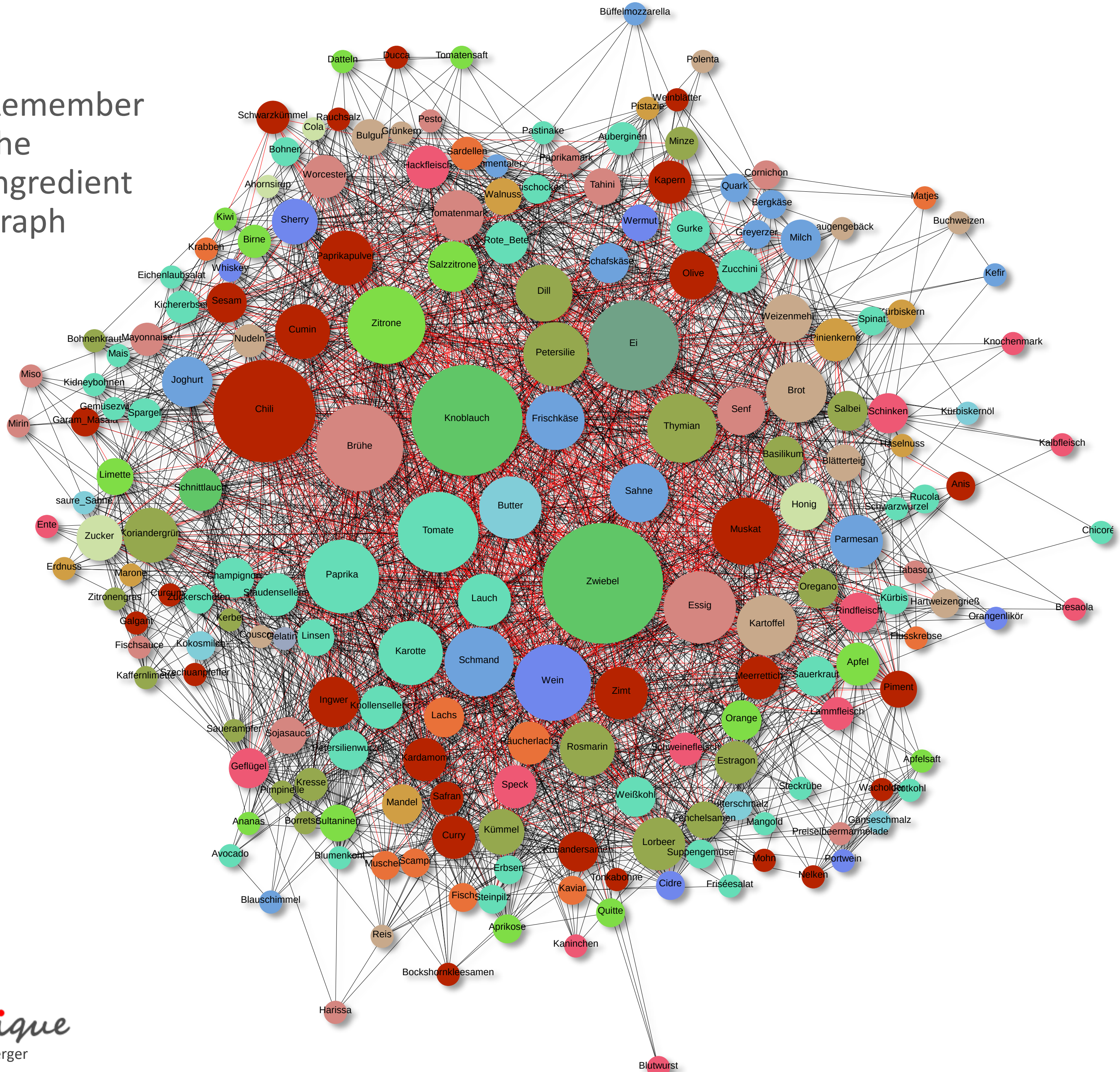
fruschtique

by Norbert Luttenberger

Culinary Interpretation

What ingredient graphs can tell

Remember the ingredient graph

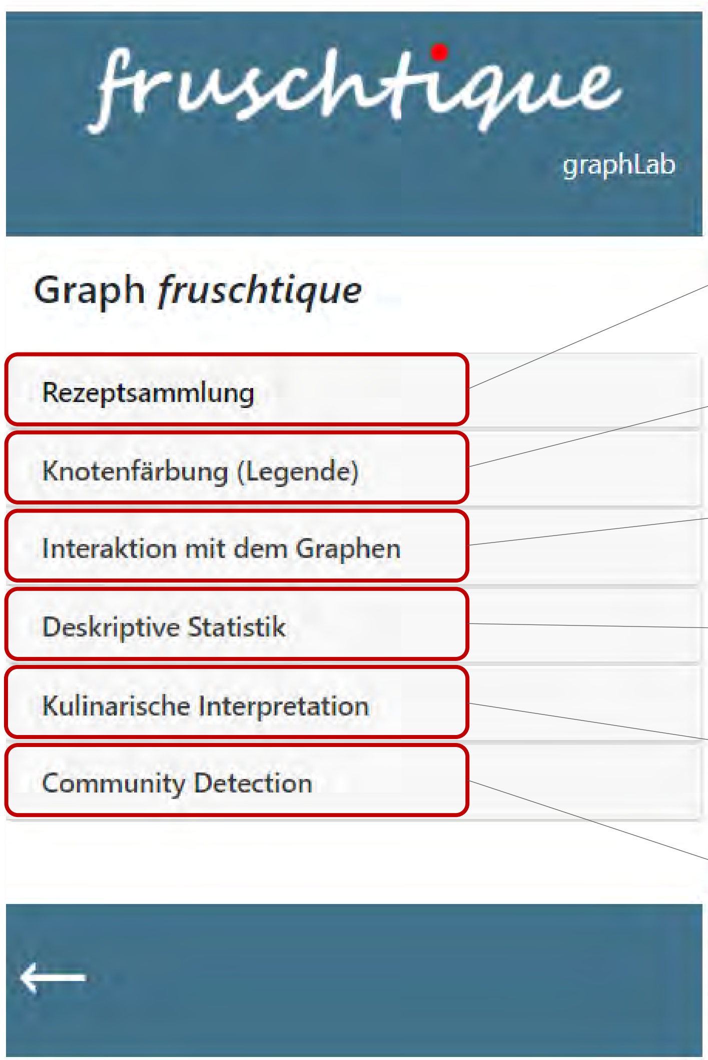


fruschtique
by Norbert Luttenberger

Descriptive statistics for ingredient graphs

Dedicated statistics for culinary interpretation

Viewer page



short description of the graph data source

legend for graph coloring

modes for direct user interaction with graph

descriptive statistics for the graph, see below

culinary interpretation of the graph, see below

community detection, under preparation


Graph-theoretic interlude

Basic definitions

A graph G is an ordered pair (V, E), where V is called the set of vertices, and E is called the set of edges of G.

graph

V is a finite set and $E \subseteq \binom{V}{2}$ or $E \subseteq \{X \subseteq V: |X| = 2\}$ is a set of pairs of elements in V.

vertices, edges

The edge $e = \{u, v\} \in \binom{V}{2}$ is also denoted by $e = uv$.

If $e = uv \in E$ is an edge of G, then u is called adjacent to v.

adjacent

If the vertex u is on edge e, the vertex u is said to be incident on e.

incident

Some graph types

A directed graph is a pair $G = (V, A)$, where V is a finite set and $A \subseteq V \times V$. The edges of a directed graph are also called arcs. A directed graph is different from an undirected graph in that an arc is defined by an ordered pair (u, v) of two nodes. This definition allows us to distinguish arc (u, v) going from node u to node v from arc (v, u) going from v to u.

directed graph

A weighted or edge-labeled graph can be defined as a triple $G = (V, E, w)$ where $w: E \rightarrow val$ is a function mapping edges to their values. For edge-labeled graphs val can be any type. We talk about edge-colored graphs, if val is restricted to $\mathbb{N}\backslash\{0\}$ The same applies to node-labeled graphs.

weighted graph

A simple graph (or just graph), is an unweighted, undirected graph containing no graph loops or multiple edges.

simple graph

A complete graph is a graph in which each pair of vertices is connected by an edge, i.e., a graph in which each pair of vertices are adjacent is a complete graph. The complete graph with n vertices is denoted K_n and has $\binom{n}{2} = \frac{1}{2}\, n \cdot (n-1)$ edges.

complete graph

A multigraph is a pair $G = (V, E)$ where V is a finite set and E is a multiset of elements from $\binom{V}{2} \cup \binom{V}{1}$, i.e., we also allow multiedges and loops.

multigraph

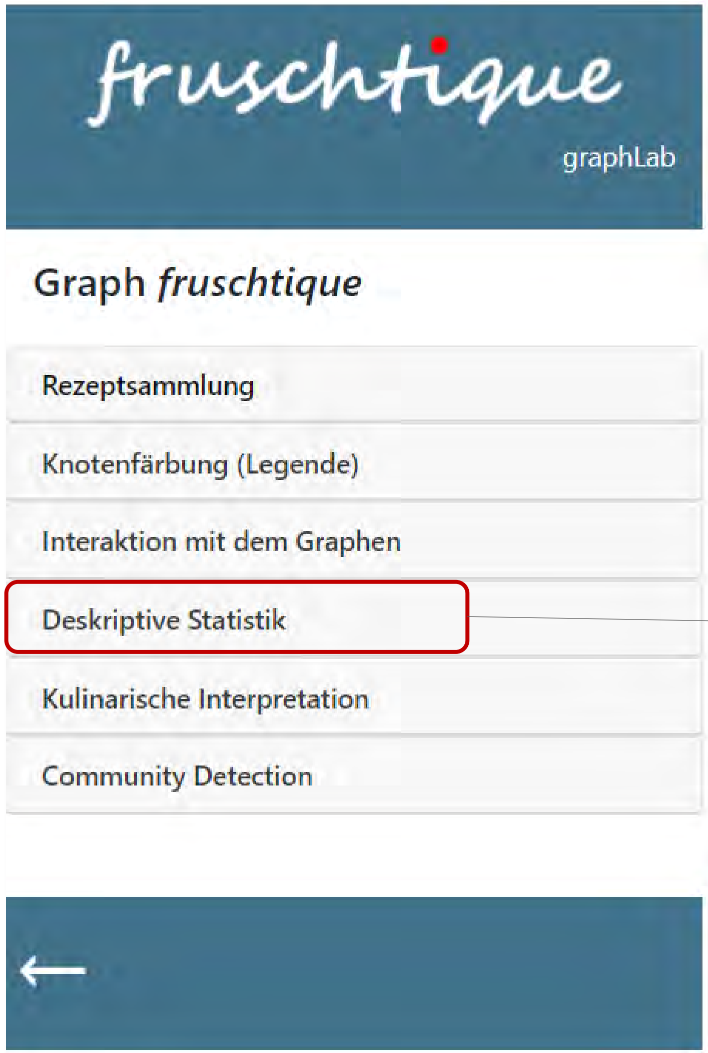
Ingredient graph

An ingredient graph $IG = (V, E, w, c, o)$ is an undirected, loop-free, weighted, and node-labeled graph where

- V is a finite set of ingredient nodes identified by their ID and
- $E \subseteq \binom{V}{2}$ is a set of edges.
- w is a function that assigns a weight value from $\mathbb{N}\backslash\{0\}$ to each edge,
- c is a function that assigns a class value to each node, where this value comes from $class = \{alc, carb, condi, egg, etc, fat, fish, fruit, herb, meat, milk, nuts, onion, spice, sweet, veg\}$
- o is a function that assigns an occurence value ω to each node with $\omega \in \mathbb{N}\backslash\{0\}$.
- The occurence value can easily be transformed into a prevalence value by dividing it by the number of recipes in the collection under scrutiny.

