

Recipe recommendation using ingredient networks

Chun-Yuen Teng¹, Yu-Ru Lin^{2,3}, Lada A. Adamic¹

¹School of Information, University of Michigan

²IQSS, Harvard University

³CCS, Northeastern University

Online recipes

Example: cupcakes More searches: [Ingredient](#) | [Nutrition](#) | [Advanced](#)

new at  [recipes »](#) [videos »](#) [menus »](#) [holidays »](#)



[Spicy Corn Salad](#) Corn, sweet onions, and jalapenos combine into a spicy summer salad. »



Crispy Herb Baked Chicken

By: DCANTER
"The secret ingredient is instant mashed potatoes, used to make the crispy coating."

 Rate/Review | Read Reviews (438)

 133  0 

1 of 23 Photos

Prep Time: 15 Min Cook Time: 45 Min Ready In: 1 Hr

Servings (Help) US Metric Original Recipe Yield 4 - 5 servings

Ingredients

2/3 cup dry potato flakes
1/3 cup grated Parmesan cheese
1 teaspoon garlic salt
1 (3 pound) chicken, skin removed, cut into pieces
1/3 cup butter, melted

Directions

kitchenapproved 

- Add to Recipe Box
- Add to Shopping List
- Print this Recipe
-
- supportingmembers**
- Create Menu
- Customize Recipe
- Kitchen-friendly View

What to Drink?

 [Summer Beers](#)

Our online recipes



- 46,337 recipes
- Each recipe includes directions, ingredients, nutrition info, cooking time, and regional information
- 1,976,920 reviews include ratings and text

Research questions

- ❑ What patterns emerge from the collective cooking knowledge aggregated in recipes?
- ❑ How can ingredient networks be used for predicting recipe ratings?

Recipe mining

- ❑ Cooking methods
 - ❑ Regional preferences
- ❑ Ingredients
 - ❑ Combination of ingredients
 - ❑ Modification of ingredients
- ❑ Predicting ratings:



OR



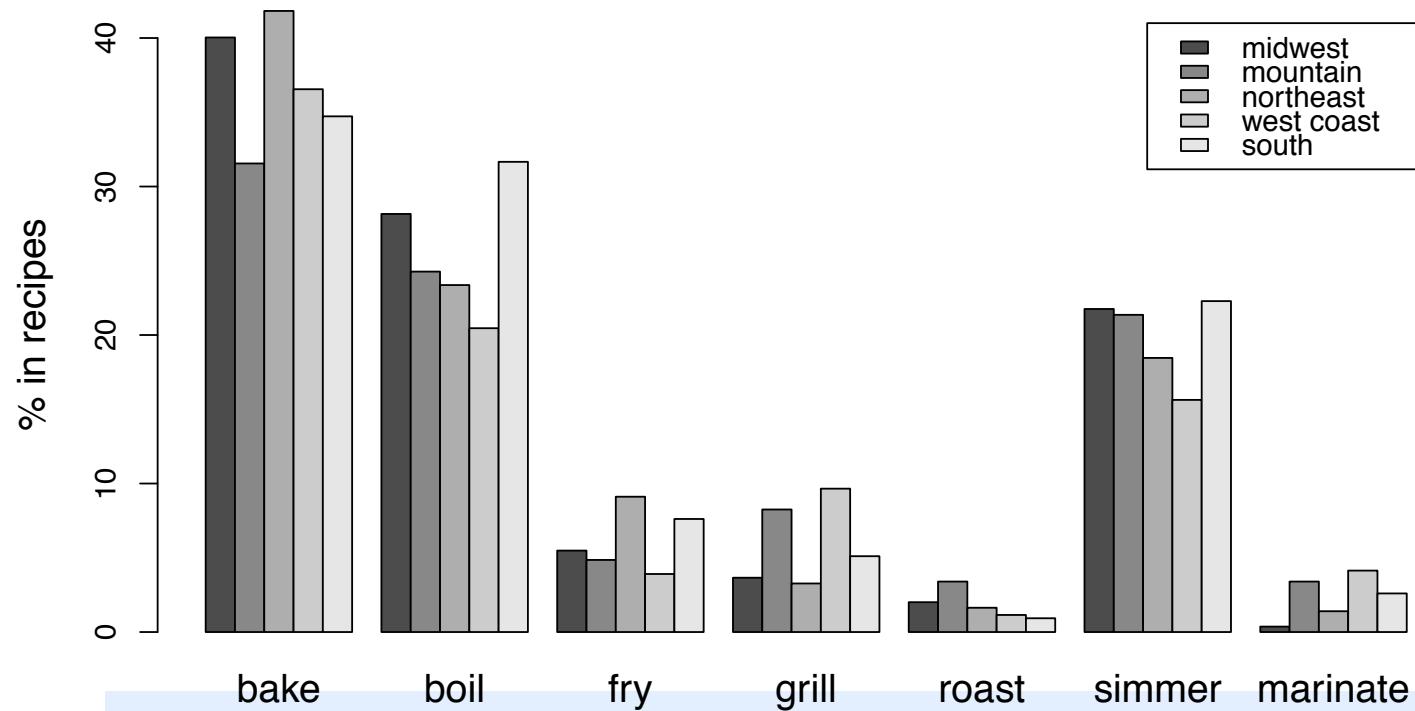
Credit: sonomaorganics.com

Cooking methods



US Regional Preferences

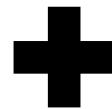
- Significantly varying preferences
- In the West, 42% of grilling recipes involve seafood relative to other regions (6%)



