Towards Personalized Diet Using Linked Data

By Aleksandra Kovachev

AI & Nutrition - AMLD 2020, Lausanne

Senior DS, DeliveryHero SE, Berlin, Germany
T&RA, Faculty of Computer Science and Engineering, Skopje, North Macedonia
aleksandra.kovachev@deliveryhero.com

Motivation

We are what we eat!

Let food be thy medicine and medicine be thy food - Hippocrates

IBS, Lactose Intolerance, Fructose Malabsorption, Celiac Desease

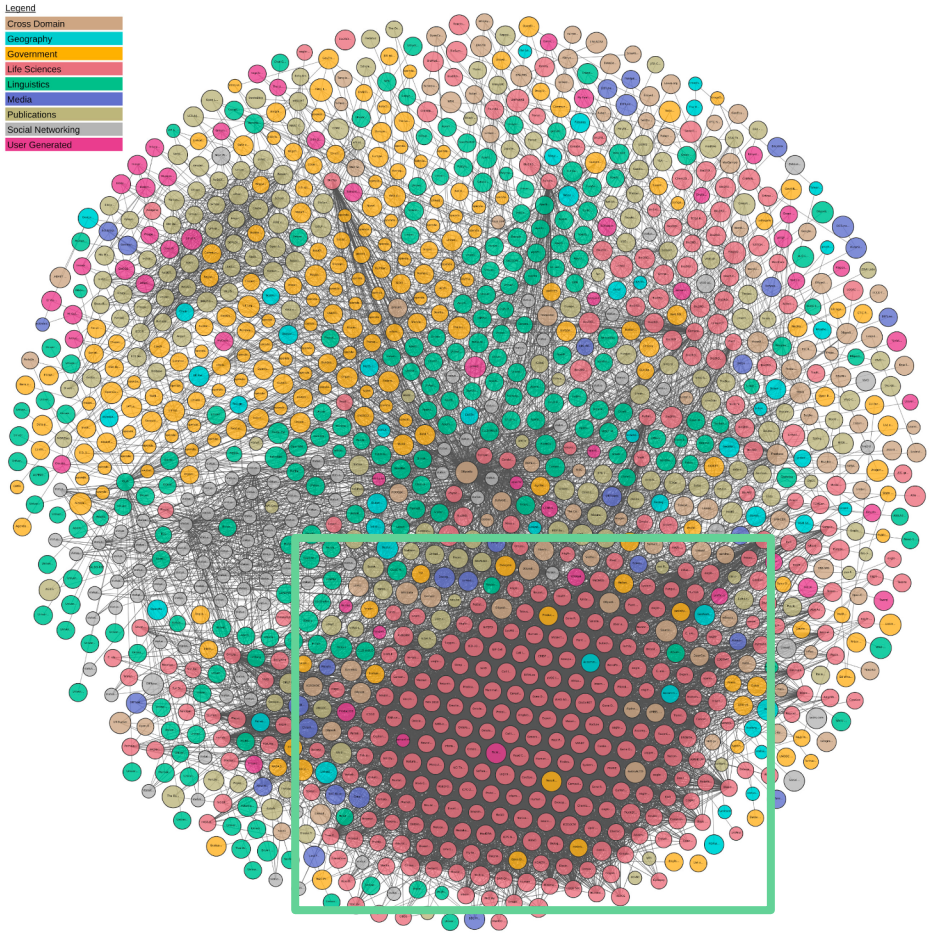
How Was The Idea Born?

