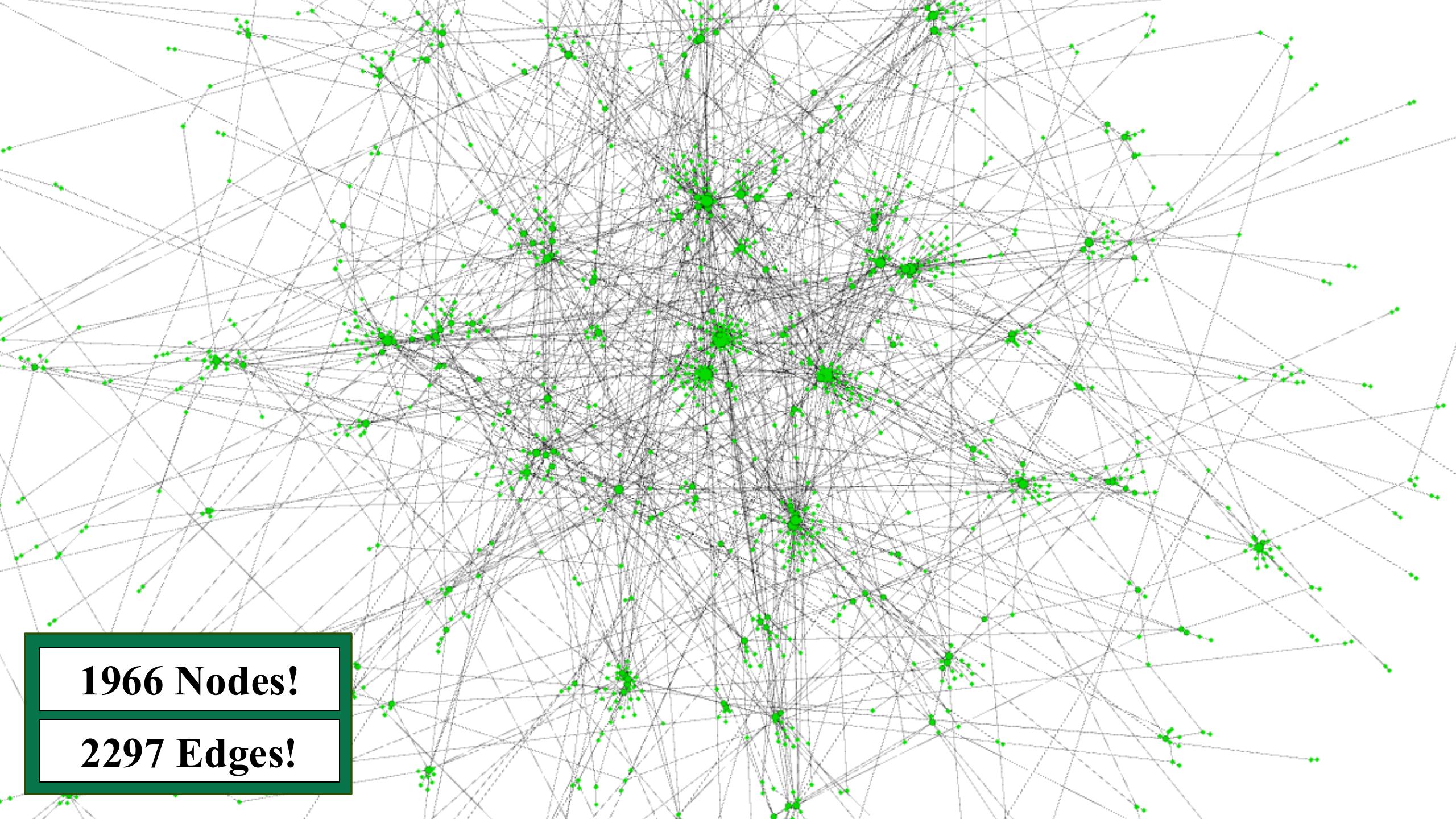
The **edges** between any two ingredients mean that they show up in a food together. The edges are weighted (the weight will be how many times that pair shows up in the set of foods)

The graph and its measures were built and calculated by Gephi. Excel was used for formatting the data.



*Let's analyze*

# The Graph

A complex network graph visualization showing numerous small green dots representing nodes and a dense web of gray lines representing edges connecting them. The graph is highly interconnected, with many clusters of nodes and some prominent hubs. The overall structure is organic and sprawling.