Ingredients



Combining ingredients

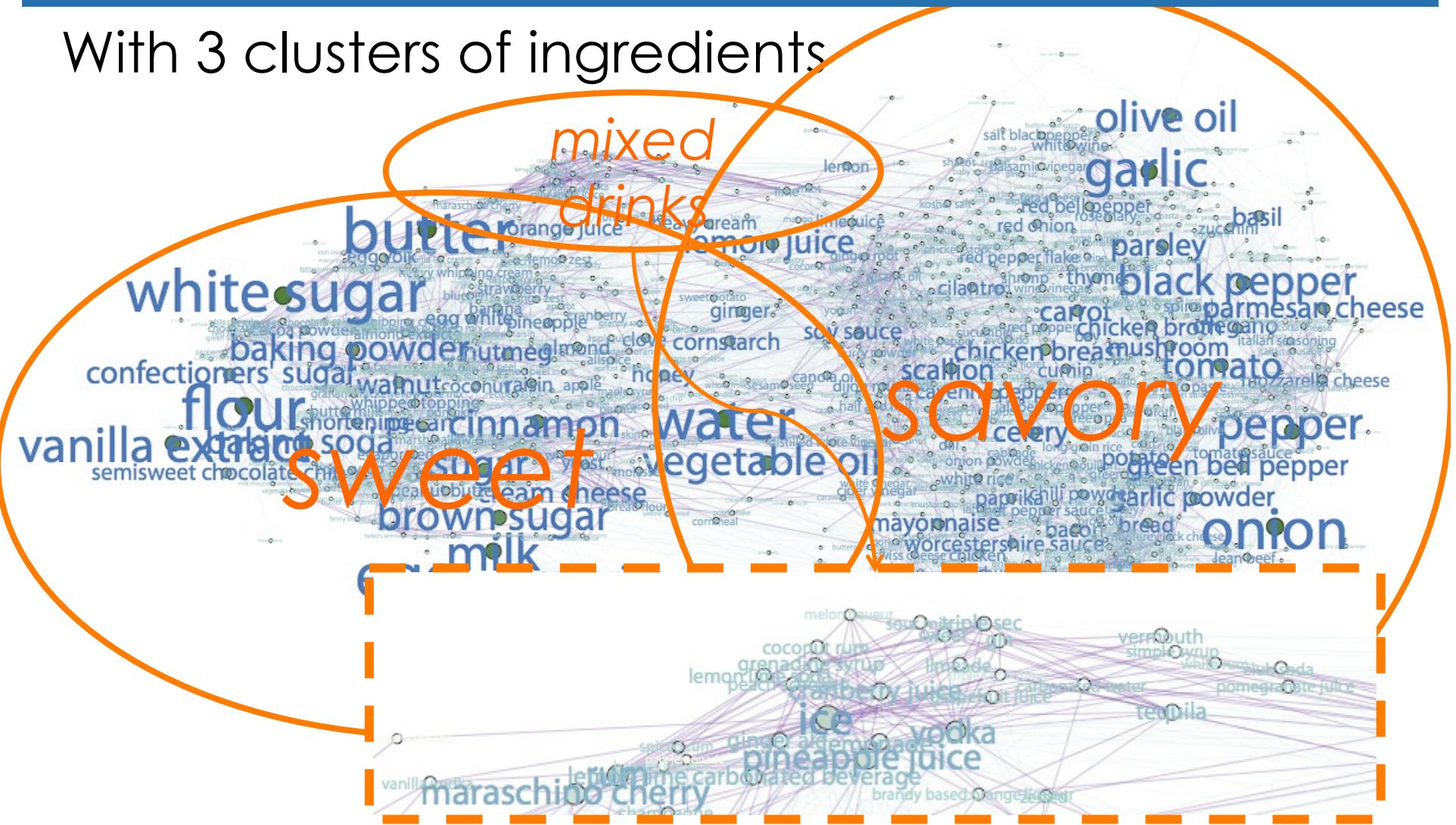


Complement network

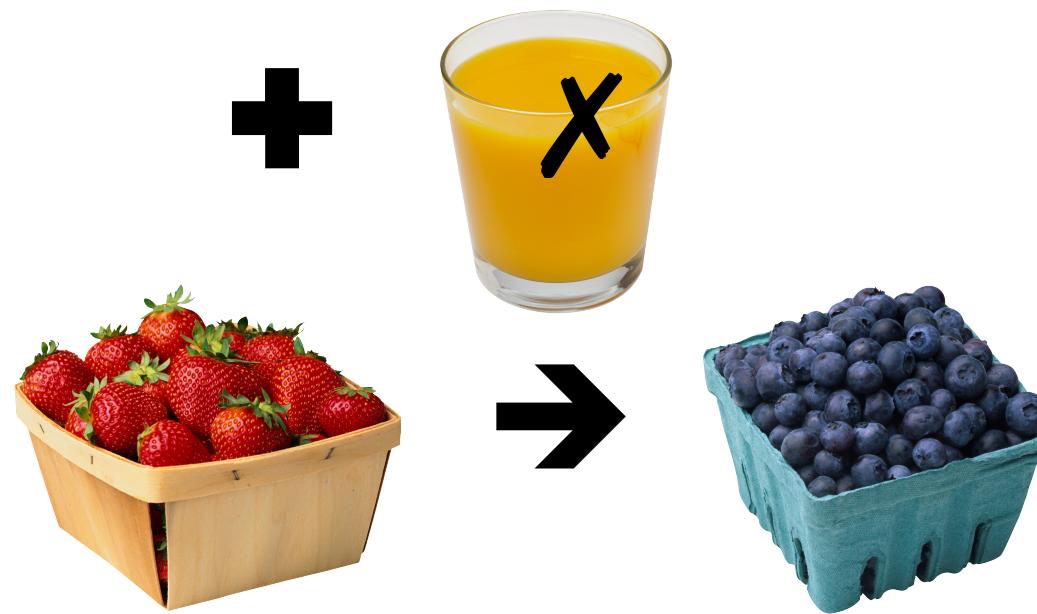
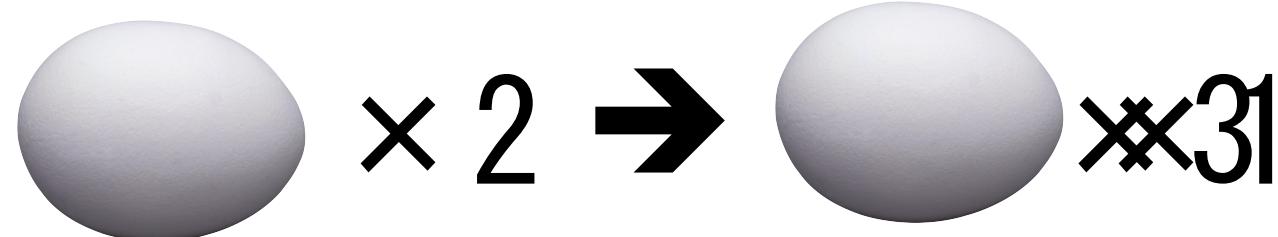
- Nodes: ingredients
- Undirected edges:
 - Weighted by pointwise mutual information
 $PMI = \log \left(P(a,b) / P(a)P(b) \right)$
 - $P(a,b) = (\# \text{ of recipes containing } a \text{ and } b) / (\# \text{ of recipes})$
 - $P(a) = (\# \text{ of recipes containing } a) / (\# \text{ of recipes})$
 - $P(b) = (\# \text{ of recipes containing } b) / (\# \text{ of recipes})$
 - Recipe rating and PMI of its ingredient pairs
 - Mean and minimum of PMI (no correlation with rating)
 - Max of PMI ($\rho=0.09$, $p < 0.001$)