Descriptive statistics

Viewer page



descriptive statistics for the ingredient graph

Two views on descriptive statistics for ingredient graphs

Graph-oriented

- Exploit structural properties of the ingredient graph
- Examples: Degree, betweenness, … of ingredient nodes

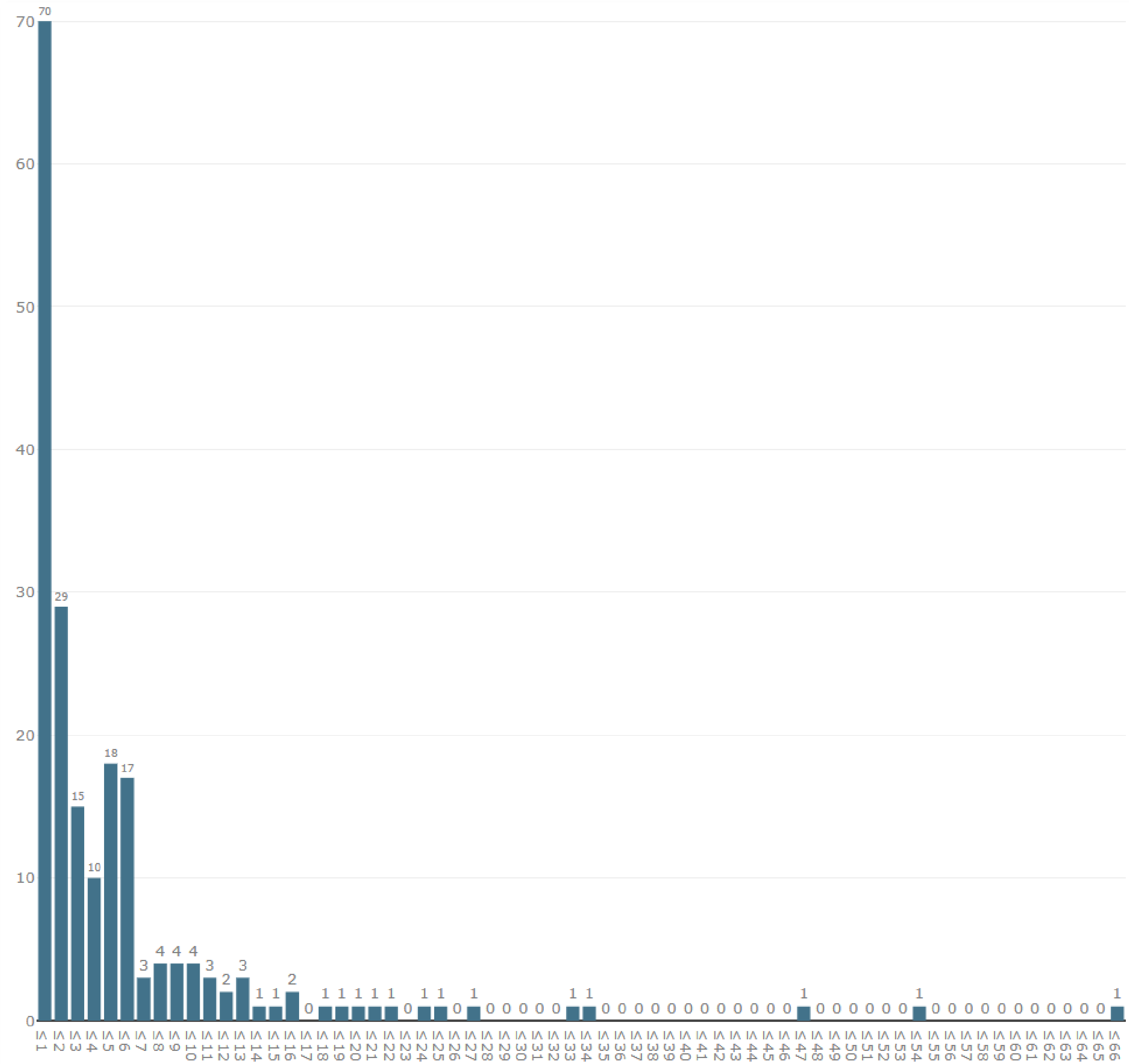
Recipe-oriented

- Relate back to contributing recipe(s) (kind of inverted index approach)
- Example: Ingredient prevalence

Ingredient prevalence

ingredient prevalence	$prev_{igt} = \frac{\#rcps\ using\ igt}{\#rcps}$
	computed from `occurs` value contained in `ingredients` section of `combinedGraph.json`
	typically: few high, many low values; "long tail"
	gives hint to diversity of collection

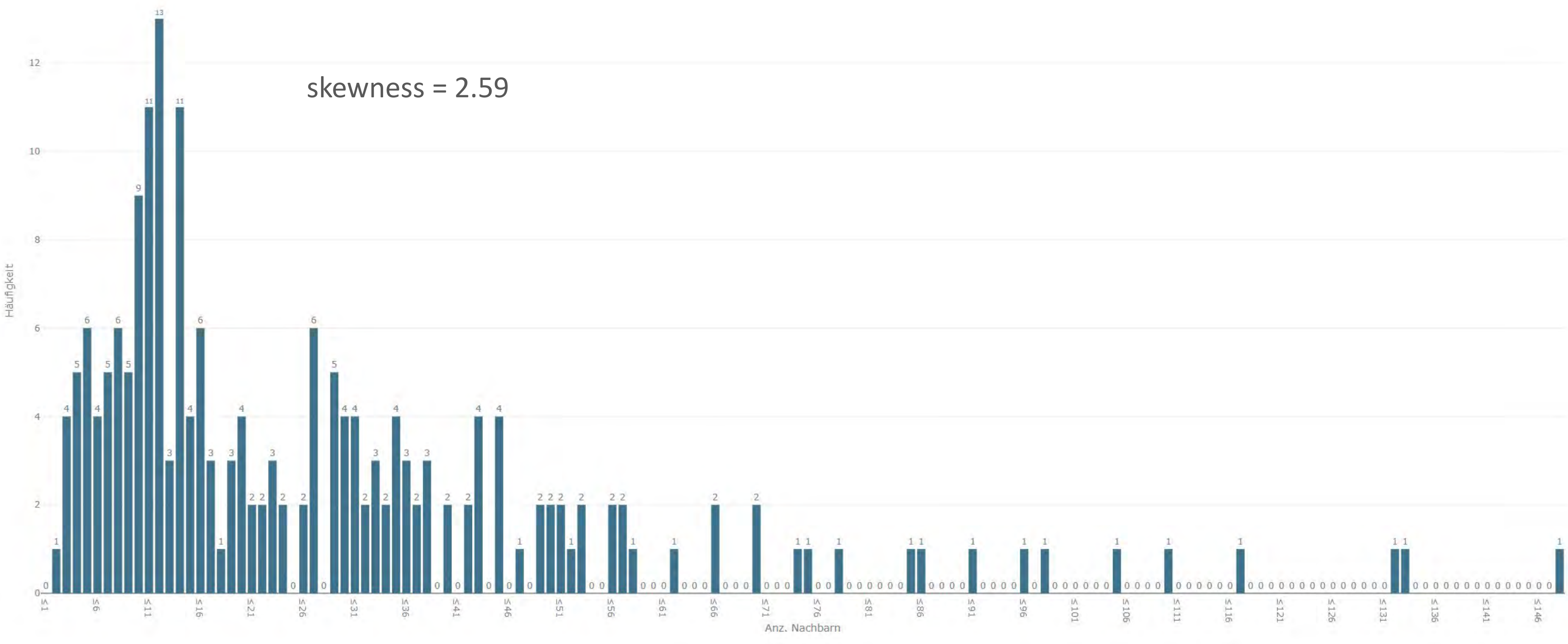
occurs values histogram



Ingredient node neighbors

ingredient node neighbours	#nodes adjacent to selected node
	computed from ingredient graph
	typically: more high than low values; right-skewed
	gives hint to which ingredients are combined with a selected ingredient