The Linking Open Data (LOD) Project

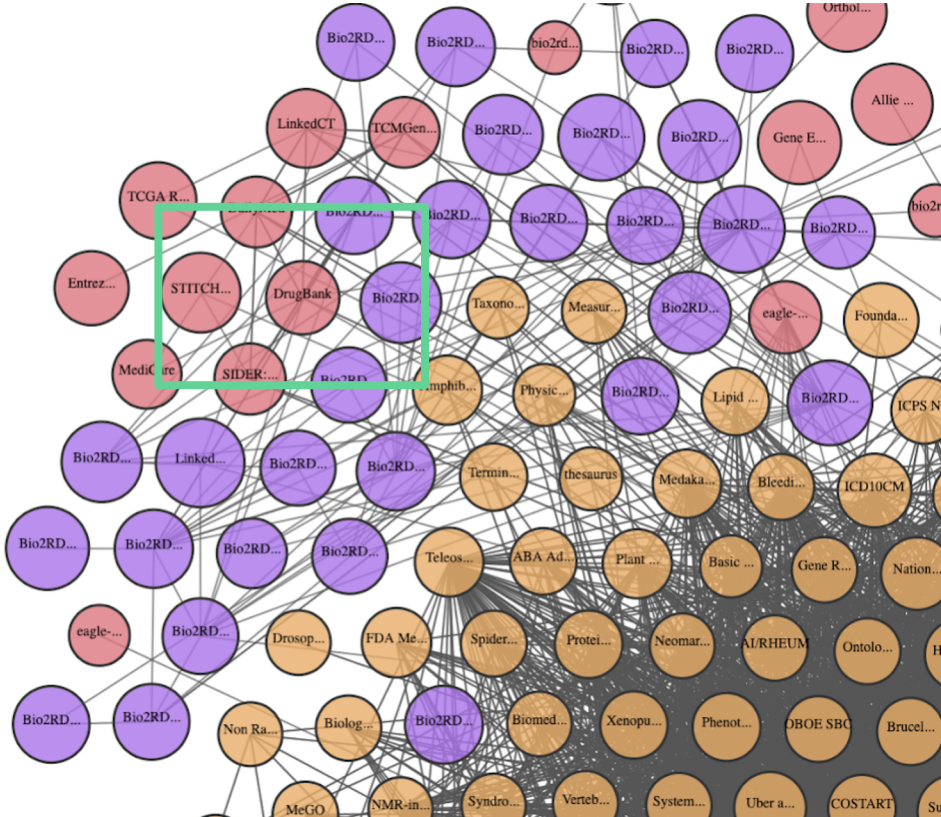
- The LOD project is a community activity started in 2007 by W3C.
- The project's stated goal is to "make data freely available to everyone."
- More info about the LOD cloud is available at http://lod-cloud.net/.