Complement network

With 3 clusters of ingredients

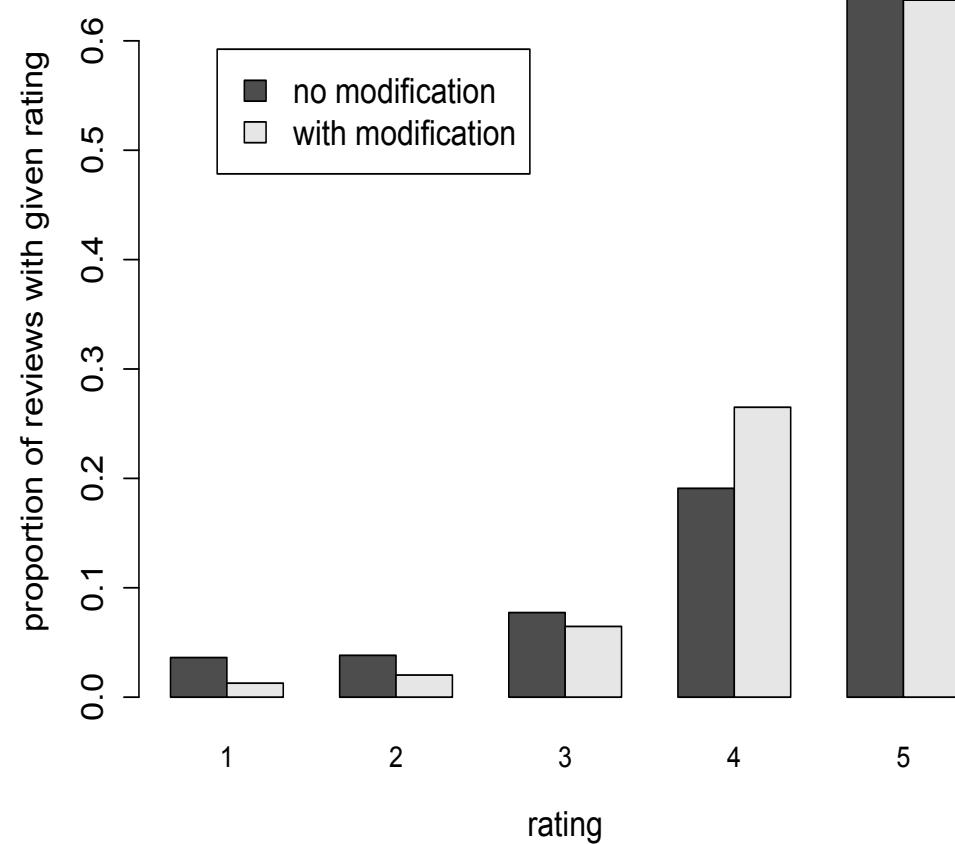


Recipe modification

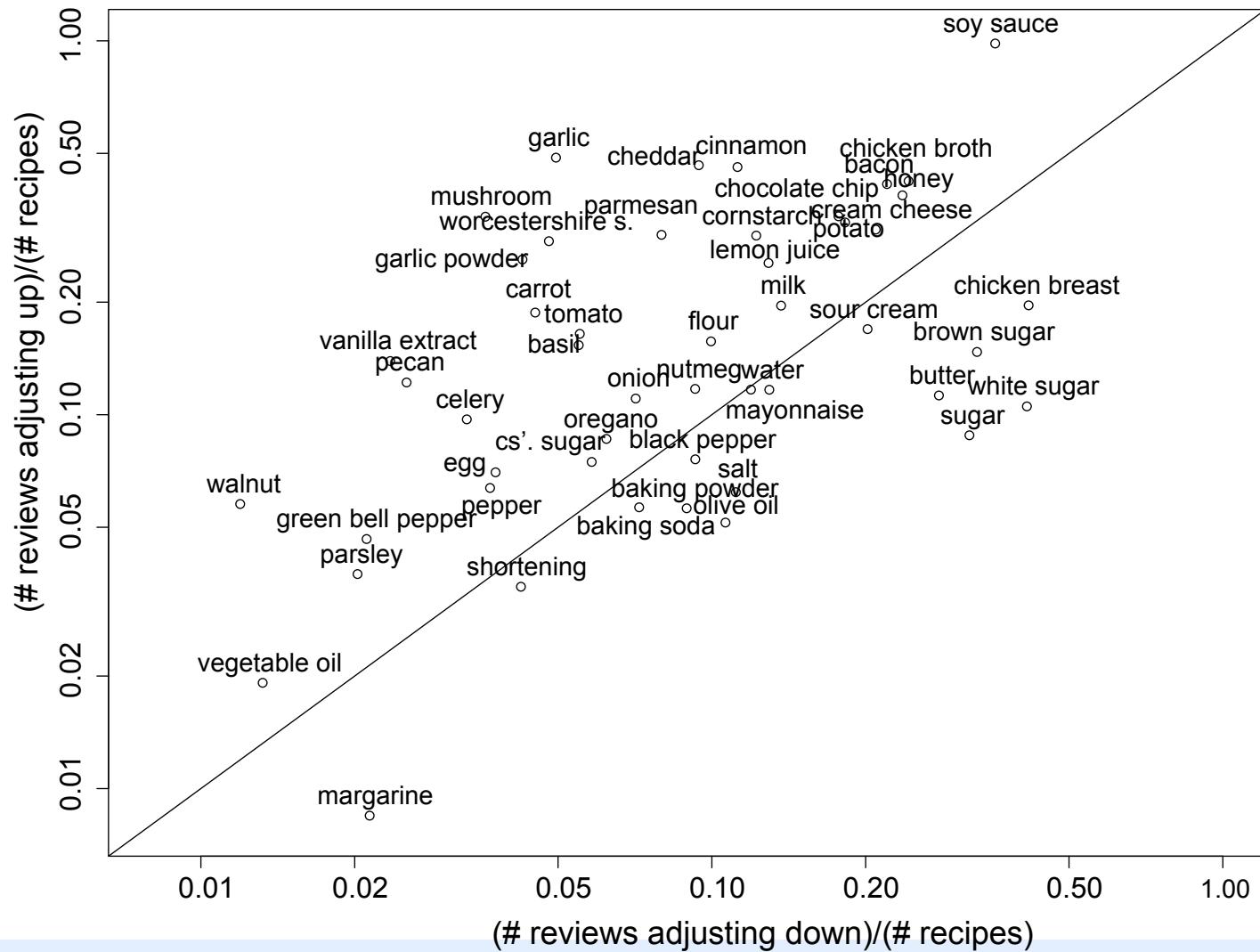


Recipe modification

- 60% of reviews contain “add”, “omit”, “instead”, “extra”, and 14 others.
- Reviews that include changes assign higher star ratings (4.49 vs. 4.39, $p<10^{-10}$)
- almost perfect but not quite (4 star) reviews often suggest modifications



