

# SOCIAL AND INFORMATION NETWORK ANALYSIS: PROBLEM 0

---

NGUYEN T. Hoang

2015-10-20

## Problem 0.1

This data was derived from comments on recipes on the Allrecipes website. An edge exists between ingredient  $e$  and ingredient  $j$ , if  $j$  was recommended as a substitute for  $i$  in at least 5% of the comments recommending substitutions, e.g. “Great recipe, but I used Brussels sprouts instead of broccoli...”

The network and the name for each nodes are contained in 2 files: **ingredient\_key.txt** and **ingredient\_substitutes.txt**.

## Questions and Answer

1. Find the size of the largest SCC. What percentage of the nodes is in the largest SCC?  
The largest strongly connected component is calculated in Snap.py with function `GetMxScc`.

**Answer:** The largest SCC of ingredient network has 244 nodes and it is 43% of all nodes in the network.

---

Listing 1: Load the network from file name and return the size of the largest SCC

---

```
1 import snap
2
3 IG = snap.LoadEdgeList(snap.PNGraph, "ingredient_substitutes.txt↵
    ", 0, 1)
4 maxScc = snap.GetMxScc(IG)
5 snap.PrintInfo(maxScc)
6 # Return 244 nodes for the larges SCC.
7 # There is total of 562 nodes in the network.
```

---

1. Find the ingredient most distant (via an undirected path within the WCC) from *co-coa powder*. What is the distance? You could use the Snap.py function `GetNodesAtHop`

to find nodes at a given distance.