Node neighbours histogram

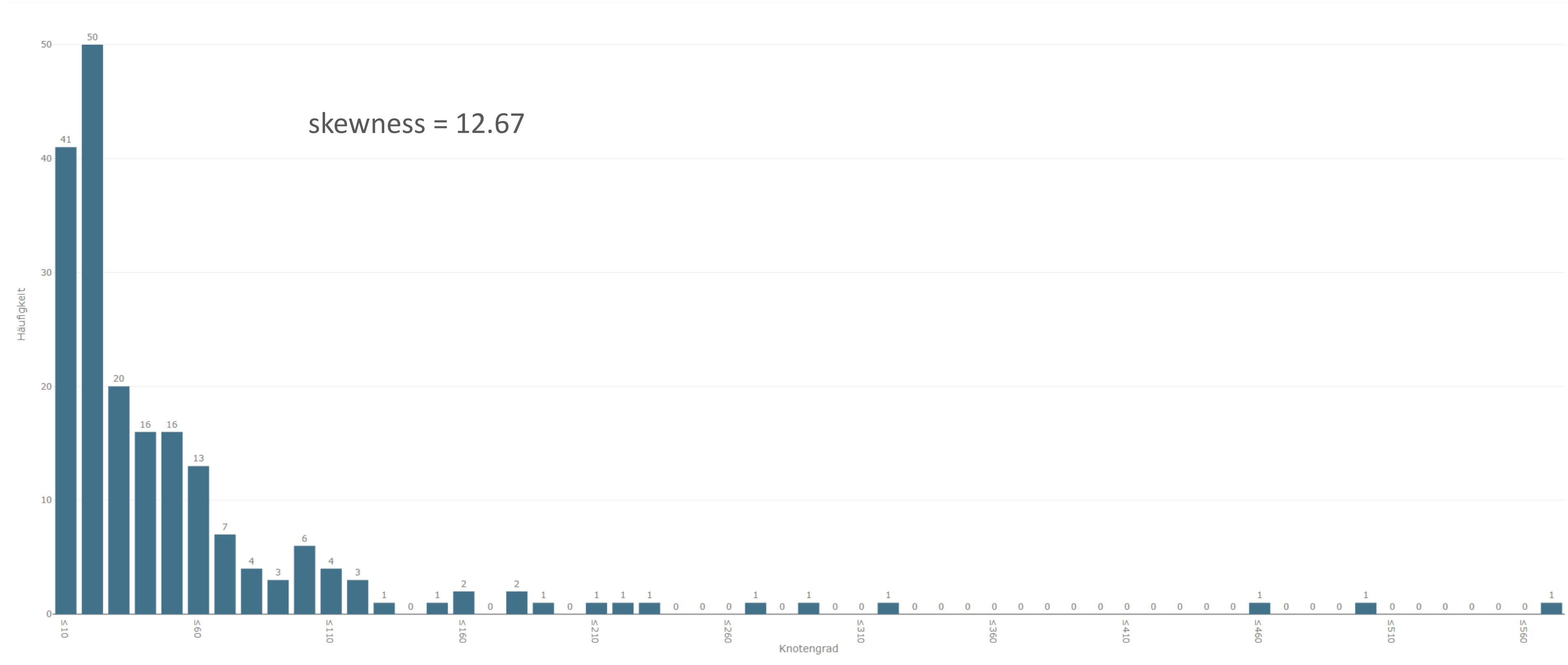


fruschtique
by Norbert Luttenberger

Ingredient node degree

ingredient node degree	The degree of a vertex v of G, denoted by $d(v)$ or $deg(v)$, is the number of edges incident to v.
	computed from ingredient graph
	typically: more high than low values; highly right-skewed
	gives hint to which ingredients are strongly/weakly related to others

Node degree histogram

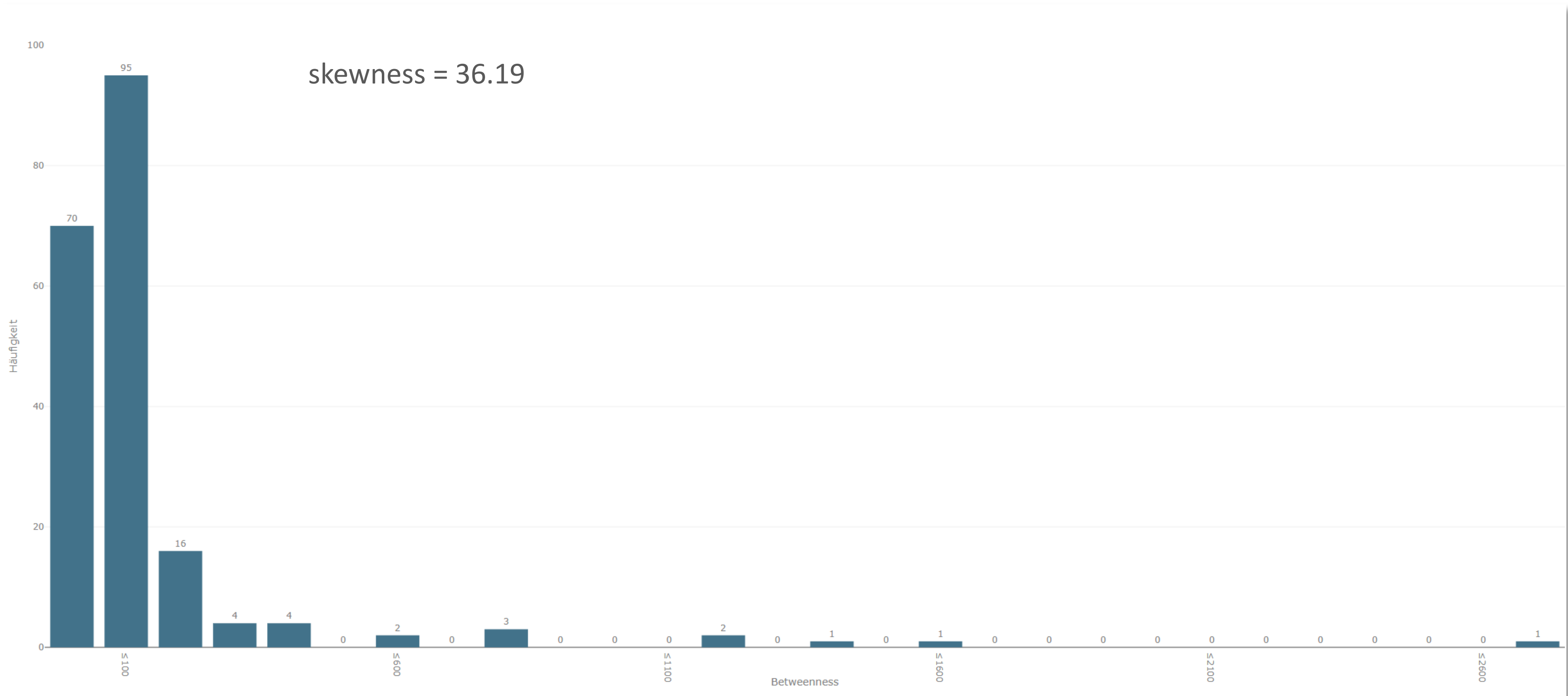


fruschtique
by Norbert Luttenberger

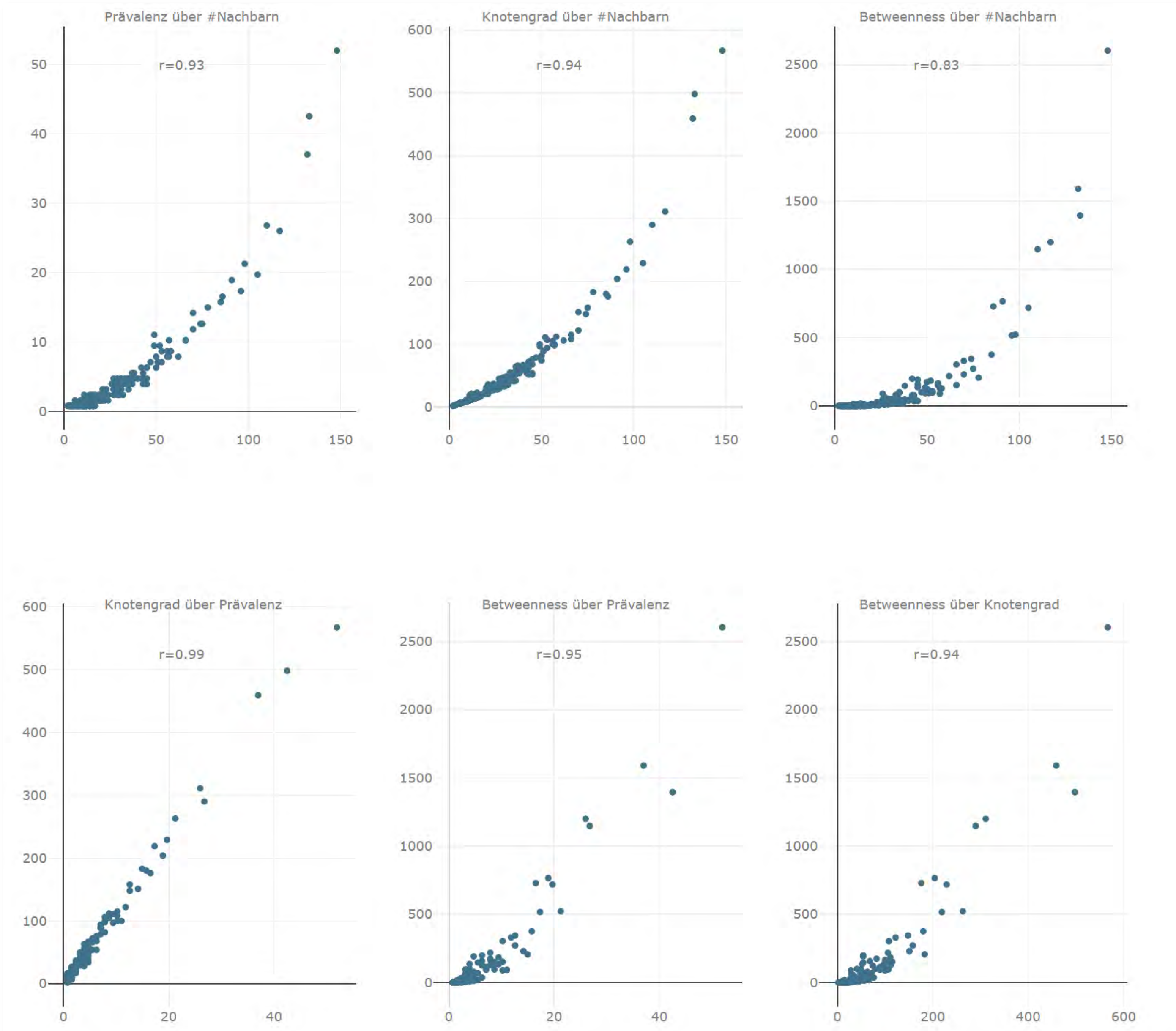
Ingredient node betweenness

ingredient node between-ness	Betweenness Centrality for a vertex $v \in V$ in a connected graph $G = (V, E)$ is given by $\sum_{s,t \in V \wedge s \neq v \wedge t \neq v} n_{s,t}^{v} / n_{s,t}$, where $n_{s,t}$ is the number of shortest paths from s to t and $n_{s,t}^{v}$ is the number of shortest paths from s to t passing through v.
	computed from ingredient graph
	typically: more high than low values; highly right-skewed
	gives hint to which ingredients may serve as "intermediators" between others

Node betweenness histogram, sample



Mutual correlations



fruschtique
by Norbert Luttenberger

Mind the high correlation coefficient values!