Suggested modifications of quantities



Correlations between ingredient modifications

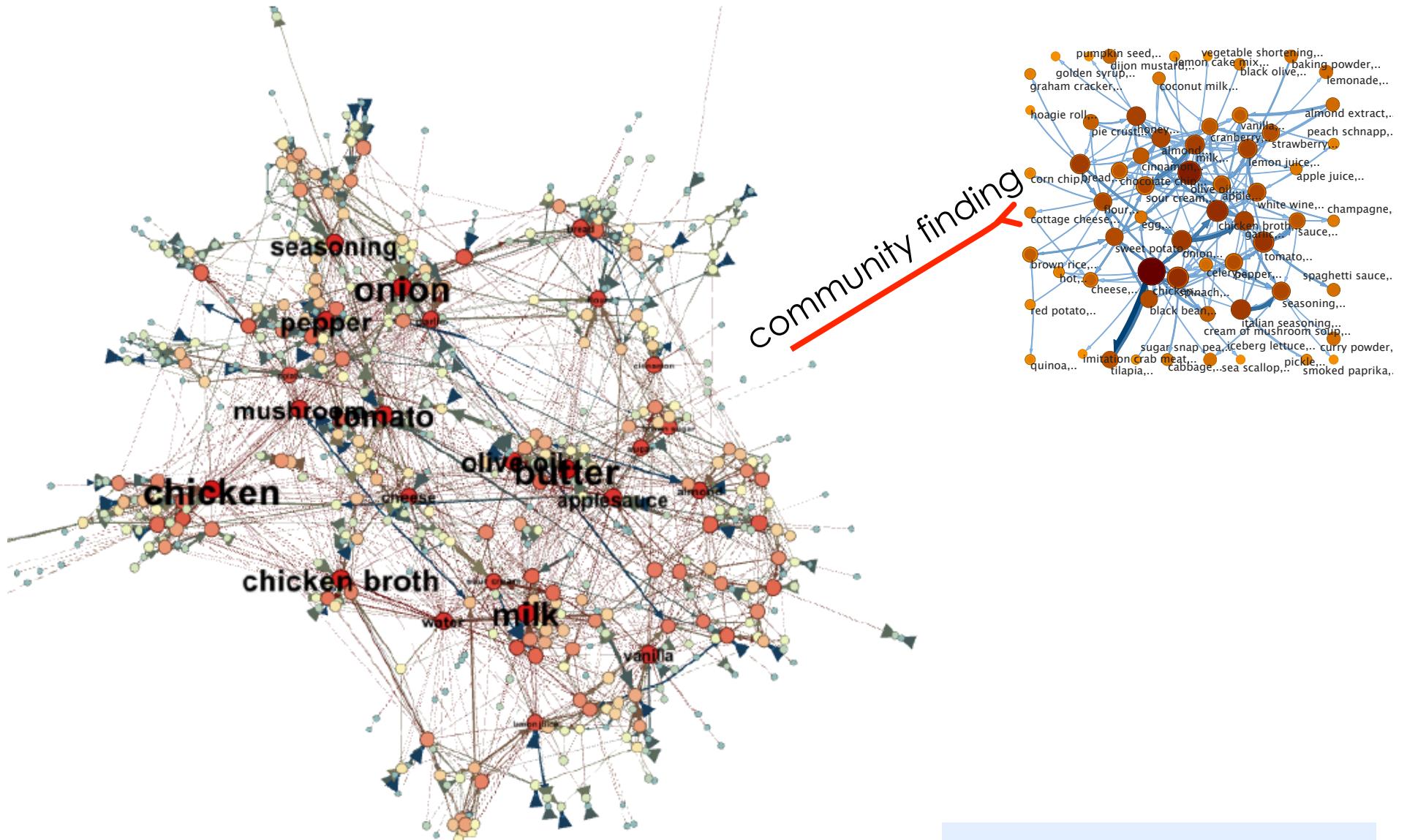
- Recipe freq. vs. deletion/recipe ($\rho = -0.22$)
- Recipe freq. vs. addition/recipe ($\rho = -0.25$)
- Recipe freq. vs. increase/recipe ($\rho = -0.26$)
- Correlations between ingredient modifications

| | addition | deletion | increase | decrease |
|-----------|----------|----------|----------|----------|
| # recipes | 0.41 | 0.22 | 0.61 | 0.68 |
| addition | | -0.15 | 0.79 | 0.11 |
| deletion | | | 0.09 | 0.58 |
| increase | | | | 0.39 |

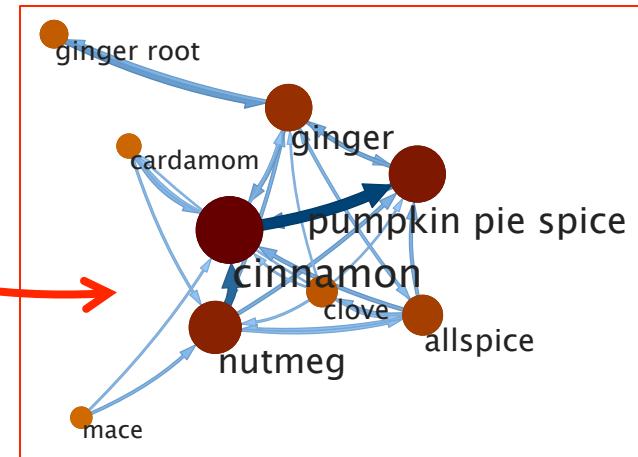
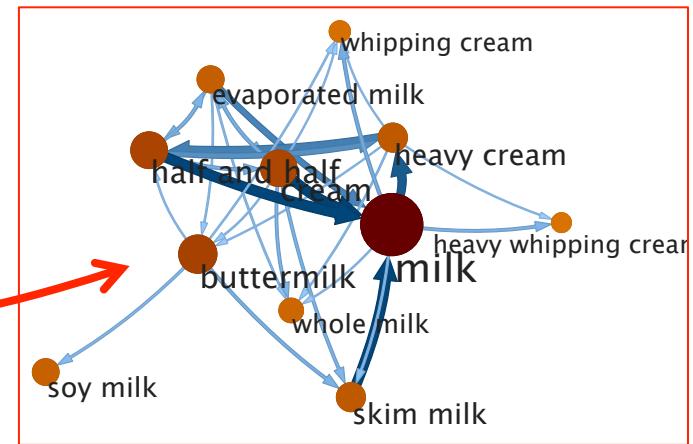
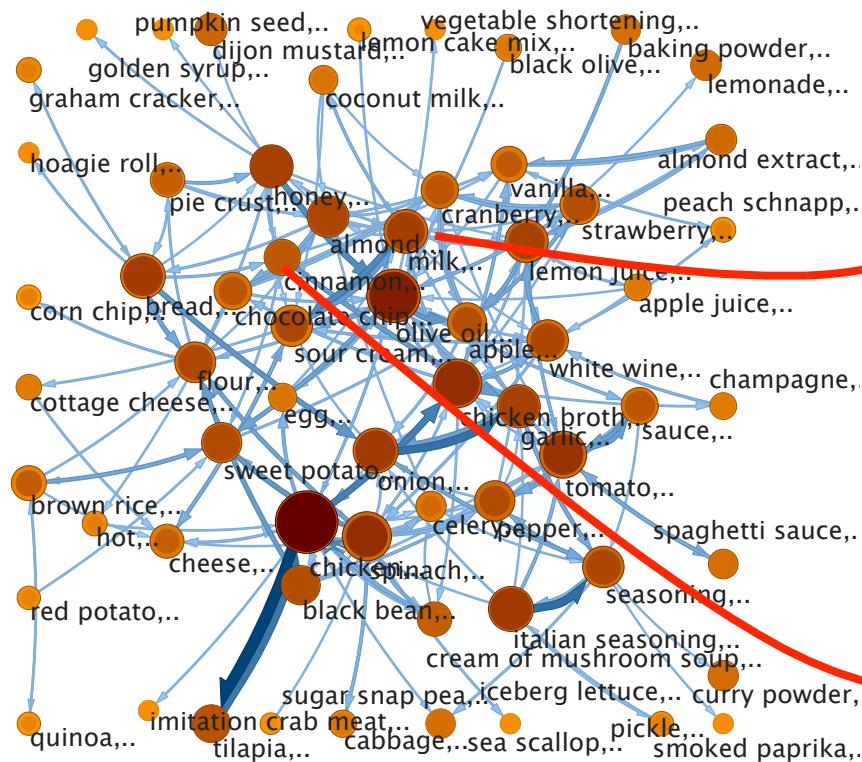
Substitution network