**1966 Nodes!**

**2297 Edges!**

**1966 Nodes!**

**2297 Edges!**

**Modified  
Food Starch**

**Vegetable Oil**

**Soybean Oil**

**Milk**

**Citric Acid**

**Colorants**

**Reduced Iron &  
Mono-diglycerides**

**Folic Acid**

**Water**

**Spice**

**Guar-Gum**

**Salt**

**Sugar**

**Corn  
Starch**

**Niacin**

**Artificial  
Flavor**

**Caramel Color**

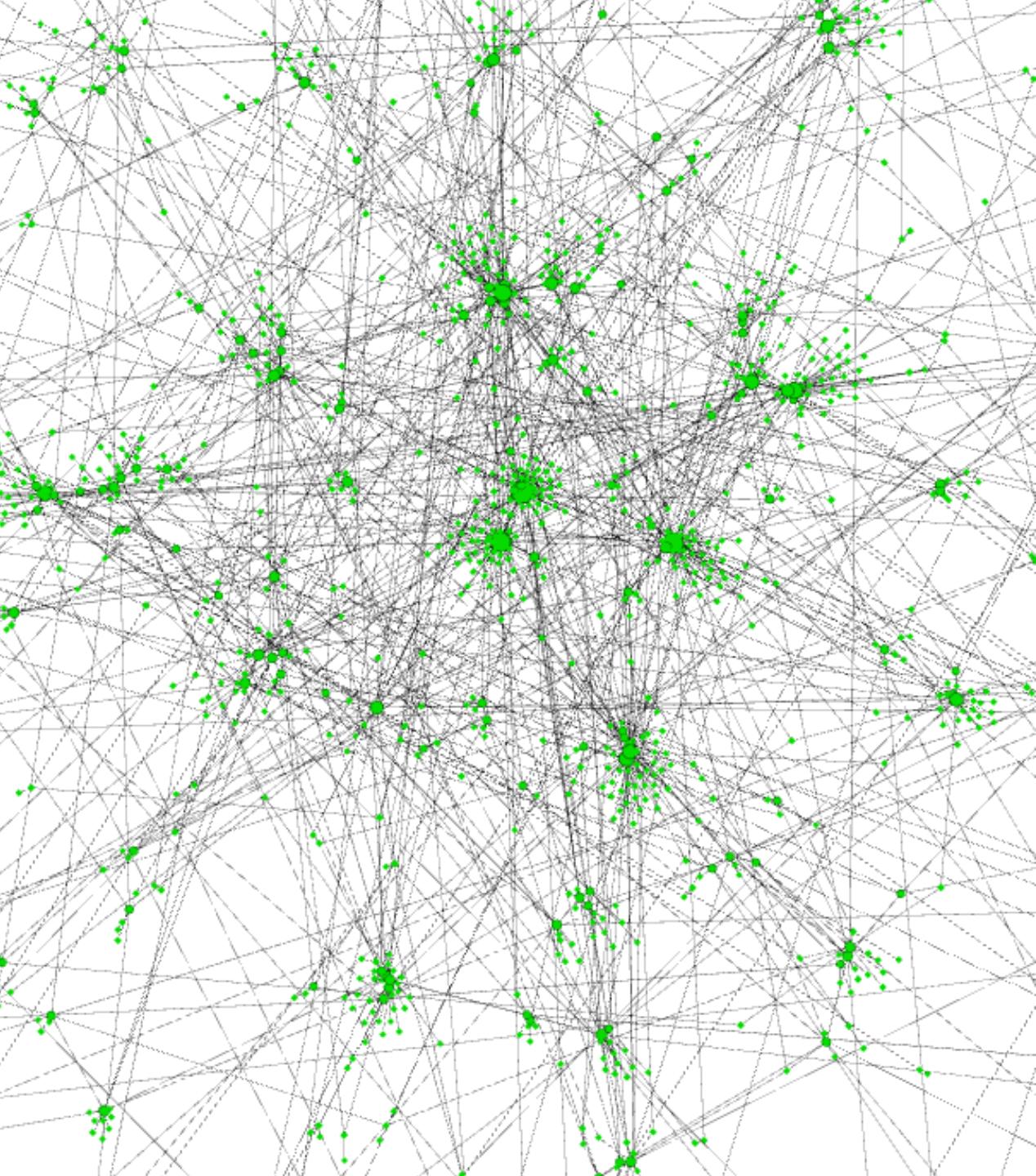
**Dextrose**

**Yeast**

**Wheat Flour**

# Global Graph Measures:

- **Number of Nodes:** 1966
- **Number of Edges:** 2297
- **Average Degree:** 2.337 (*small world*)
- **Average Weighted Degree:** 7.046
- **Network Diameter:** 16
- **Graph Density:** 0.001
- **Connected Components / Cliques:** 136
- **Modularity:** 0.842
- **Average Path Length:** 5.881





# **Node Specific Measures**

# General Overview

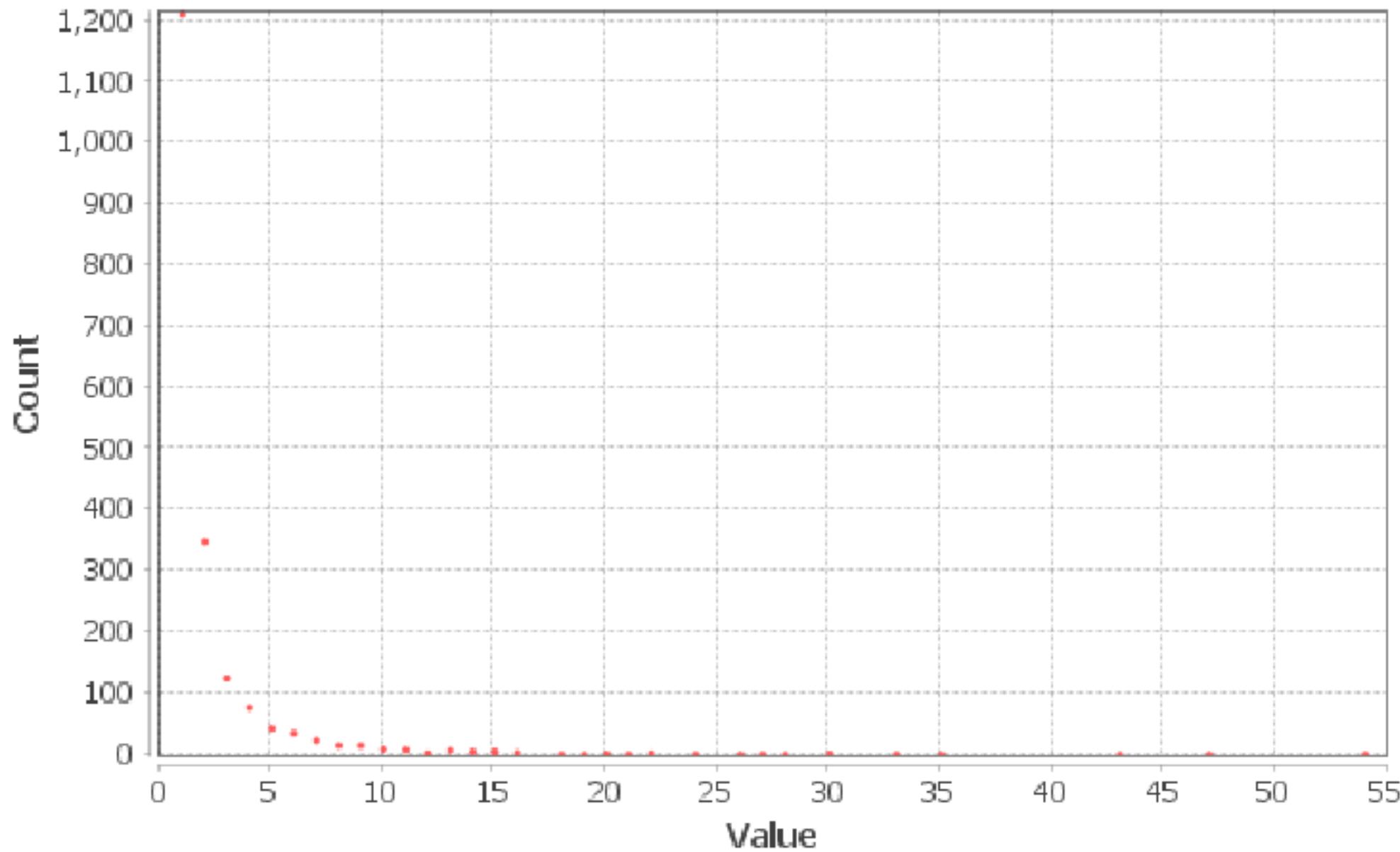
Id	Degree ▾	Weighted Degree	Eigenvector Centrality	Betweenness Centrality	Closeness Centrality	Hub	PageRank	Clustering Coefficient
salt	54	614.0	1.0	158889.82773	0.249661	0.621468	0.009481	0.000699
sugar	47	467.0	0.647249	100088.24948	0.227716	0.108811	0.008545	0.0
water	43	292.0	0.576341	90247.602916	0.228155	0.146529	0.007949	0.0
folic-acid	35	329.0	0.587741	144339.277841	0.247944	0.095337	0.005902	0.001681
wheat-flour	33	346.0	0.348914	67129.909793	0.22336	0.025551	0.006293	0.001894
spices	30	163.0	0.378189	77160.251316	0.225425	0.074216	0.005234	0.0
soybean-oil	30	56.0	0.336541	75604.28147	0.221836	0.033323	0.005112	0.002299
citric-acid	28	153.0	0.320378	84209.52022	0.226782	0.02584	0.004904	0.002646
enzymes	27	117.0	0.31864	76911.582349	0.228406	0.03883	0.004562	0.002849
riboflavin	26	256.0	0.282107	54299.855383	0.215633	0.036332	0.00471	0.0
niacin	24	169.0	0.396855	65591.76702	0.238733	0.158345	0.004298	0.0
flavor	22	88.0	0.277078	47306.231381	0.229449	0.052484	0.004176	0.0
dextrose	22	85.0	0.234622	60680.55497	0.219312	0.037316	0.00381	0.0
vitamin-b	22	32.0	0.393787	114915.93516	0.24318	0.064409	0.00346	0.0
soy-lecithin	21	312.0	0.35501	112095.021981	0.241023	0.050502	0.003485	0.004762
garlic	20	117.0	0.223684	87366.61275	0.229037	0.028422	0.003414	0.0
caramel-color	20	63.0	0.126676	39075.658064	0.191322	0.004915	0.00388	0.0
black-pepper	19	39.0	0.20795	69856.970018	0.221037	0.015333	0.003249	0.005848
milk	18	98.0	0.17372	42832.989299	0.213991	0.016245	0.00292	0.0
yellow	16	138.0	0.285828	46088.449998	0.221155	0.070663	0.002655	0.008333
sea-salt	16	43.0	0.153017	25567.27355	0.211399	0.020232	0.002807	0.0
red	16	44.0	0.307369	82605.216147	0.244723	0.0493	0.002633	0.008333
flavors	16	40.0	0.111374	29063.260682	0.196422	0.010424	0.003056	0.0
sodium-bicarbonate	15	123.0	0.194991	36566.316633	0.224297	0.036828	0.002684	0.0
ascorbic-acid	15	38.0	0.081705	28382.550765	0.184735	0.001698	0.002915	0.0
lactic-acid	15	51.0	0.225659	44697.182923	0.227123	0.040515	0.002521	0.0
modified-food-starch	15	54.0	0.132991	35143.006422	0.211372	0.009304	0.002535	0.0
corn-starch	15	32.0	0.099074	27872.924454	0.195565	0.006145	0.002724	0.0
vinegar	15	34.0	0.14983	38163.50263	0.218849	0.011512	0.002511	0.0
vegetable-oil	14	59.0	0.168206	61958.782247	0.22327	0.01248	0.002359	0.010989
thiamine-mononitrate	14	28.0	0.162033	38378.836563	0.222401	0.017546	0.002542	0.010989
flour-wheat-flour	14	30.0	0.309427	52812.940277	0.233949	0.123954	0.002447	0.0
color	14	35.0	0.180833	43160.558758	0.224631	0.022109	0.002398	0.0
garlic-powder	14	26.0	0.121712	24480.019093	0.201287	0.033888	0.002476	0.0
thiamin-mononitrate	14	38.0	0.1748	49450.747916	0.224448	0.031122	0.002295	0.0
corn-syrup	13	147.0	0.164194	29341.883129	0.220479	0.029896	0.002167	0.0
reduced-iron	13	153.0	0.098373	22164.582682	0.192746	0.005713	0.00252	0.0
potassium-sorbate	13	63.0	0.200424	35882.325642	0.227154	0.037258	0.002202	0.0
yeast-extract	13	79.0	0.306341	56581.966172	0.239112	0.053181	0.002086	0.0
vitamin-palmitate	13	61.0	0.113121	26838.909019	0.201507	0.005836	0.002182	0.0
iron	13	35.0	0.123067	22481.938097	0.210326	0.023689	0.002161	0.0