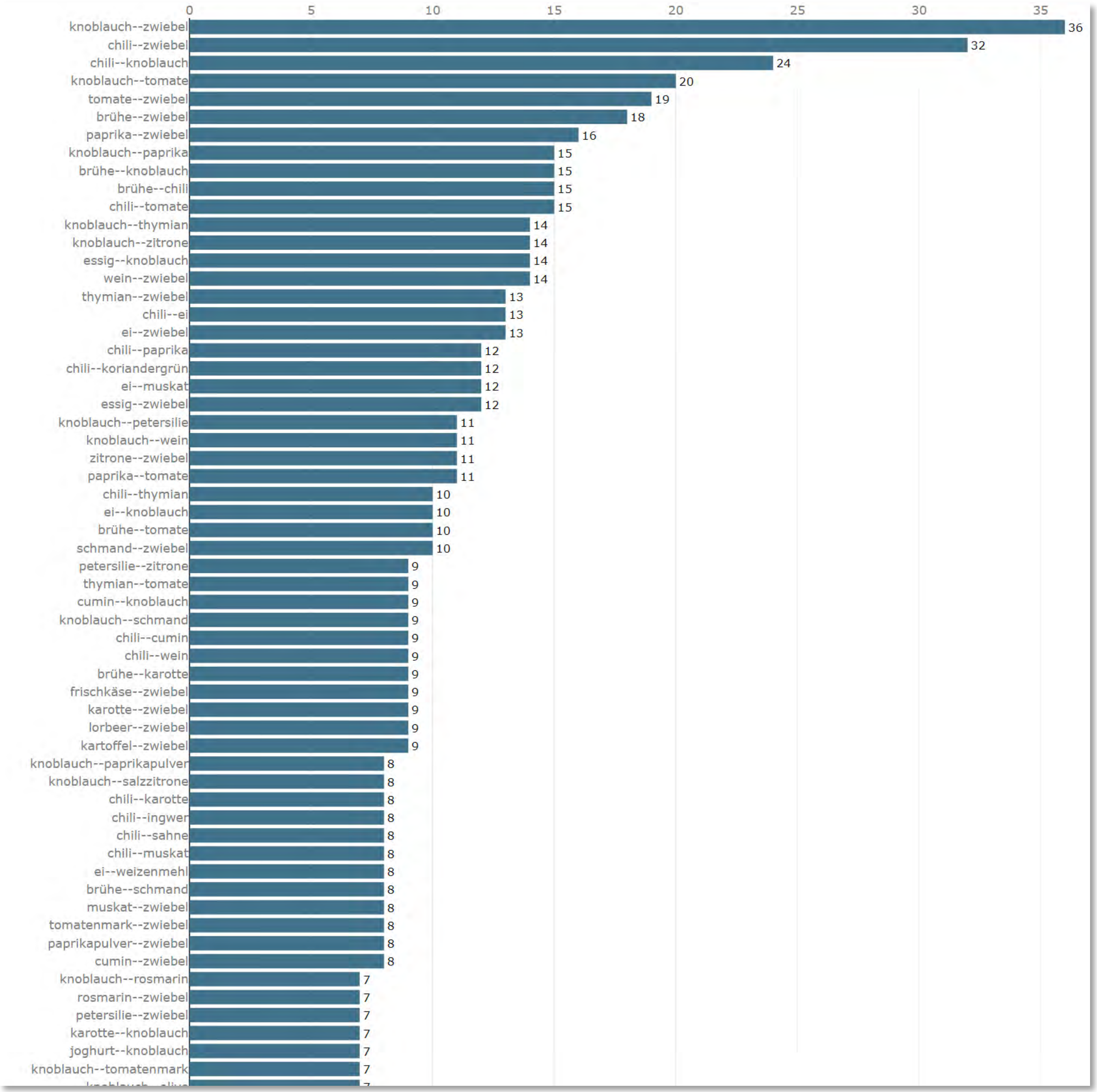
for above sample:

prevalence	–	#node neighbours	0.93
node degree	–	#node neighbours	0.94
betweenness	–	#node neighbours	0.83
node degree	–	prevalence	0.99
betweenness	–	prevalence	0.95
betweenness	–	node degree	0.94

Ingredient relations

ingredient relations	Ingredient graph edges including their weight values
	computed from ingredient graph
	typically: more high than low values; "long tail"
	gives hint how ingredients are related to one another

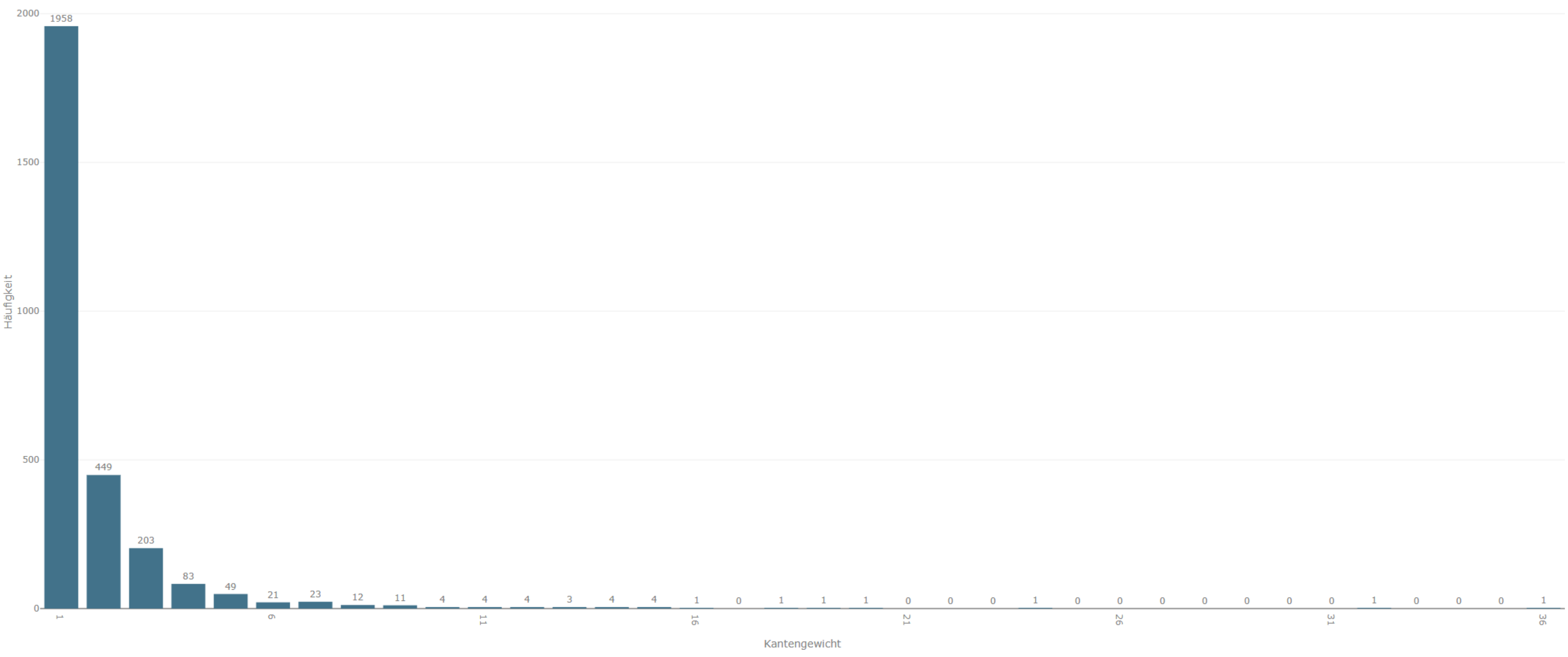
Ingredient relations bar chart
(highest values)



fruschtique
by Norbert Luttenberger

Ingredient relations bubble chart (highest values)