- Extract substitution relationships from comments
 - e.g. “I replaced the butter with sour cream”
 - “replace **a** with **b**”, “substitute **b** for **a**”, “**b** instead of **a**”
- Nodes: ingredients
- Edge weights = $p(\mathbf{b} \mid \mathbf{a})$, which is the proportion of substitutions of ingredient **a** that suggests ingredient **b**

Substitution network



Substitution network: communities



Examples of substitution

| main | other ingredients |
|-------------------|--|
| chicken | turkey, beef, sausage, chicken breast, bacon |
| olive oil | butter, apple sauce, oil, banana, margarine |
| sweet potato | yam, potato, pumpkin, butternut squash, parsnip |
| baking powder | baking soda, cream of tartar |
| almond | pecan, walnut, cashew, peanut, sunflower s. |
| apple | peach, pineapple, pear, mango, pie filling |
| egg | egg white, egg substitute, egg yolk |
| tilapia | cod, catfish, flounder, halibut, orange roughy |
| spinach | mushroom, broccoli, kale, carrot, zucchini |
| italian seasoning | basil, cilantro, oregano, parsley, dill |
| cabbage | coleslaw mix, sauerkraut, bok choy napa cabbage |

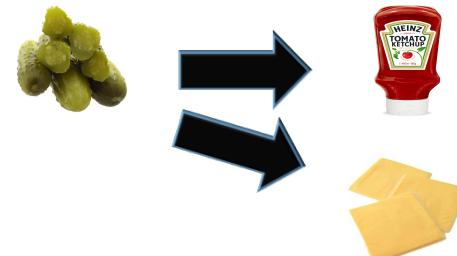
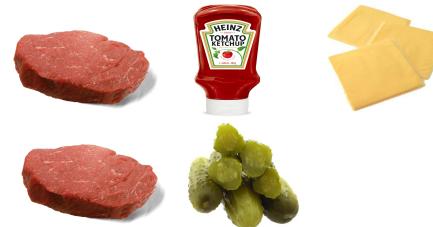
Substitution network and users' preference

Preference network

- Create an edge from ingredient **a** to **b** if $\text{rating}(a) < \text{rating}(b)$

- ex:

- Recipe X contains
- Recipe Y contains
- $\text{Rating}(X) > \text{Rating}(Y)$



Substitute network and users' preference

- ❑ Weight of preference network
 - ❑ $\text{PMI}(a \rightarrow b) = \log(p(a \rightarrow b) / p(a)p(b))$
 - ❑ where $p(a \rightarrow b) = (\# \text{ of recipe pairs from } a \text{ to } b) / (\# \text{ of recipe pairs})$
- ❑ Correlations between preference network and substitute network ($\rho = 0.72$, $p < 0.001$)

Prediction task

- Given a recipe pair with overlapped ingredients, determine which one has the higher rating



Prediction task

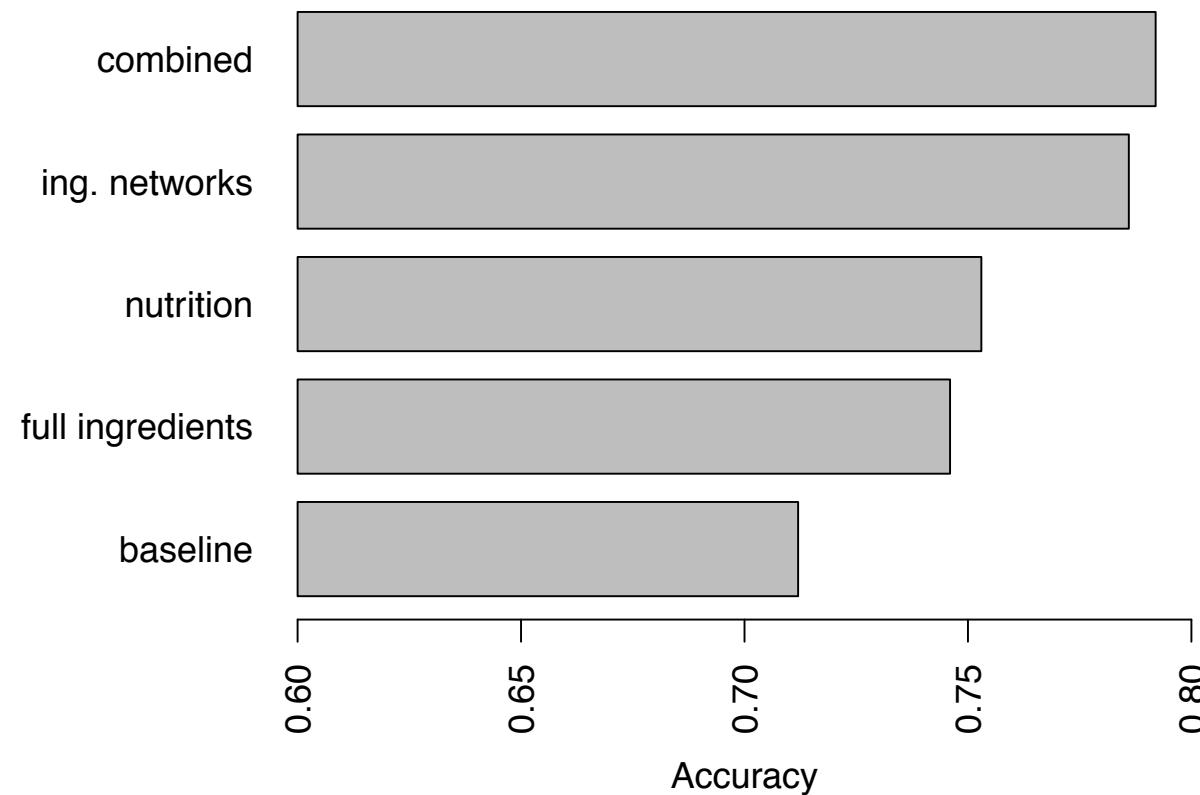
- Features
 - Baseline
 - Cooking methods, preparation time, the number of servings
 - 1000 popular ingredient list
 - Binary vector indicating the occurrence of ingredients
 - Nutrition
 - Calories, carbohydrates, fat, etc.
 - Ingredient networks
 - Network positions (centrality) and communities (SVD)
 - Combined set
 - Everything listed above