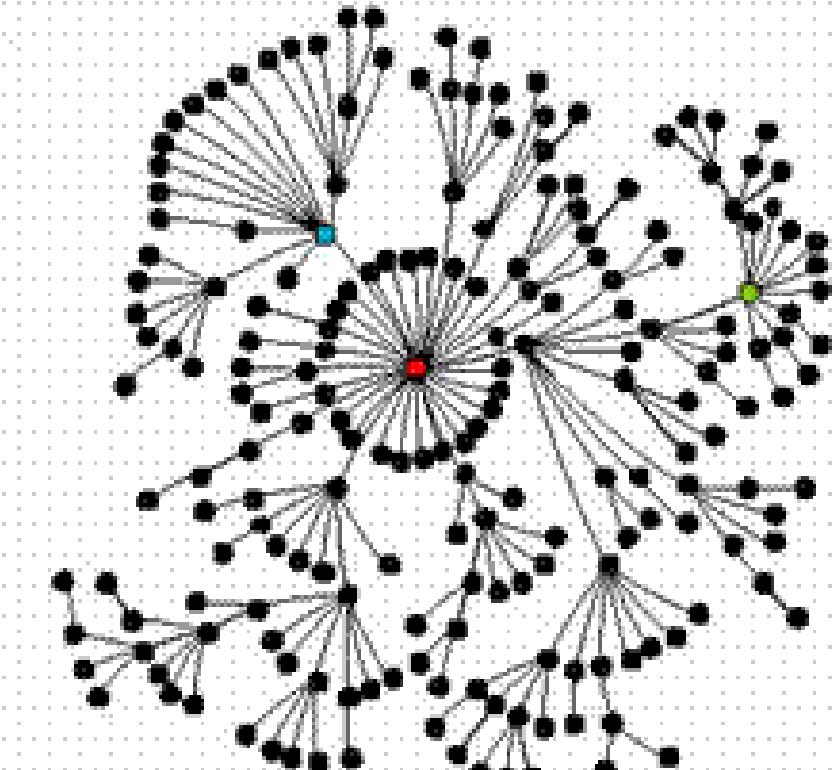
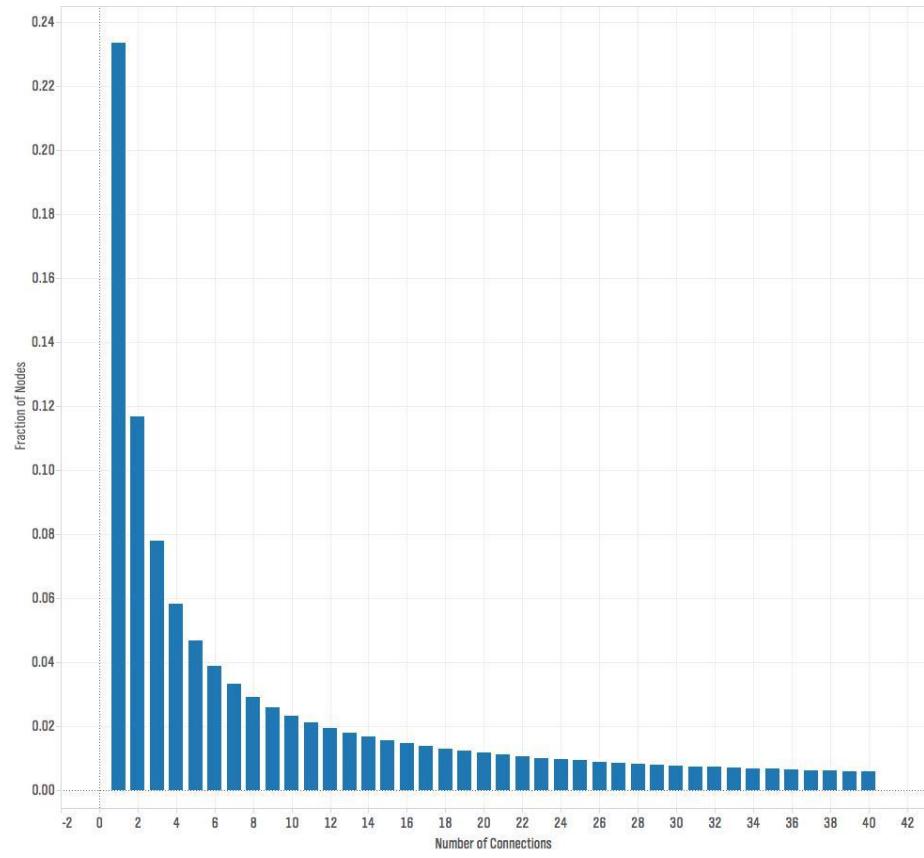
# Degree Centrality