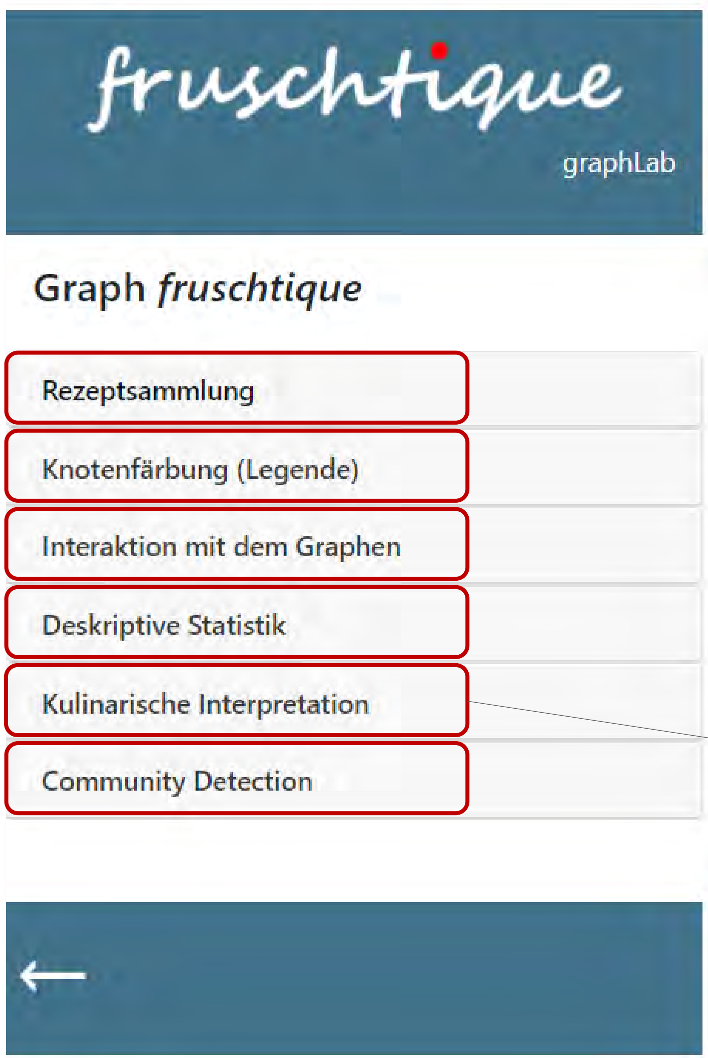
Ingredient relation weight histogram



fruschtique
by Norbert Luttenberger


Culinary interpretation

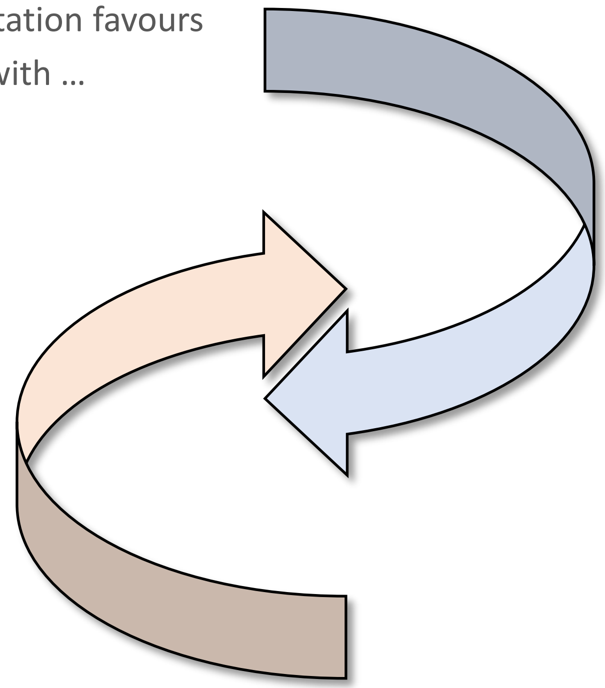
Viewer page



culinary interpretation of the ingredient graph

Caveat

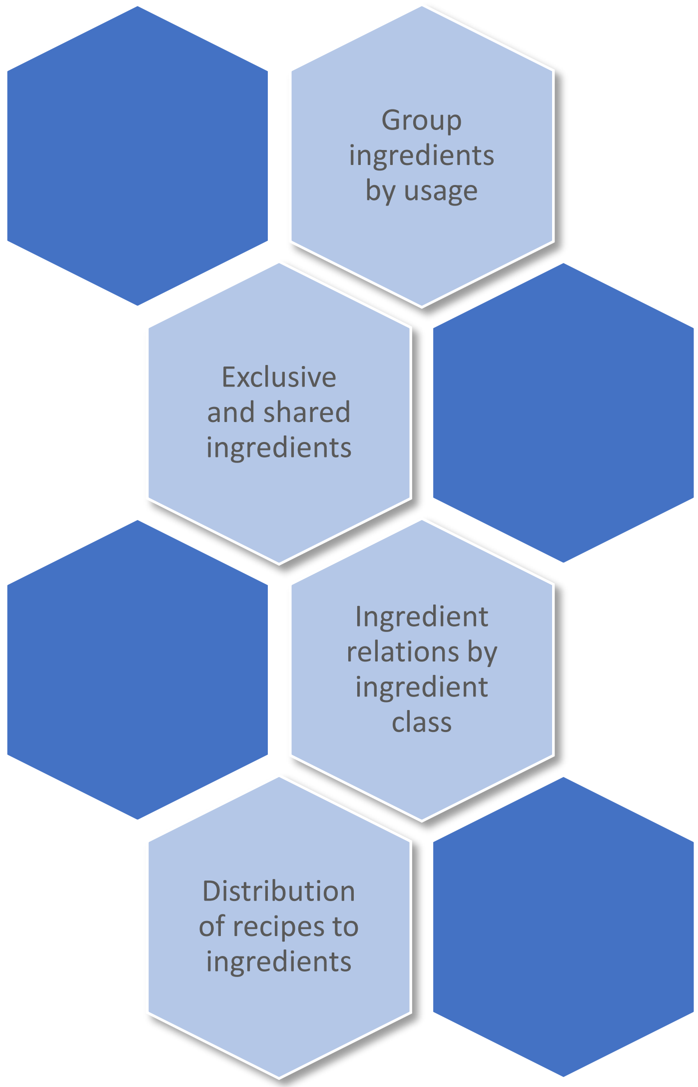
My culinary interpretation favours culinary collections with …



diversity in ingredients and ingredient combinations

diversity in seasoning

Making it visible



Making it visible

- Group ingredients by usage
- Exclusive and shared ingredients
- Ingredient relations by ingredient class
- Distribution of recipes to ingredients

"Degree of usage" groups

big five

- Highest five `occurs` values
- Ingredients needed for nearly all recipes

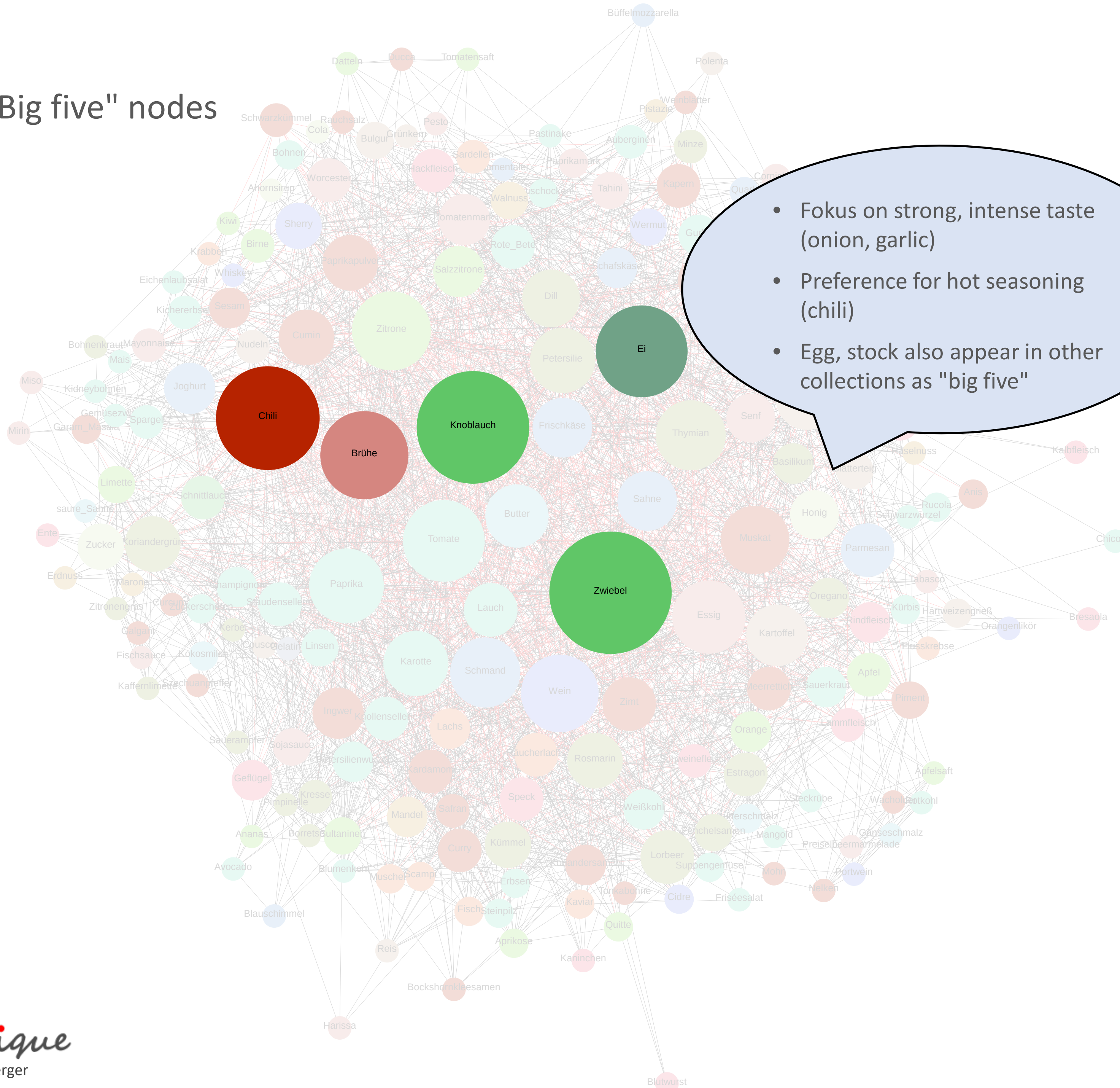
middle-earth

- $1 < \texttt{occurs} < \texttt{occurs}_{\texttt{big five}}$
- "Bread and butter" ingredients

zero-nodes

- `Occurs` value = 1
- Ingredients needed to make recipes unique

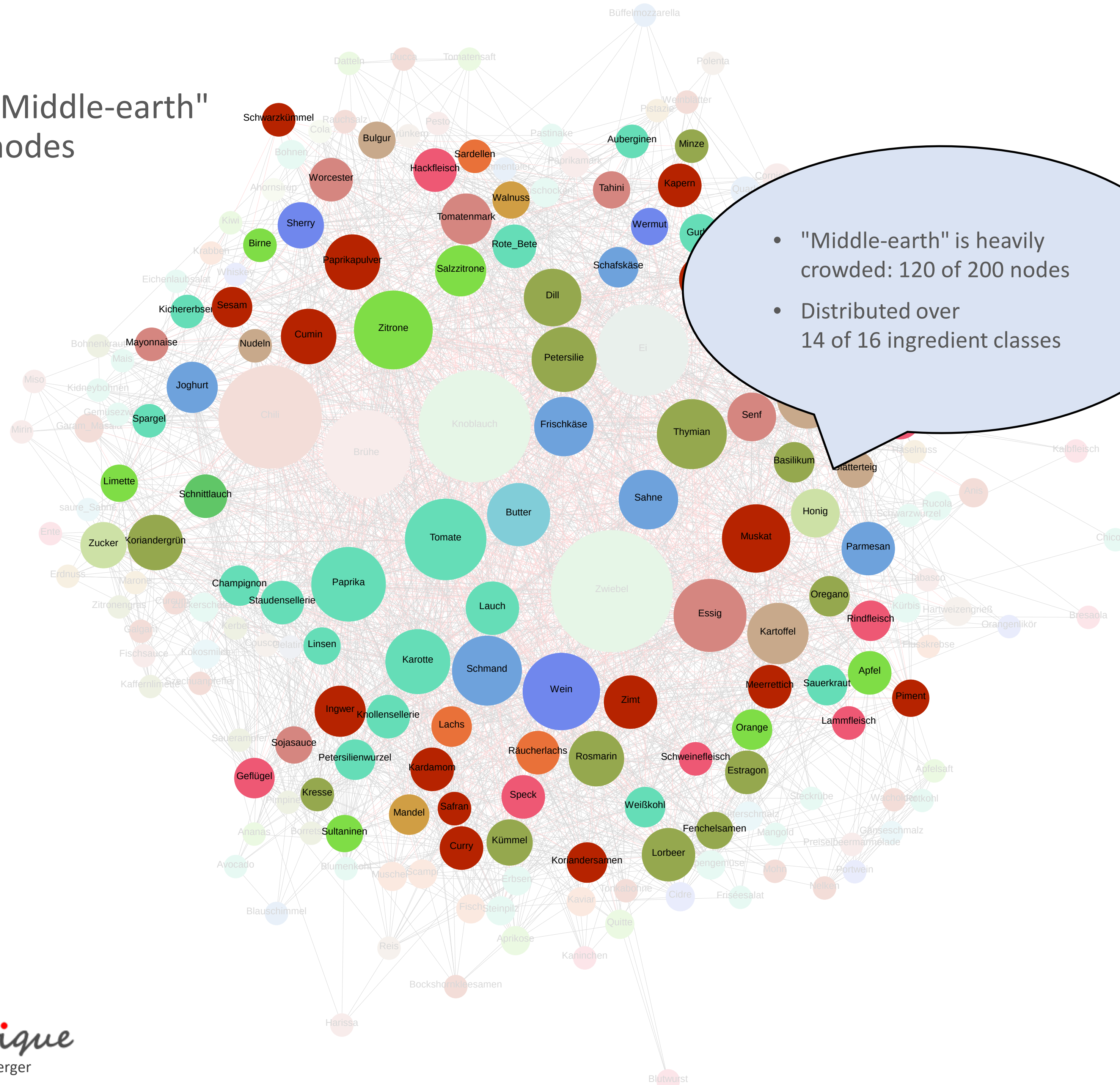
"Big five" nodes



- Fokus on strong, intense taste (onion, garlic)
- Preference for hot seasoning (chili)
- Egg, stock also appear in other collections as "big five"