Prediction task

- 62,031 recipe pairs (X,Y)
 - where rating(X) > rating(Y)
 - ≥ 10 user reviews
 - $\geq 50\%$ users have rated both recipes
 - Cosine similarity of ingredients $(X,Y) > 0.2$
- Train with gradient boosting tree
 - balanced dataset
 - 2/3 for training, 1/3 for testing
 - Evaluate based on accuracy

Prediction performance

- Ingredient network features lead to impressive performance

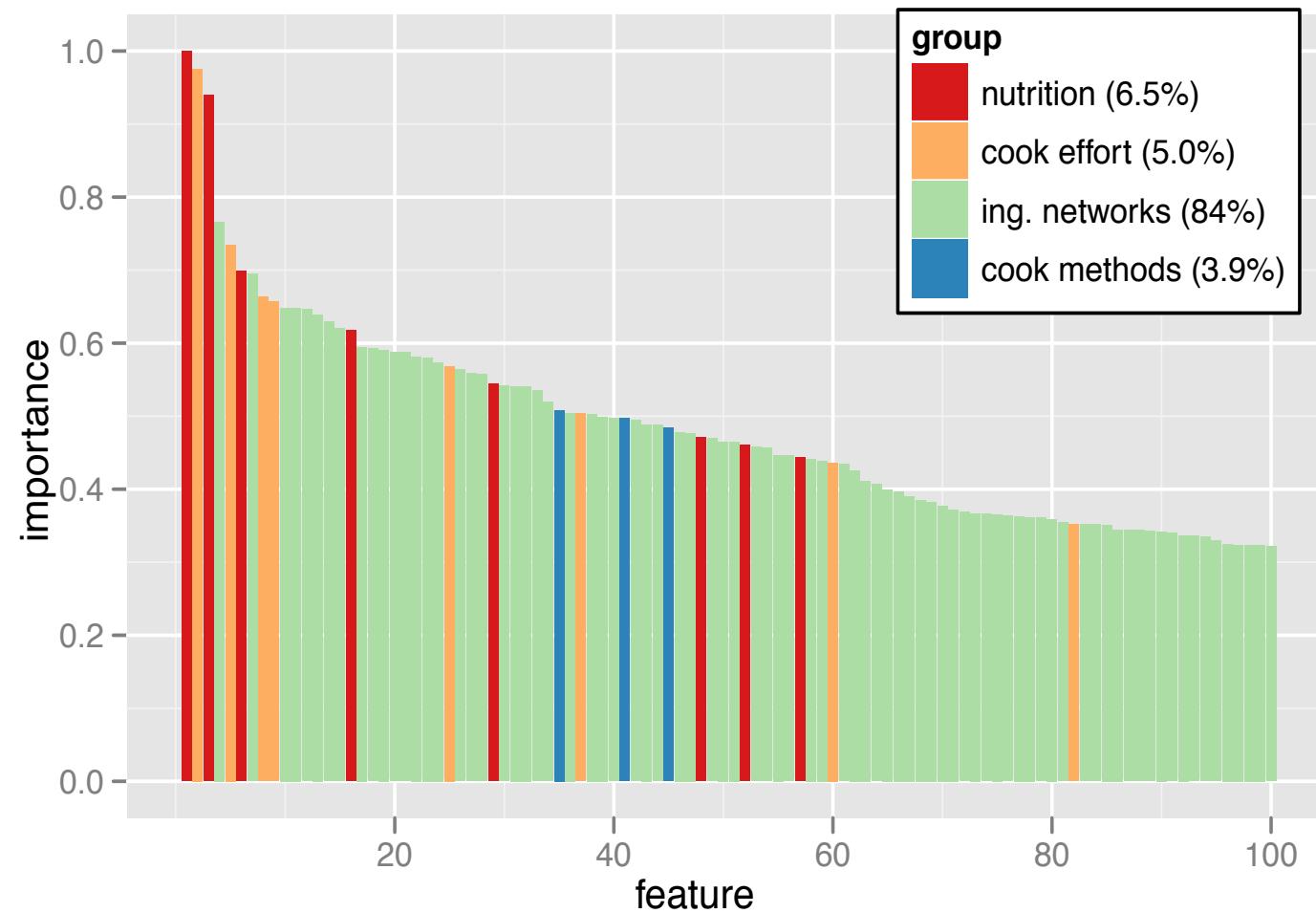


Relative importance of features

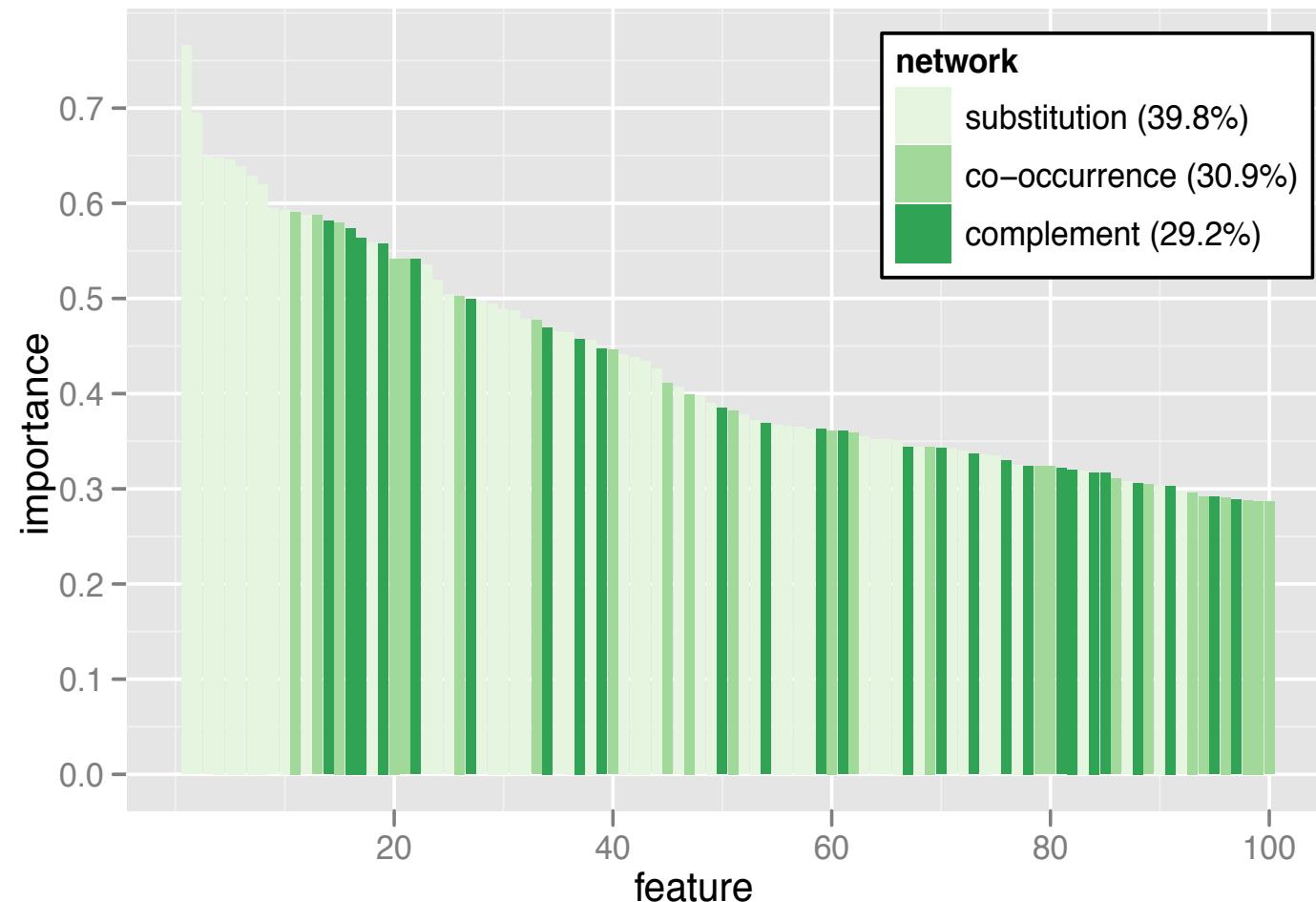
- Relative importance of feature x^j is the sum of squared improvement over all internal nodes
- i^k is the empirical improvement by the k-th node splitting on x^j