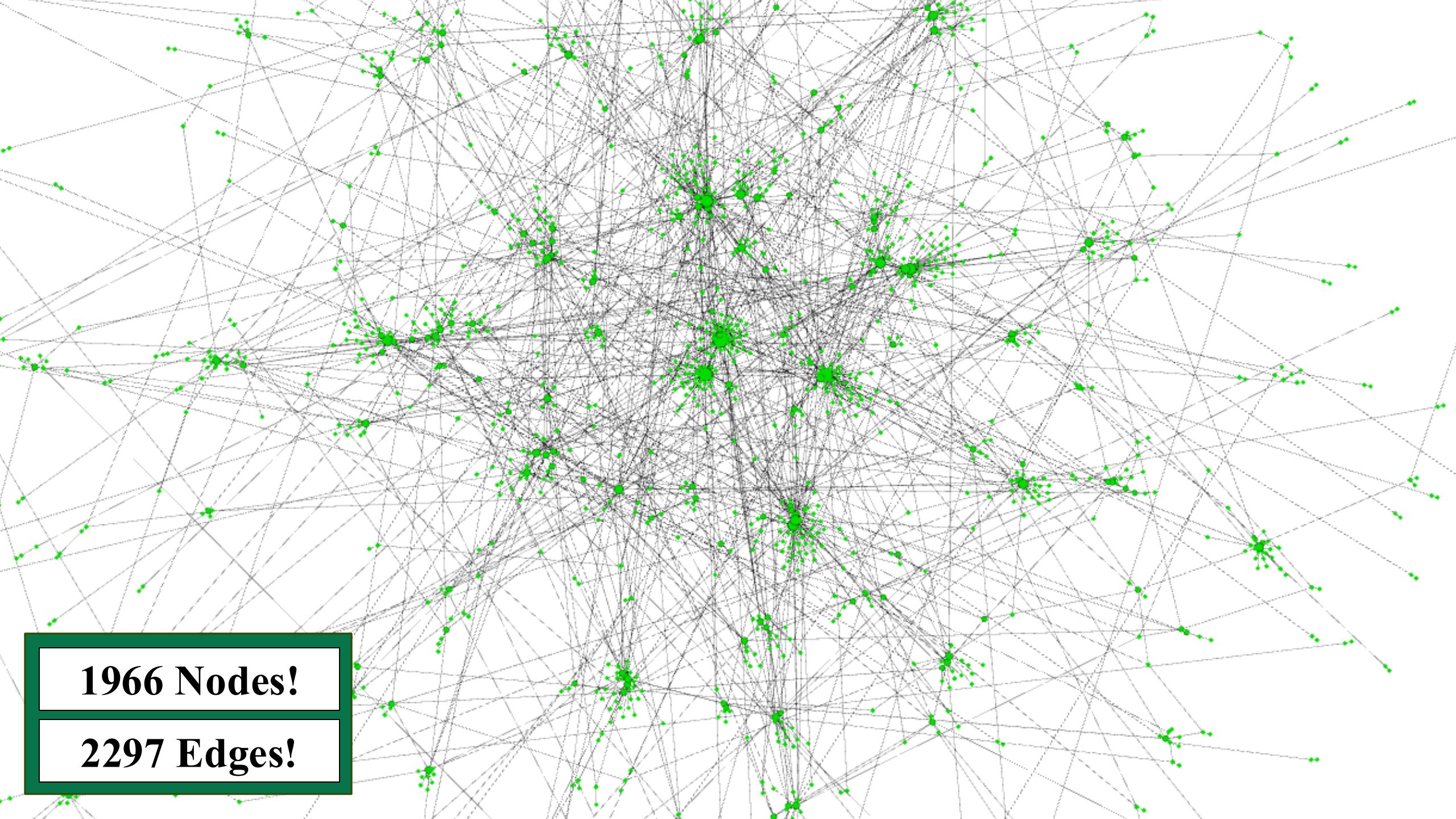
- These are the top 15 nodes in terms of degree centrality and weighted degree centrality
- Note the difference between Degree centrality and Weighted Degree centrality
  - Degree centrality shows the size of the range of ingredients one ingredient can match with
  - Weighted Degree centrality shows how many different foods that ingredient is a part of
- Swaps in the lists demonstrate how some ingredients are used in less contexts, but more frequently and vice versa

Id	Degree	Id	Weighted ...
salt	54	salt	614.0
sugar	47	sugar	467.0
water	43	wheat-flour	346.0
folic-acid	35	folic-acid	329.0
wheat-flour	33	soy-lecithin	312.0
spices	30	water	292.0
soybean-oil	30	riboflavin	256.0
citric-acid	28	malted-barley-flour	223.0
enzymes	27	nonfat-dry-milk	188.0
riboflavin	26	niacin	169.0
niacin	24	palm-oil	168.0
flavor	22	sodium-benzoate	164.0
dextrose	22	spices	163.0
vitamin-b	22	oregano	158.0
soy-lecithin	21	titric-acid	153.0

# Degree Distribution





A complex network graph visualization showing numerous small green dots representing nodes and a dense web of gray lines representing edges connecting them. The graph is highly interconnected, with many clusters of nodes and several prominent hubs. The overall structure is organic and sprawling.