Open Linked Data Cloud

1,239 datasets
with **16,147** links in 2019

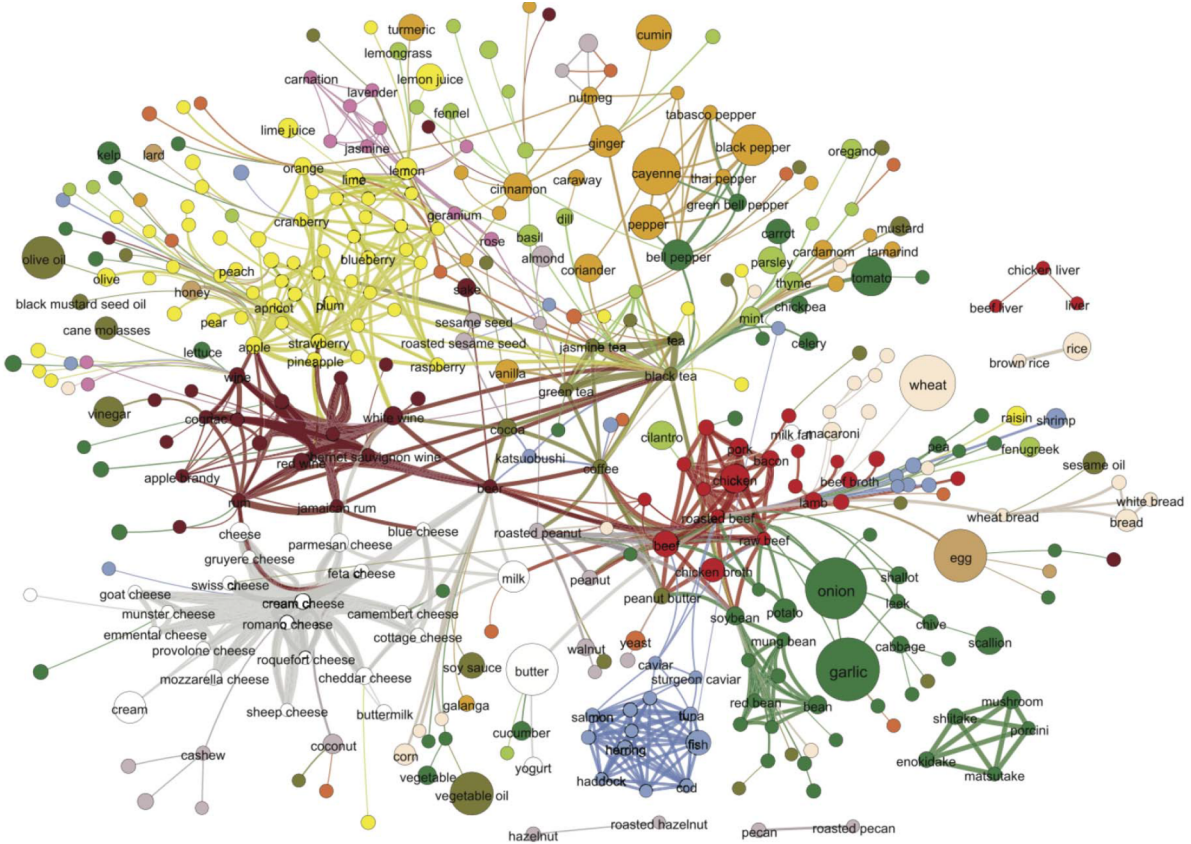


DrugBank Dataset



The Flavour Network

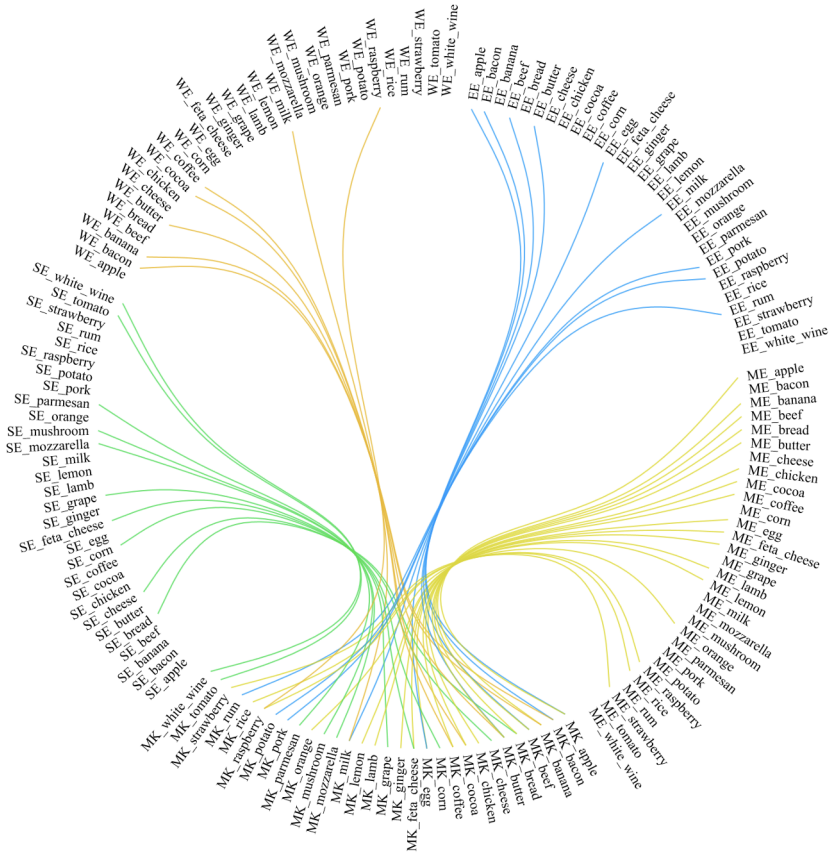
Ahn Y-Y, Ahnert SE, Bagrow JP, Barabási A-L (2011) Flavor network and the principles of food pairing. Scientific Reports 1