$$imp(j) = \sum i_k^2 I(\text{splits on } x^j)$$

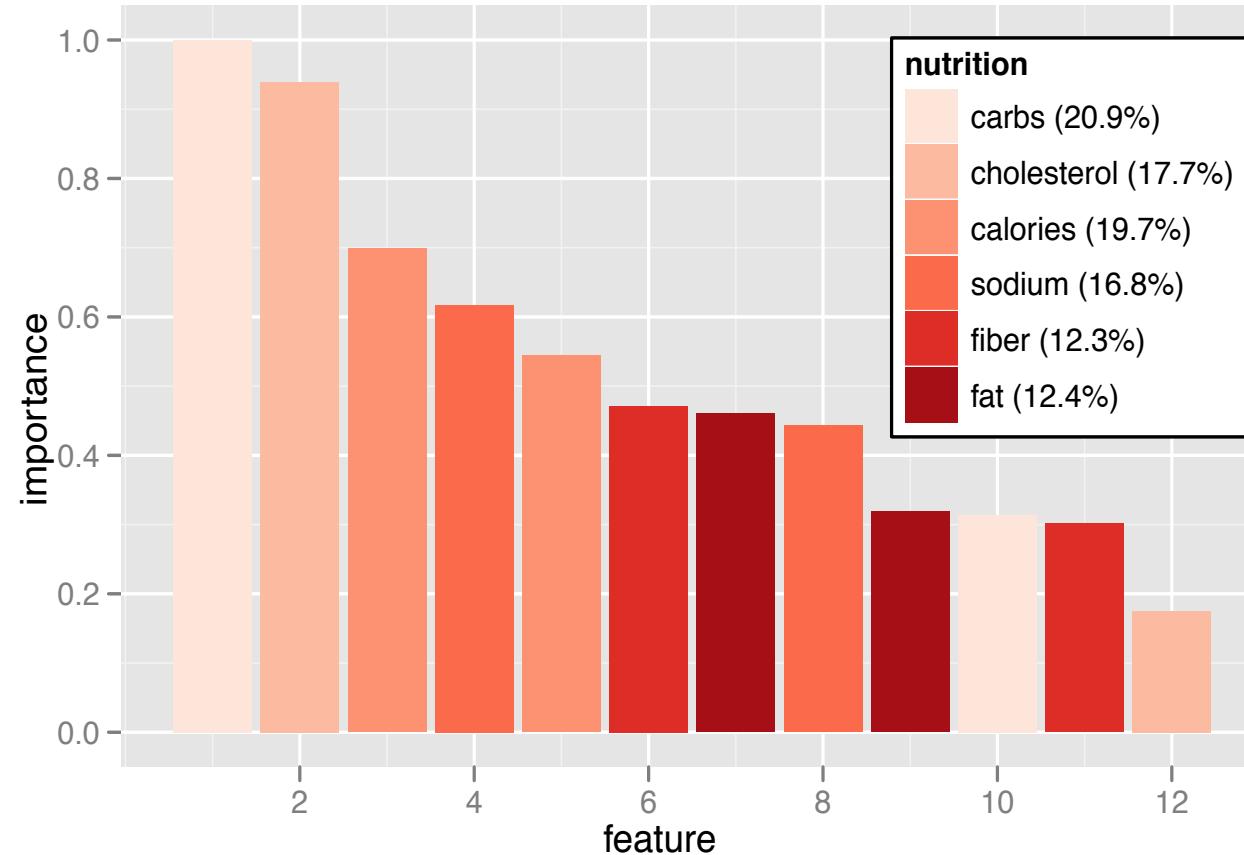
Relative importance of features



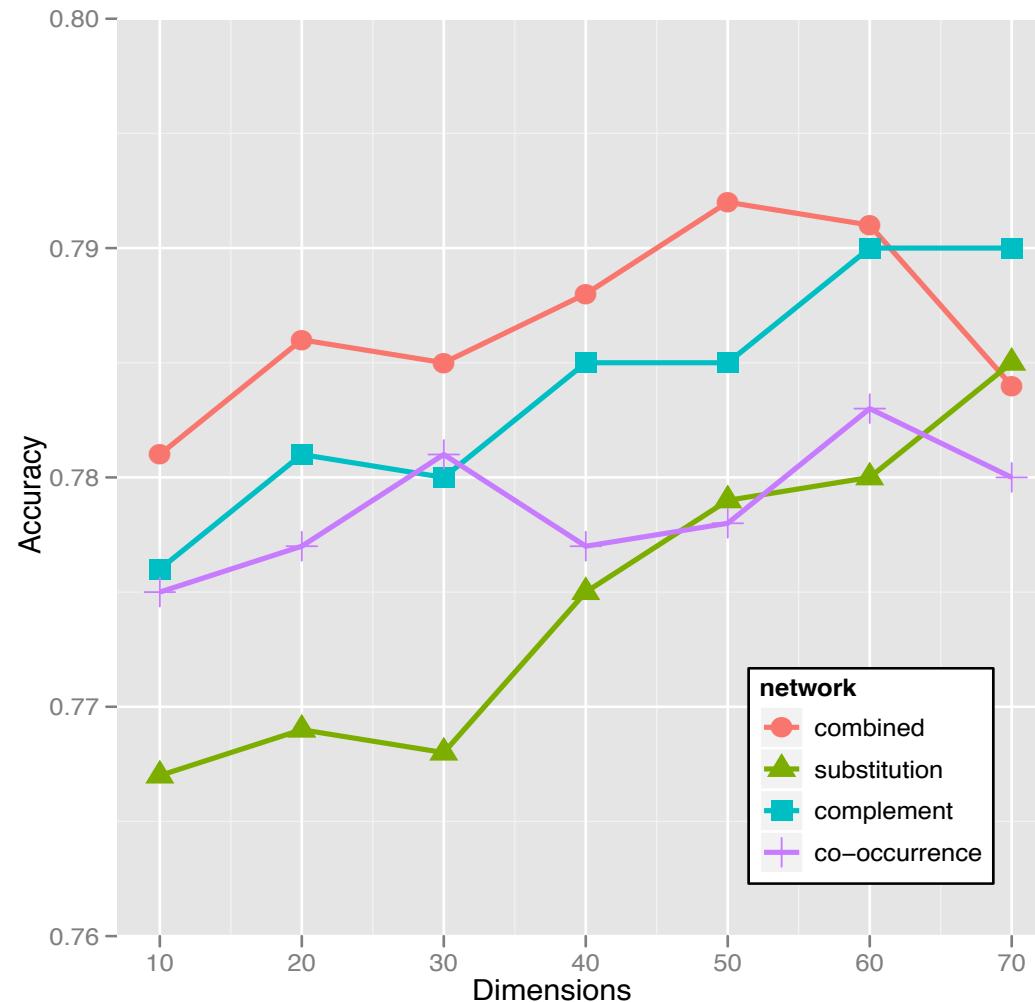
Relative importance of network features



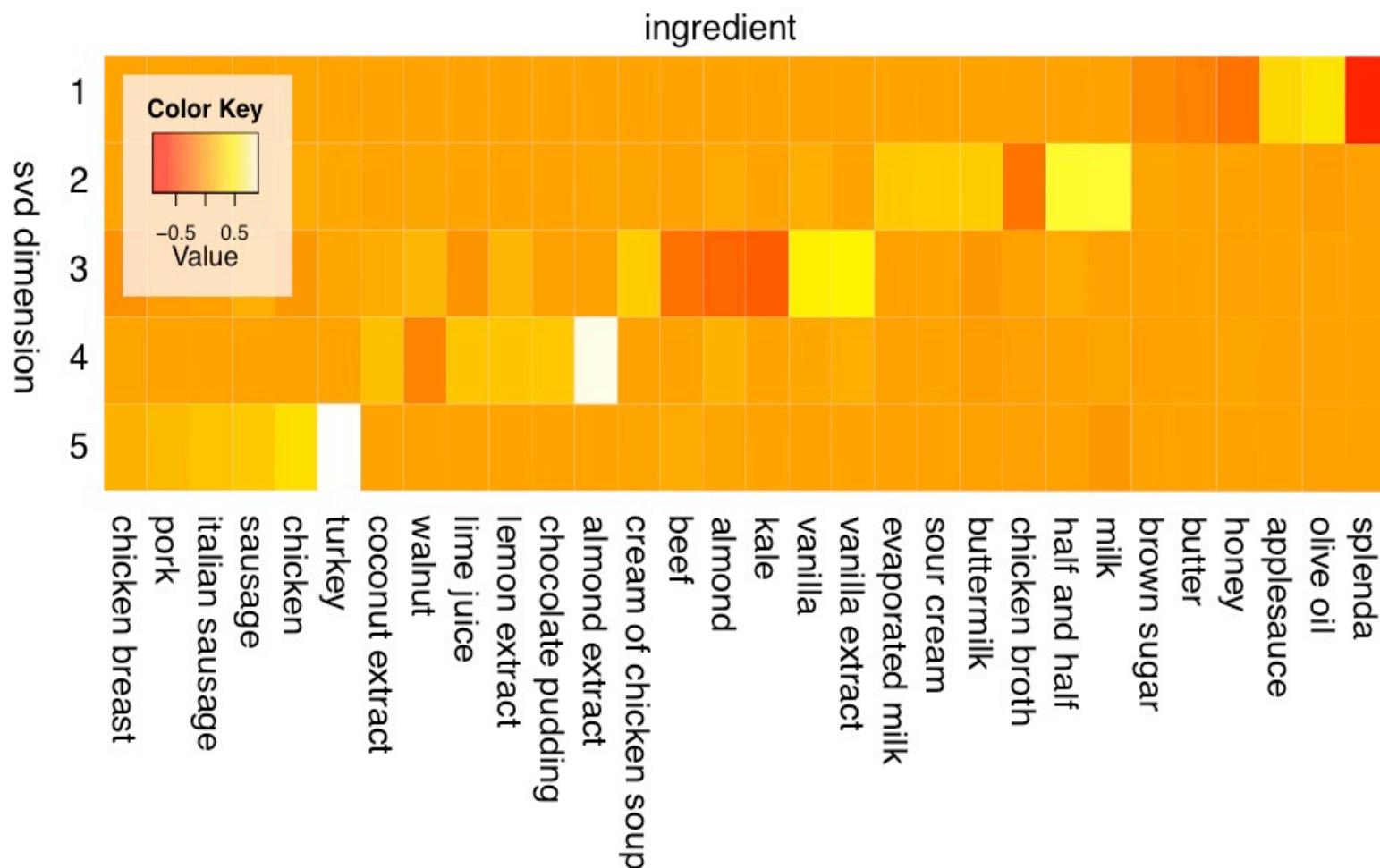
Relative importance of features from nutrition



Prediction performance vs. dimensionality



Influential substitution communities



Conclusion: recipes encode our collective cooking knowledge

- regional preferences for cooking methods
- complementarity of ingredients
- substitutability of ingredients
- complement and substitute networks encode users preferences and can be used to effectively predict recipe ratings

Future work

- ❑ Extend ingredient networks by incorporating cooking methods
- ❑ Build a recommender system that incorporates users' background info (ex: region & diet)
- ❑ More info
 - ❑ <http://netsi.org>