fruschtique
by Norbert Luttenberger

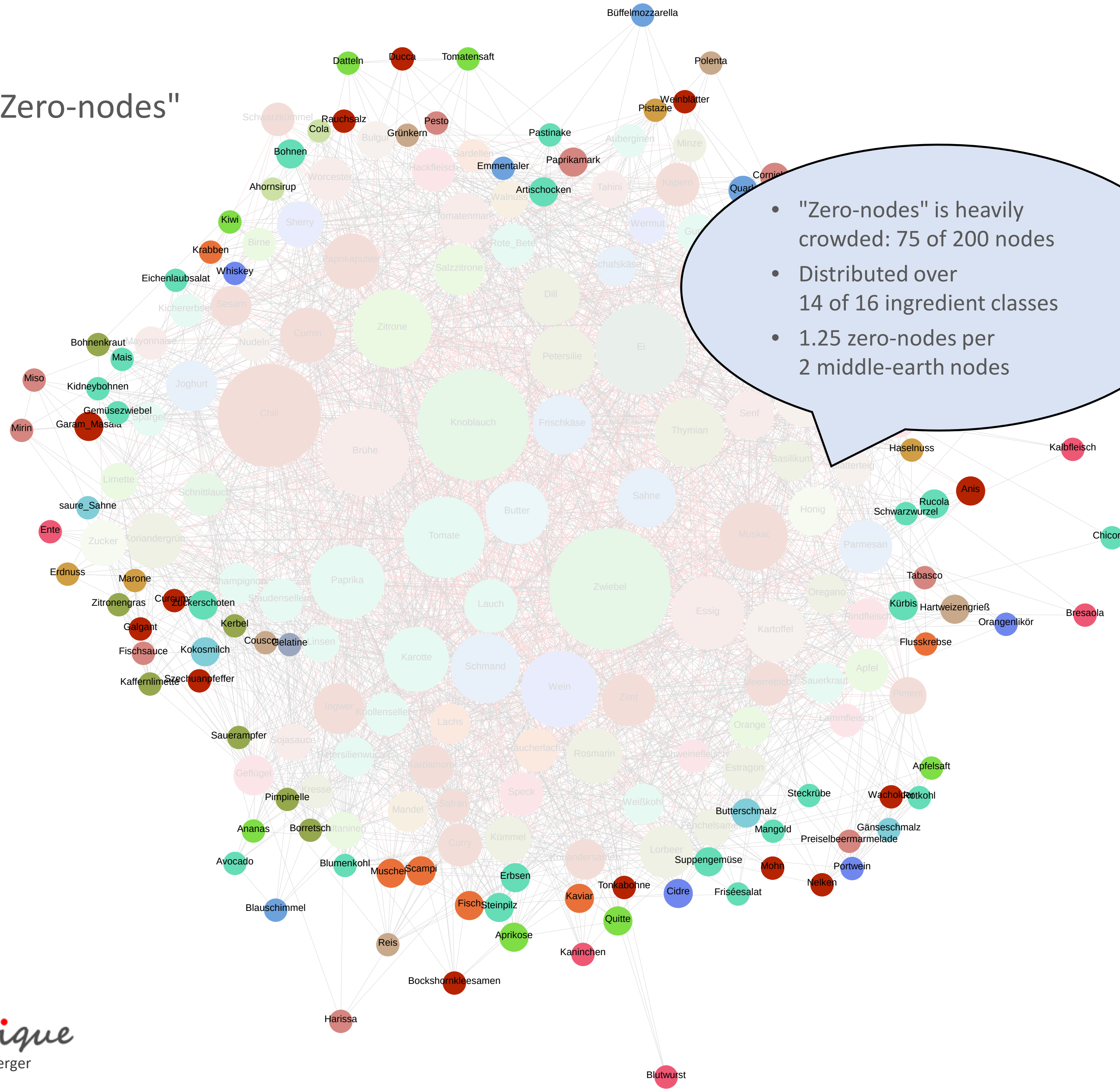
"Middle-earth" nodes



- "Middle-earth" is heavily crowded: 120 of 200 nodes
- Distributed over 14 of 16 ingredient classes

fruschtique
by Norbert Luttenberger

"Zero-nodes"



- "Zero-nodes" is heavily crowded: 75 of 200 nodes
- Distributed over 14 of 16 ingredient classes
- 1.25 zero-nodes per 2 middle-earth nodes

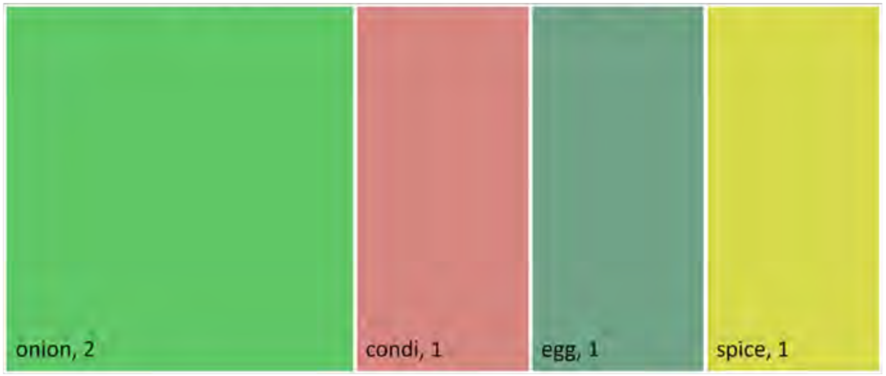
fruschtique
by Norbert Luttenberger

Ingredient classes vs. degree of usage

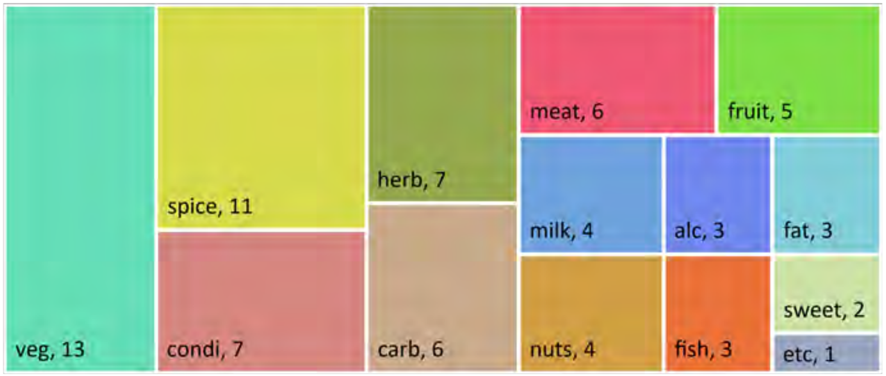
Ingredient Class Population total

Ingredient Class Population "Middle-earth"

Ingredient Class Population "Big Five"



Ingredient Class Population "Zero-nodes"



fruschtique
by Norbert Luttenberger

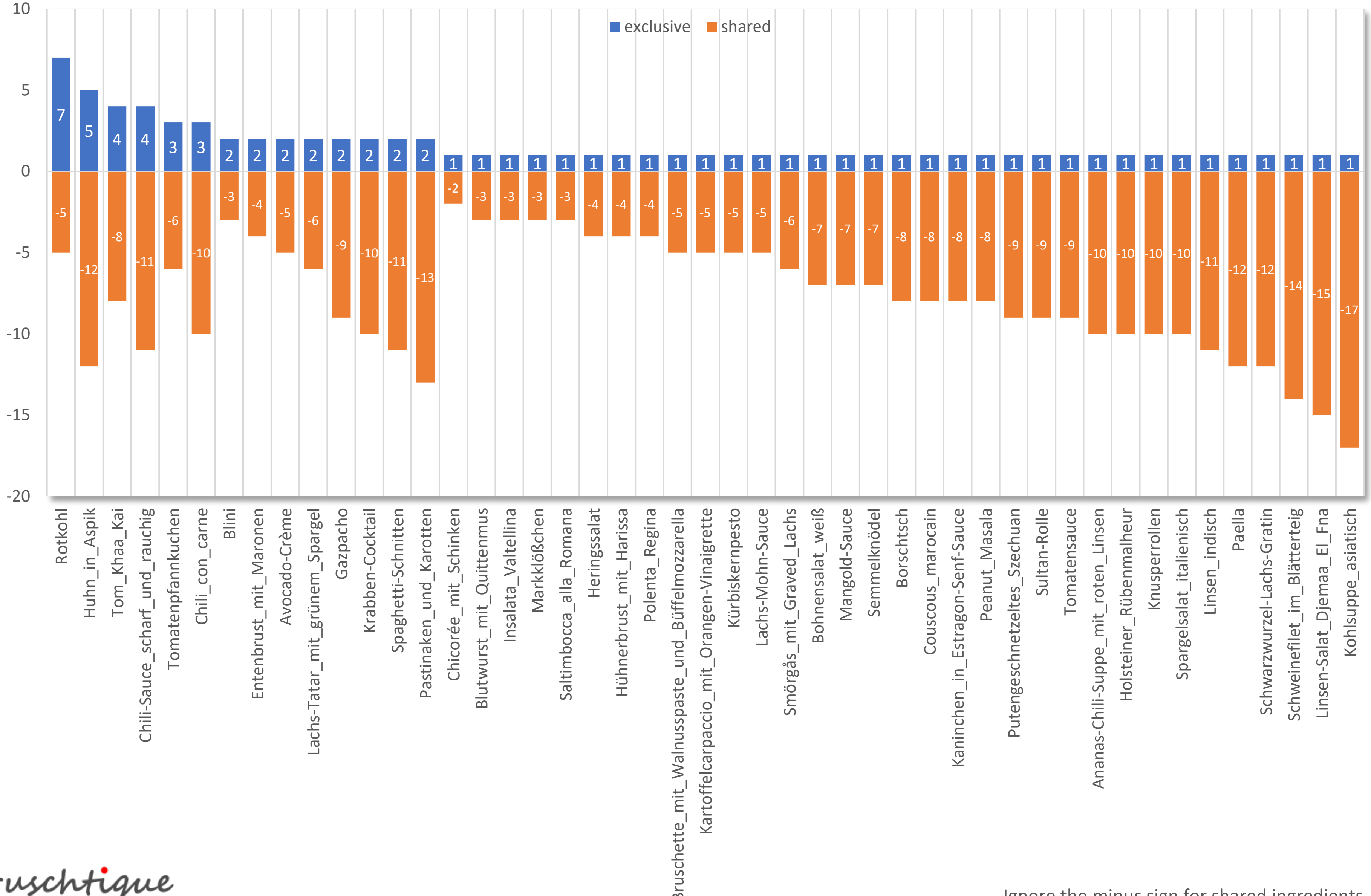
Making it visible

- Group ingredients by usage
- Exclusive and shared ingredients
- Ingredient relations by ingredient class
- Distribution of recipes to ingredients

Exclusive vs. shared ingredients

- If an ingredient is used in only one recipe, it is called an exclusive ingredient.
  Exclusive ingredients have an occurrence value = 1.
- If an ingredient is used in more than one recipe, it is called a shared ingredient.
  Shared ingredients have an occurrence value > 1.
- Favor recipes that use a high number of exclusive ingredients
  and a low number of shared ingredients.