**1966 Nodes!**

**2297 Edges!**

# Eigenvalue Centrality

- **Note:** Salt has an eigenvalue centrality of 1!!
- With eigenvalue centrality, we can truly see which nodes are of maximum importance since the ranking is based on connection importance
- This means that the top ingredients in this list are both used frequently, and in the most foods as well. The ranking is a middle ground for what we say in the degree centrality rankings
- The top four nodes, (*salt, sugar, folic acid, and water*), co-exist in the majority of foods

Id	Eigenvector Centrality
salt	1.0
sugar	0.647249
folic-acid	0.587741
water	0.576341
niacin	0.396855
vitamin-b	0.393787
spices	0.378189
soy-lecithin	0.35501
wheat-flour	0.348914
soybean-oil	0.336541
citric-acid	0.320378
enzymes	0.31864
flour-wheat-flour	0.309427
red	0.307369
yeast-extract	0.306341
yellow	0.285828
riboflavin	0.282107
flavor	0.277078
nonfat-milk	0.276591
rice	0.274137
blue	0.238151
dextrose	0.234622
lactic-acid	0.225659
garlic	0.223684
palm-oil	0.21502
dextrin	0.211543
sodium-benzoate	0.20825

# Closeness Centrality

- These ingredients are the closest on average to the other nodes around them. (*the values of closeness centrality have been normalized*)
- Even though these ingredients can reach other ingredients around them in a few steps, this does not really imply anything too important about our food structure
- The nodes listed are universal in their use cases. Even though they may not be common, they are able to connect the food world together
  - They may represent the links between different types of foods as they can be universally used.
  - **Note:** Commonly used ingredients appear in hubs as many recipes use very similar ingredients

Id	Closeness Centrality
brominated-vegetable-oil	1.0
calcium-disodium-edta	1.0
artificial-flavor-sausage-pork	1.0
whey-solids	1.0
bay-leaf	1.0
sulfiting-agents-contains-sulfites	1.0
calcium-phosphate	1.0
cultured-grade-skim-milk	1.0
candy-bits	1.0
corn-cereal	1.0
sodium-alginate	1.0
cottonseed-oil	1.0
less-than-tricalcium-phosphate-to-prevent-caking	1.0
nonfat-milk	1.0
malted-barley	1.0
fish-protein	1.0
tetrasodium-pyrophosphate	1.0
rye	1.0
lemon-pepper	1.0
lymph-nodes-fat	1.0
pork-salivary-glands	1.0
mandarin-oranges	1.0
mizuna	1.0
modified-acacia-gum	1.0
salt-less-than-rosemary-extract	1.0
yeast-nutrient	1.0
tocopherols-ascorbic-acid	1.0
acidity-regulator	1.0
chilli-pepper	1.0

# Betweenness Centrality

- Betweenness centrality demonstrates a more intuitive way of measuring closeness
  - Instead of seeing how close an ingredient is to others, we see how useful an ingredient is in creating links with other sections of the graph.
  - These ingredients will represent the ingredients that most commonly match two or more different types of foods
- The presence of processed ingredients is higher in this list, and the rankings seem to prefer processed foods as well.
- Processed foods provide a culinary bridge

Id	Betweenness Centrality
salt	158889.82773
folic-acid	144339.277841
vitamin-b	114915.93516
soy-lecithin	112095.021981
sugar	100088.24948
water	90247.602916
garlic	87366.61275
citric-acid	84209.52022
red	82605.216147
spices	77160.251316
enzymes	76911.582349
soybean-oil	75604.28147
black-pepper	69856.970018
wheat-flour	67129.909793
niacin	65591.76702
vegetable-oil	61958.782247
dextrose	60680.55497
blue	56755.75928
yeast-extract	56581.966172
riboflavin	54299.855383
flour-wheat-flour	52812.940277
thiamin-mononitrate	49450.747916
flavor	47306.231381
yellow	46088.449998
lactic-acid	44697.182923

# Page Rankings

- Page rank will measure the weighted probability of an ingredient being used in any given food
- This list is a guide to the most "generally used" foods in an average supermarket.
  - If you were to walk through a supermarket and look through the ingredients, there is a high probability that a few of these ingredients will be listed.
- Processed ingredients seem to be a common theme in page rank
  - Note that many of these ingredients show up in out degree centrality and betweenness centrality rankings as well

Id	PageRank ▾
salt	0.009481
sugar	0.008545
water	0.007949
wheat-flour	0.006293
folic-acid	0.005902
spices	0.005234
soybean-oil	0.005112
citric-acid	0.004904
riboflavin	0.00471
enzymes	0.004562
niacin	0.004298
flavor	0.004176
caramel-color	0.00388
dextrose	0.00381
soy-lecithin	0.003485
vitamin-b	0.00346
garlic	0.003414
black-pepper	0.003249
flavors	0.003056
milk	0.00292
ascorbic-acid	0.002915
sea-salt	0.002807
corn-starch	0.002724
sodium-bicarbonate	0.002684
yellow	0.002655
red	0.002633

# Comparisons Between Measures

Degree	Weighted	Eigenvector	Betweenness	Page Rank
salt	salt	salt	salt	salt
sugar	sugar	sugar	folic-acid	sugar
water	wheat-flour	folic-acid	vitamin-b	water
folic-acid	folic-acid	water	soy-lecithin	wheat-flour
wheat-flour	soy-lecithin	niacin	sugar	folic-acid
spices	water	vitamin-b	water	spices
soybean-oil	riboflavin	spices	garlic	soybean-oil
citric-acid	malted-barley-flour	soy-lecithin	citric-acid	citric-acid
enzymes	nonfat-dry-milk	wheat-flour	red	riboflavin
riboflavin	niacin	soybean-oil	spices	enzymes
niacin	palm-oil	citric-acid	enzymes	niacin
flavor	sodium-benzoate	enzymes	soybean-oil	flavor
dextrose	spices	flour-wheat-flour	black-pepper	caramel-color
vitamin-b	oregano	red	wheat-flour	dextrose
soy-lecithin	citric-acid	yeast-extract	niacin	soy-lecithin