Exclusive vs. shared ingredients



fruschtique
by Norbert Luttenberger

Ignore the minus sign for shared ingredients.

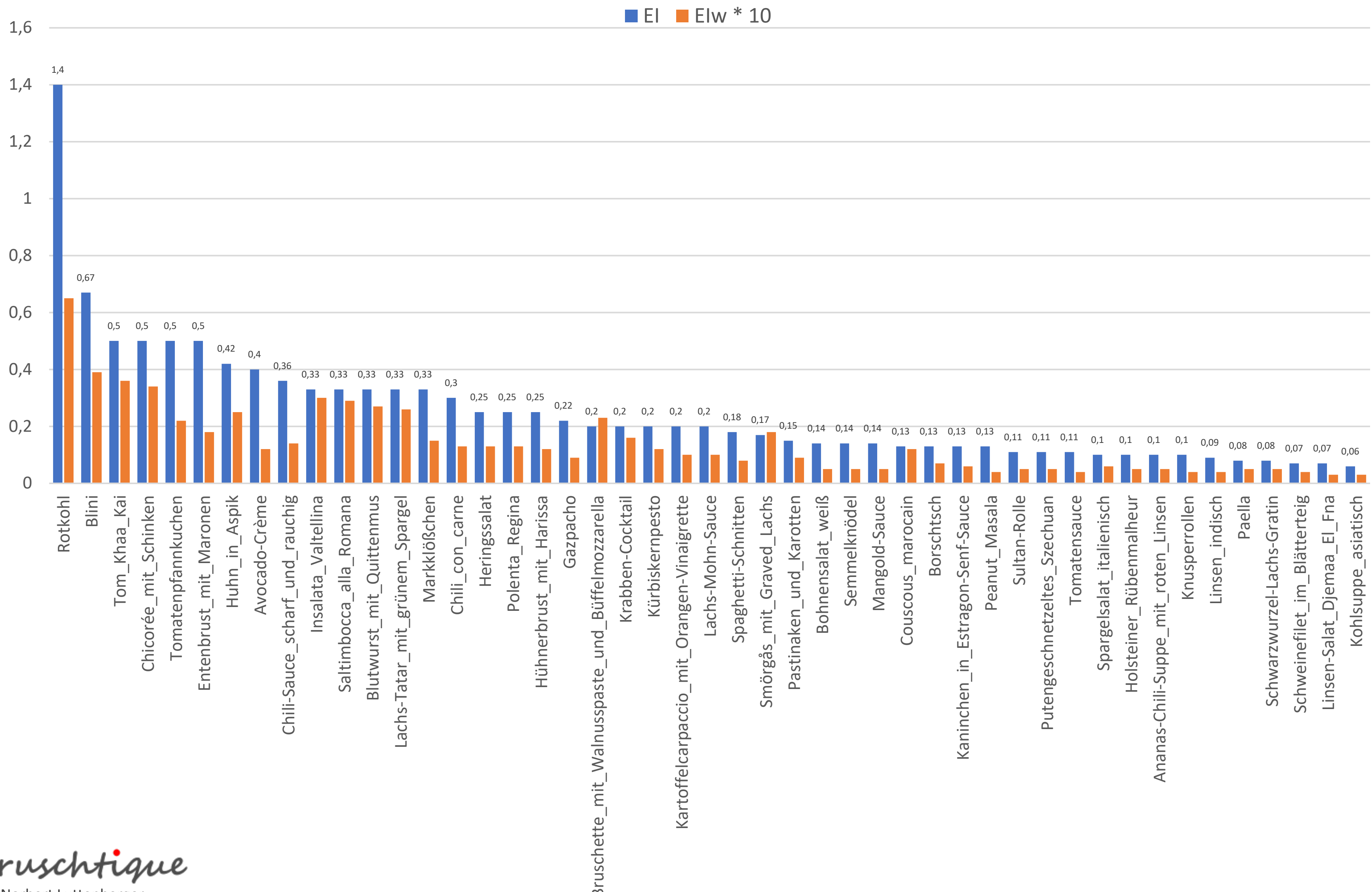
Exclusive vs. shared ingredients again

- For numerical comparison, compute a so-called EI number per recipe.
- The EI number comes in two variants:

EI number of exclusive ingredients in recipe
divided by number of shared ingredients in recipe

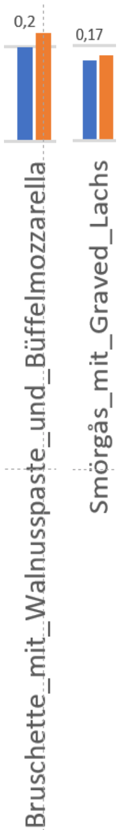
EI_w weighted EI number:
number of exclusive ingredients in recipe
divided by sum of shared ingredient occurrence values in recipe

EI and EI_w



fruschtique
by Norbert Luttenberger

Two findings



$EI_w * 10 > EI$

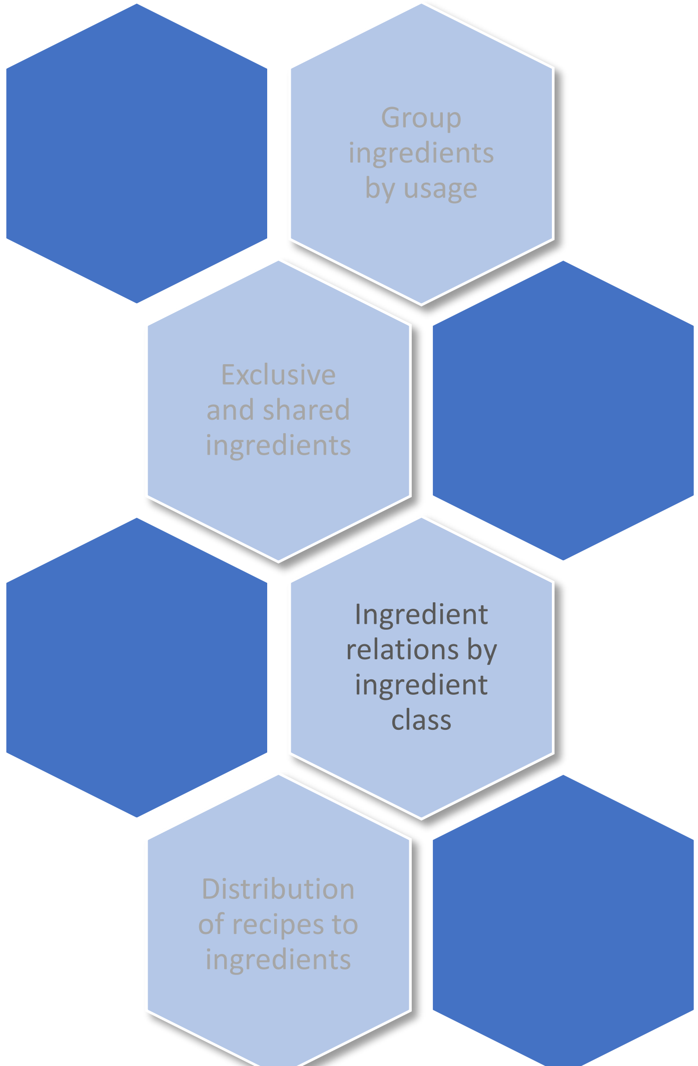
Bruschette	occ
brot	14
walnuss	4
kapern	6
petersilie	16
sardelle	3
büffelmozzarella	1

Smörgås	occ
brot	14
butter	15
dill	12
honig	8
kaviar	2
lachs	5
friséesalat	1

EI_w profits from low occurence values of shared ingredients!

fruschtique
by Norbert Luttenberger

Making it visible



Combinatorics interlude

To describe a an ingredient relation by its source and target ingredient classes, we need

$k = 2$ picks from
$n = 16$ ingredient classes.

That yields $\binom{n+k-1}{k}$ possible combinations assuming we order the picks.

Thus, with the above given values for n and k, we have
$\binom{17}{2} = 136$ possible combinations of source and target classes.

Aside:

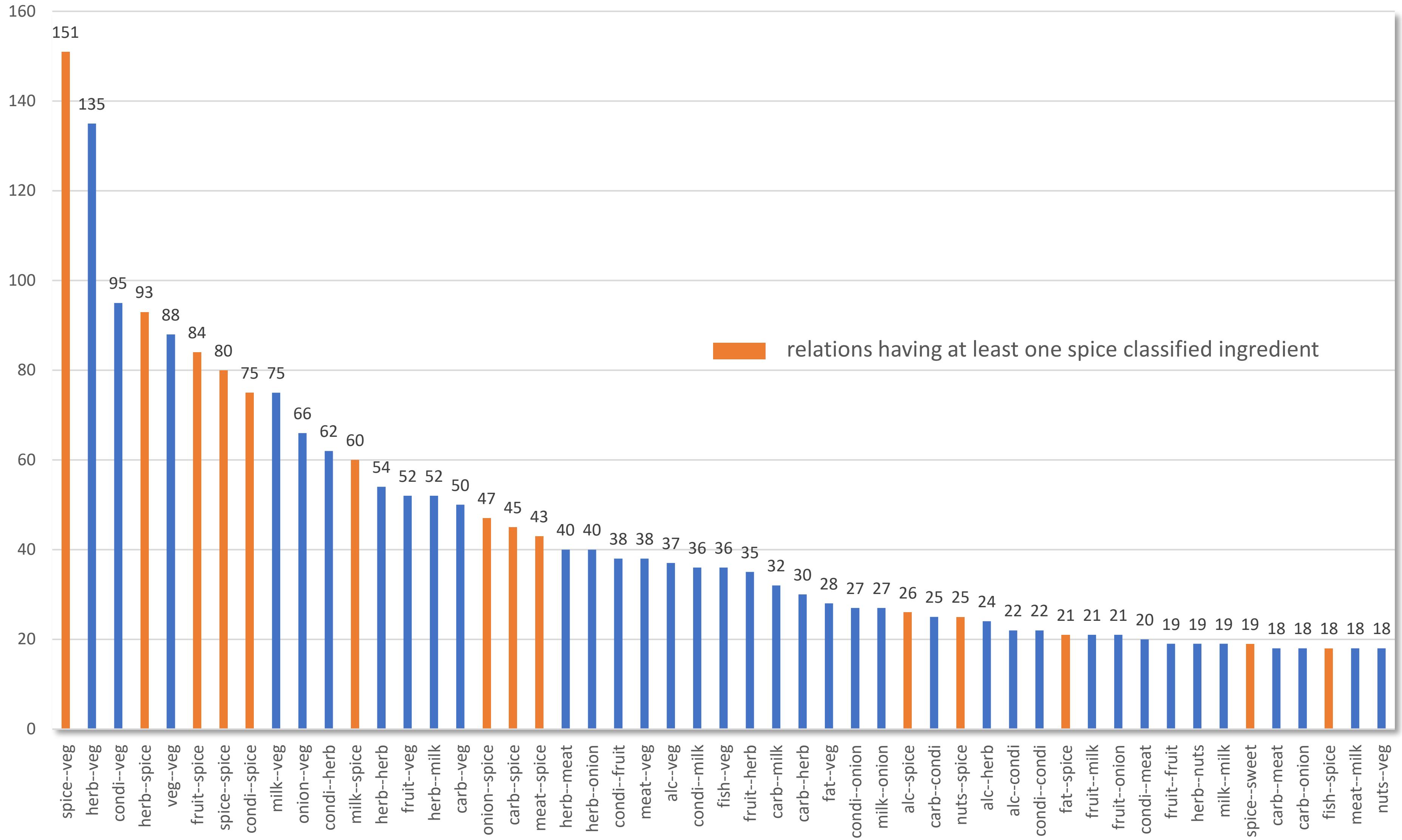
For our sample collection, to describe an ingredient relation by its source and target ingredients, we need

$k = 2$ picks from
$n = 200$ ingredients.