# Guide

## Degree

- Rank by number of unique foods it is used in

## Weighted

- Rank by total number of foods it is used in

## Eigenvector

- Rank by co-usage with other important ingredients

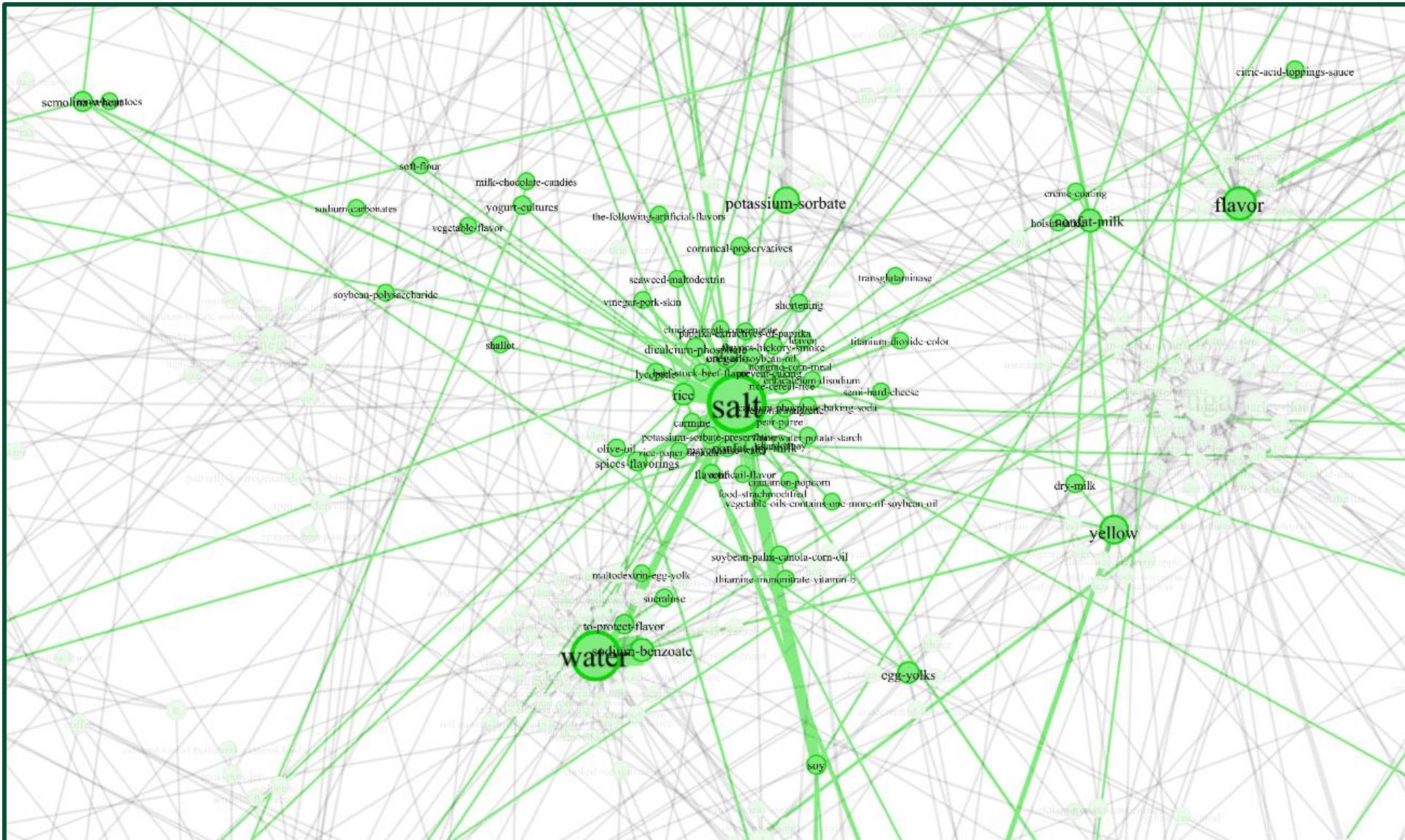
## Betweenness

- Rank by the number of links it provides between various groups of foods (shortest paths)