That yields $\binom{n+k-1}{k}$ possible combinations assuming we order the picks.

Thus, with the above given values for n and k, we have
$\binom{201}{2} = 20100$ possible combinations of source and target ingredients.

Ingredient relations by ingredient classes



highest values only

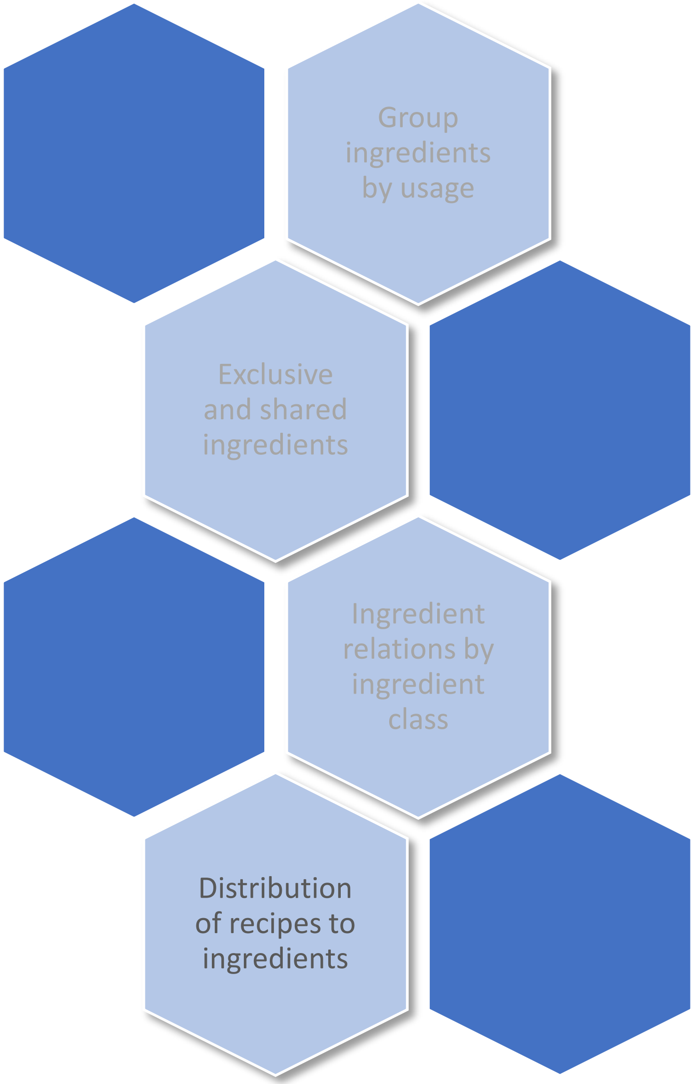
fruschtique
by Norbert Luttenberger

Missing ingredient class relations

- Total of 126 relations; potentially 136 relations
- Unpopulated relations:

carb--etc
egg--egg
etc--etc
etc--fat
etc--fish
etc--nuts
etc--fruit
etc--spice
etc--sweet
nuts--nuts

Making it visible

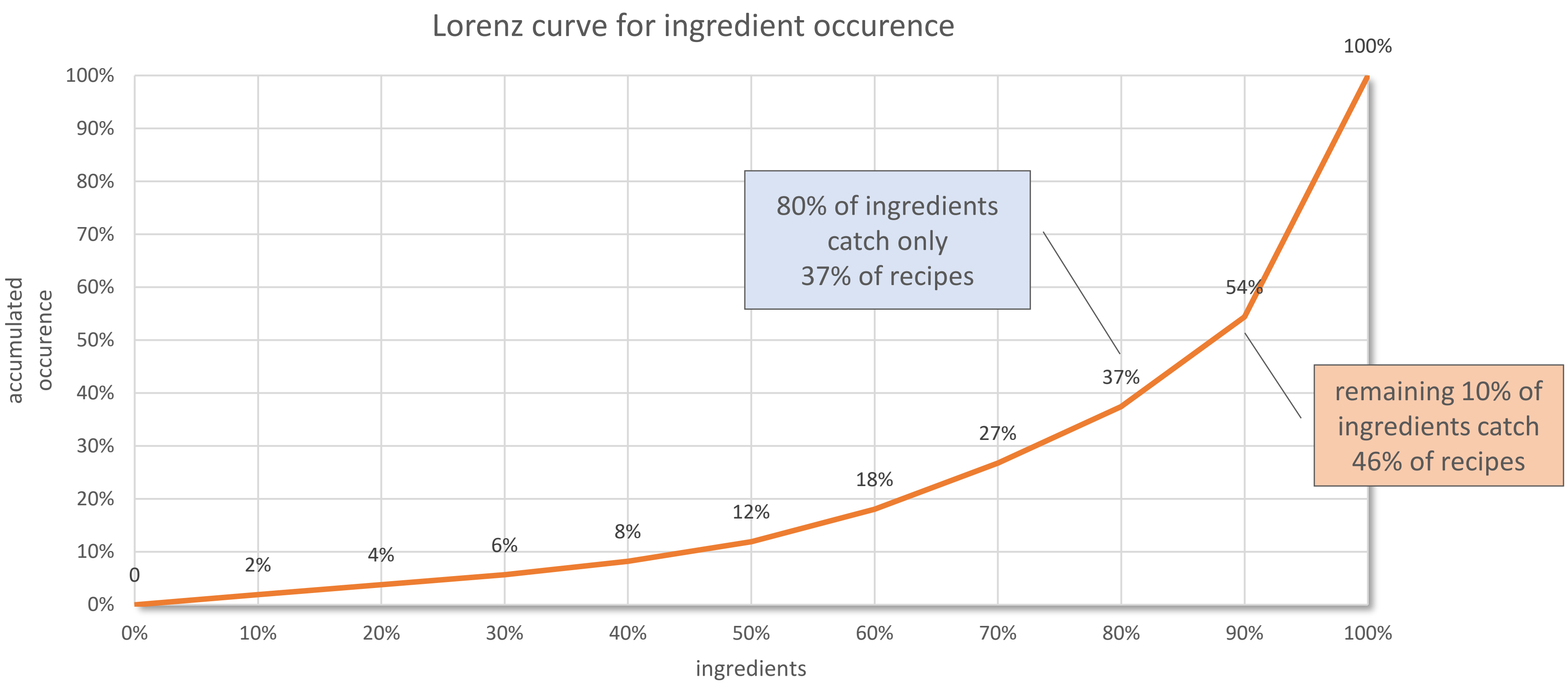


Idea

from: https://ulis-culinaria.de/kulinarische-redewendungen/

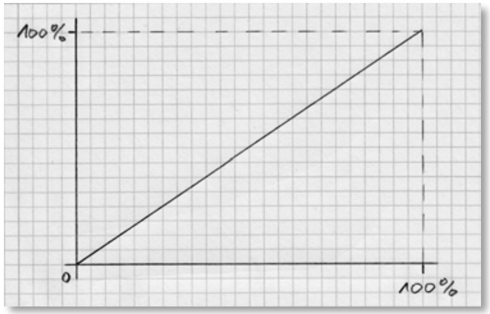
Ingredients catch recipes!

Lorenz curve and Gini coefficients

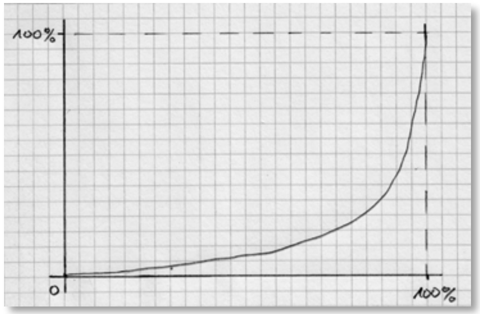


fruschtique
by Norbert Luttenberger

What the Lorenz curve can tell

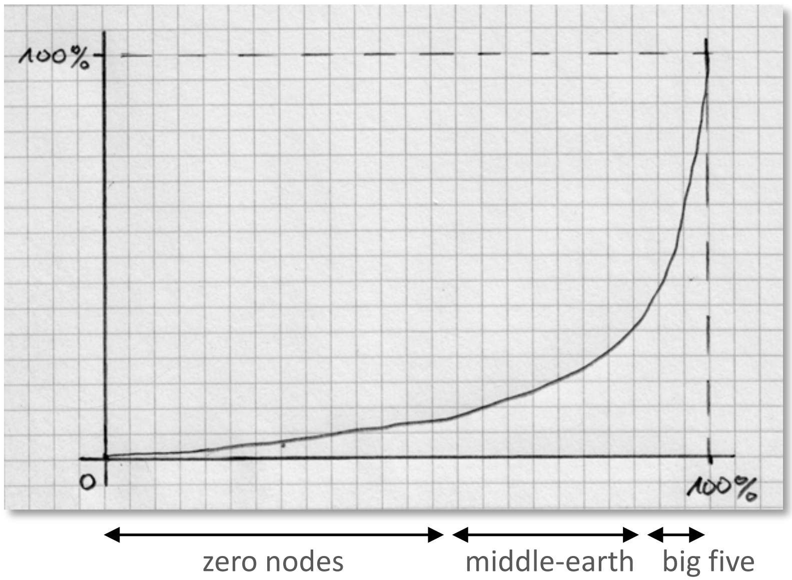


- perfect equality
- Gini coefficient = 0



- approaching perfect inequality
- Gini coefficient ≲ 1, i.e. less than, but close to 1

Lorenz curve for ingredient graphs



When comparing culinary collections:

→ Larger Gini coefficient
signals more diverse collection.

- zero nodes — curve gradient = 1
  may completely disappear in favor of middle-earth
- middle-earth — curve gradient > 1
  may completely disappear in favor of zero nodes
- big five — curve gradient ≫ 1
  always present (unless we assume a collection with less then 5 different ingredients)


Thank you!