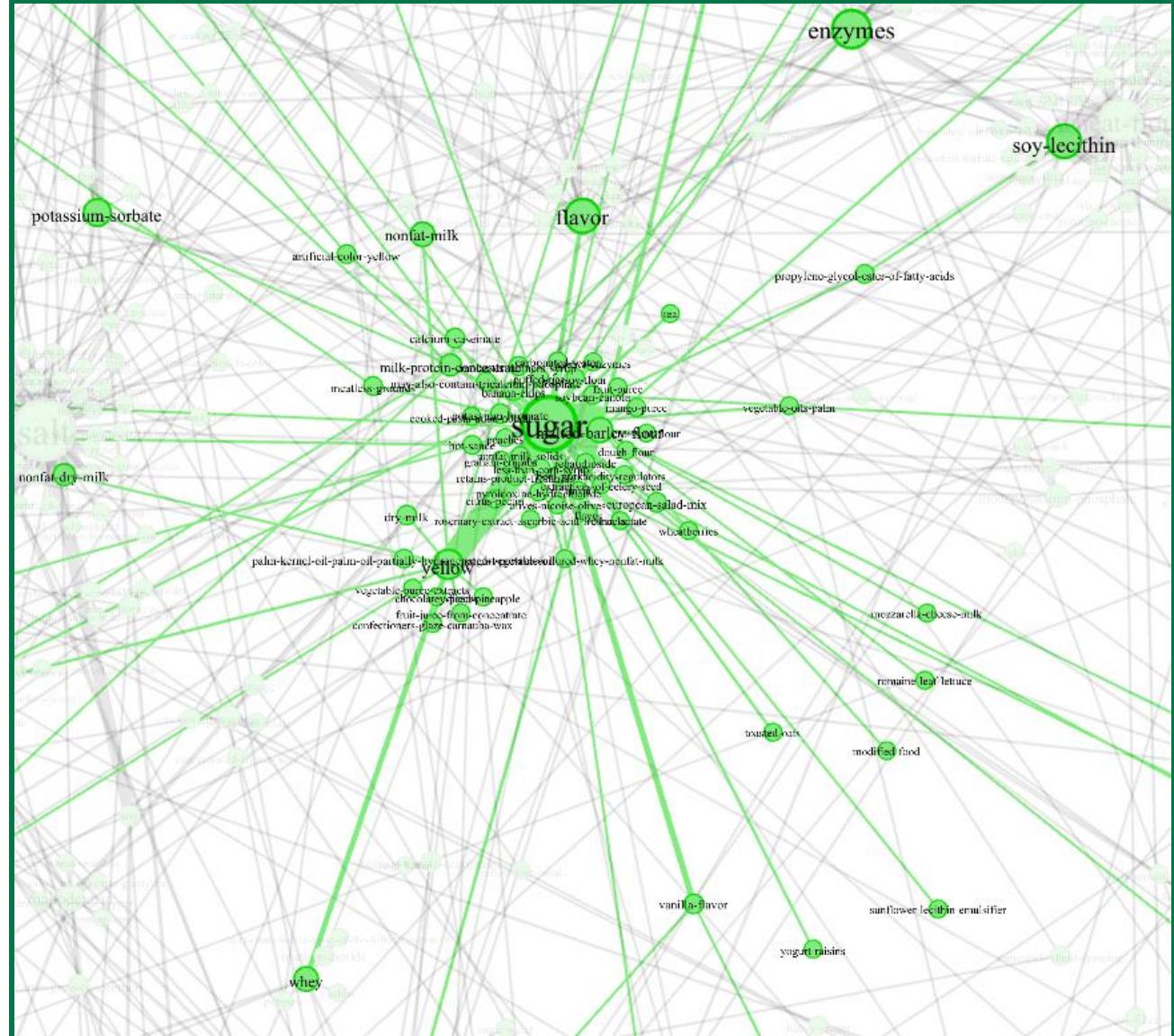
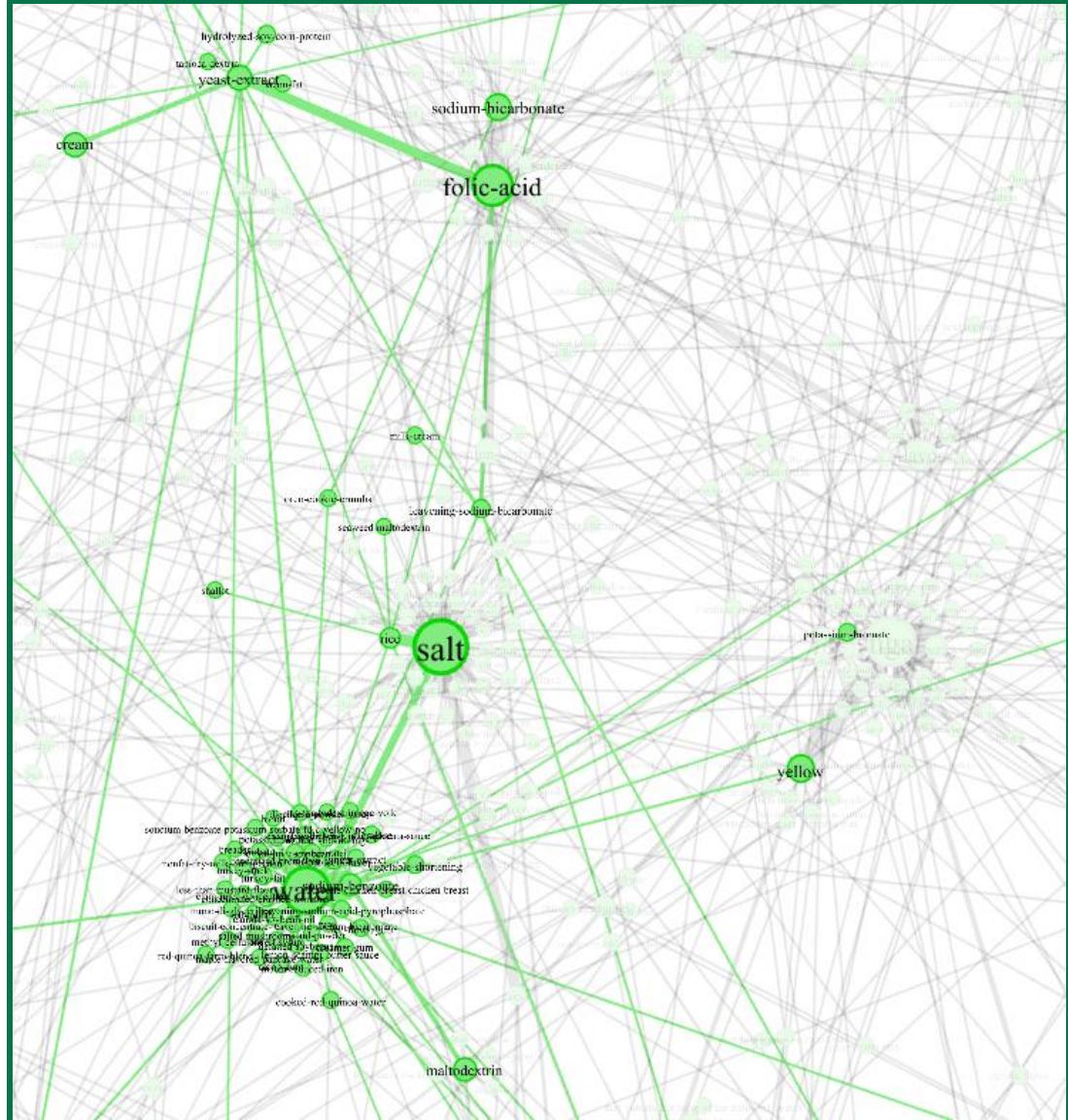
## Page Rank

- Rank by probability of actual usage (referenced by other ingredients in some food)

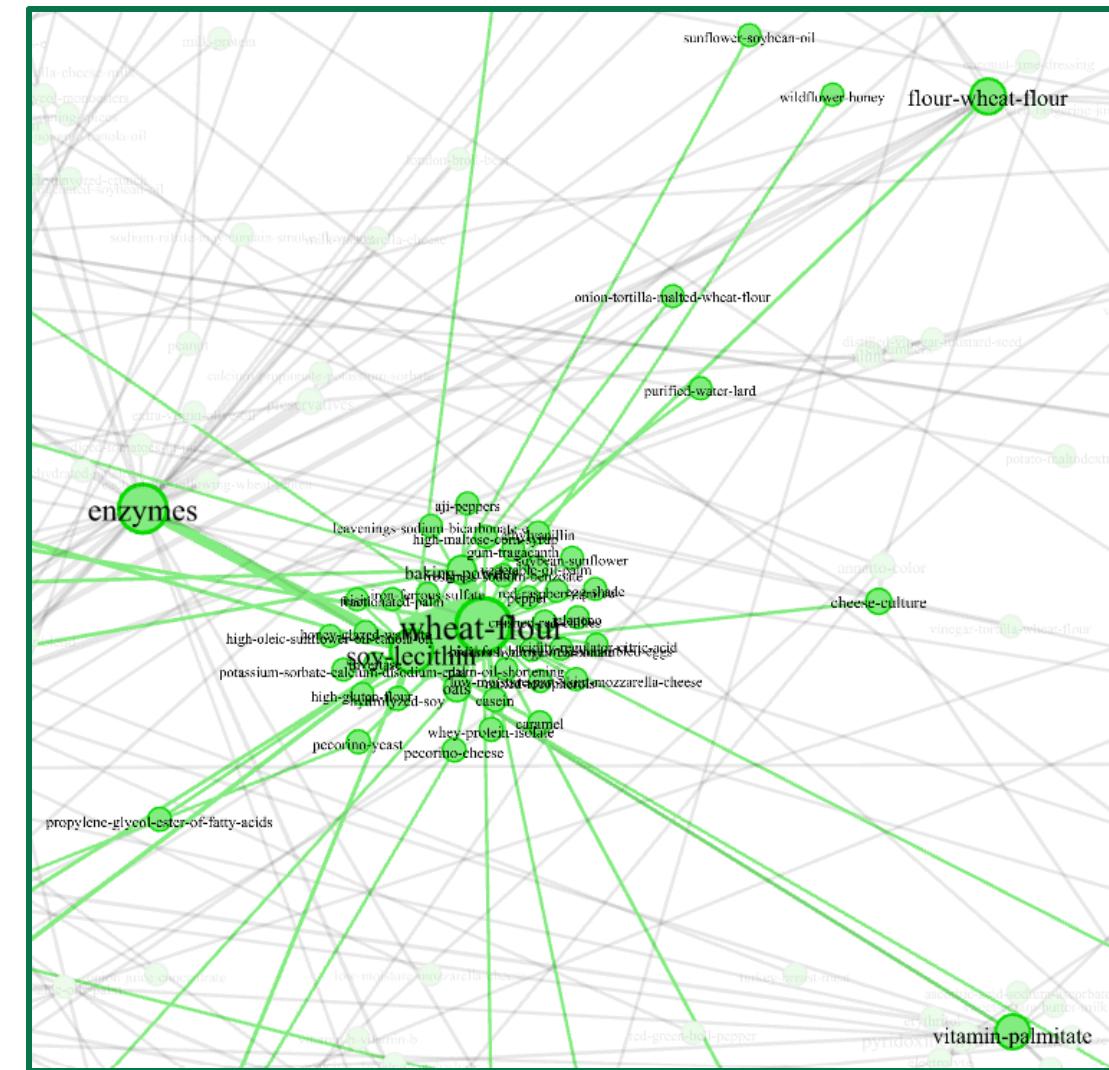
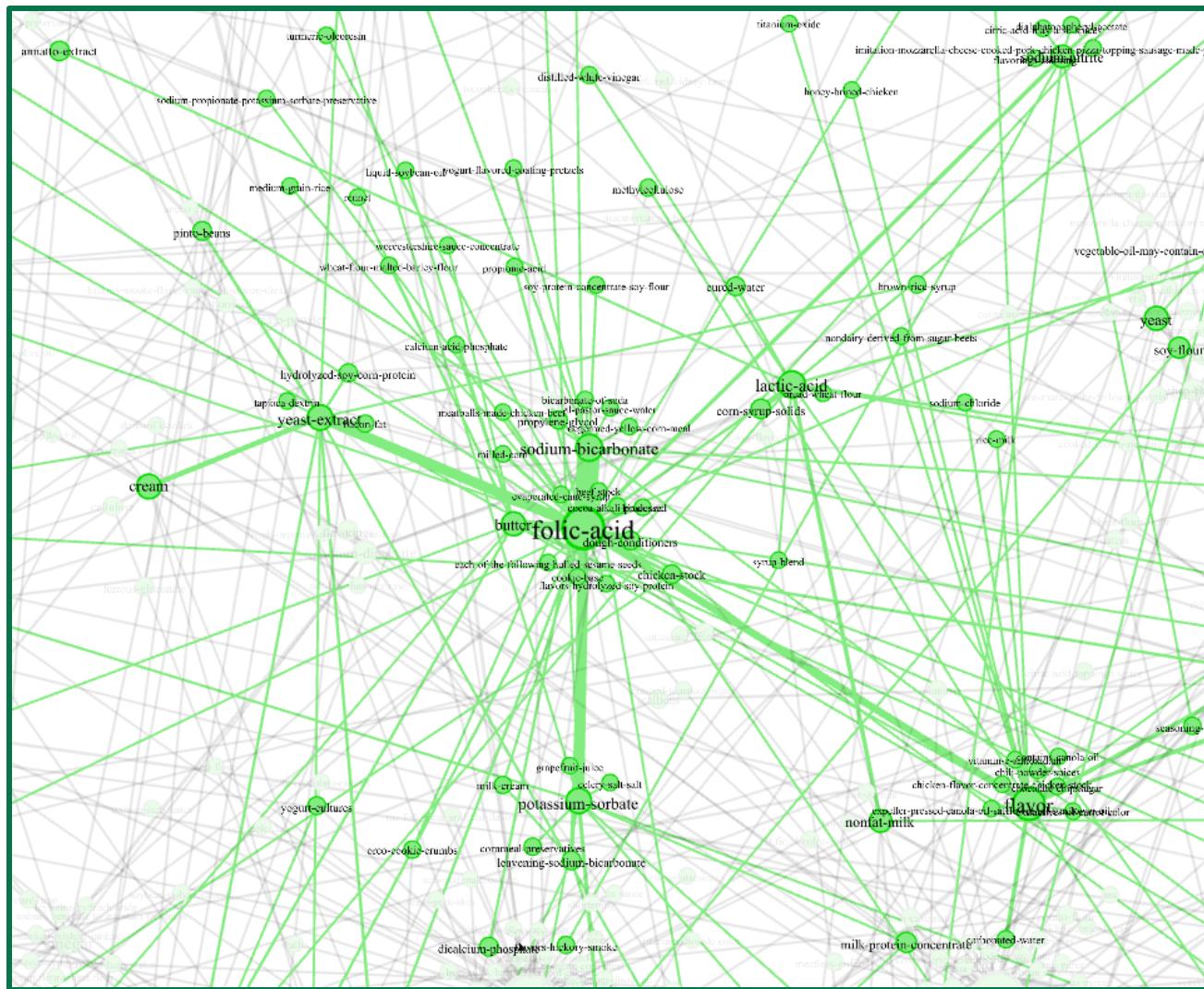
# 1.5 Ego Networks



# 1.5 Ego Networks



# 1.5 Ego Networks



# Conclusions & Implications





# Are Processed Foods Inevitable?

- In many of the rankings, processed foods dominate the twenty nodes.
    - Sugar: **Top 2** in 3 rankings, **Top 5** in 1 ranking
    - Folic Acid: **Top 5** in 4 rankings
    - Soy-based Products: **Top 10** in 4 rankings
    - Dyes (*Red, Yellow*): **Top 15** in 3 Rankings
  - Even without disregarding some non-negotiable ingredients such as salt and water, processed foods still dominate the top 5-15 ingredients used
-



# Are Processed Foods Inevitable?

- The network reveals a shift from an ingredient society to a grocery store society. High modularity (0.842) imply that traditional ingredient communities still exist. Low density and strong hubs show a few global ingredients dominating across many foods, many of which are processed. The small-world structure demonstrates how processing connects nearly all ingredients within a few steps despite the low average degree measure. Our food system shows the results of cultural globalization and the push for convenience.
  - Small-World Network
  - Scale Free Network
  - Numerical validation on the quantity of processed foods
-



# Thank You!

*Presented by Samuel Oliveira*

Please feel free to ask any questions



# Citations

<https://PMC7466326> - Food Ingredients Data

<https://claude.ai/> - Web Scraper Template

[Gephi - The Open Graph Viz Platform](#) - Graph Design



# Food For Thought:

*A Network Analysis of Supermarket Ingredients,  
a visualization of our food system, and  
a guide to understanding what we are consuming*

Presented by Samuel Oliveira