Processing and Analysis of Macedonian Cuisine and its Flavours by Using Online Recipes

A Bogojeska, S Kalajdziski, L Kocarev

International Conference on Innovations, 143-152
https://link.springer.com/chapter/10.1007/978-3-319-25733-4_15



Specificity and the Inheritance of the Positive Contributing Ingredients in the Macedonian Cuisine

Some Questions We Asked

- What recipes or cuisines should one avoid if on prescribed drugs?
- Which cuisines should be preferred when a particular disease is present?

  . . .
- What foods should I eat to stop hair loss or hair discolouring?
- How we can increase longevity and delay ageing using food?

Primary Focus on Two Aspects

Inferring Cuisine - Drug Interactions Using the Linked Data Approach

http://go.nature.com/rnPO91

- Negative interactions between drugs from a given category and recipes from a given cuisine
- Ingredients' impact on the negative food - drug interactions in different parts of the world

Let Me Ask You

- Do you read the drug guide when you are prescribed with new medication?
- When should you avoid alcohol, milk or grapefruit while taking this medication?
- What about your children medications? Had your pediatrician given you this information?
- What if you take more than one drug, and taking drugs from different categories?

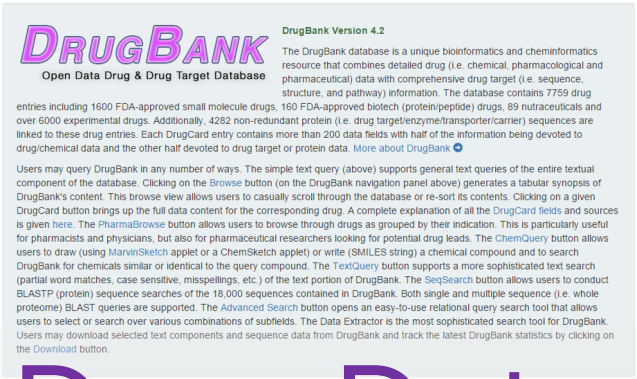
Why is This Important

- 70% of the American population consumes at least 1 prescription drug
- 20% of them are on 5 or more drugs
- 1.5 million people are harmed by medications including errors due to lack of information provided or not reading the patient drug information

Why is This Important

- Food can change the bioavailability of a drug and modify its clinical effect
- We are consuming non-native cuisines on daily basis

How Was The Analysis Performed?



DRUGBANK

Open Data Drug & Drug Target Database

DrugBank Version 4.2

The DrugBank database is a unique bioinformatics and cheminformatics resource that combines detailed drug (i.e. chemical, pharmacological and pharmaceutical) data with comprehensive drug target (i.e. sequence, structure, and pathway) information. The database contains 7759 drug entries including 1600 FDA-approved small molecule drugs, 160 FDA-approved biotech (protein/peptide) drugs, 89 nutraceuticals and over 6000 experimental drugs. Additionally, 4282 non-redundant protein (i.e. drug target/enzyme/transporter/carrier) sequences are linked to these drug entries. Each DrugCard entry contains more than 200 data fields with half of the information being devoted to drug/chemical data and the other half devoted to drug target or protein data. More about DrugBank

Users may query DrugBank in any number of ways. The simple text query (above) supports general text queries of the entire textual component of the database. Clicking on the Browse button (on the DrugBank navigation panel above) generates a tabular synopsis of DrugBank's content. This browse view allows users to casually scroll through the database or re-sort its contents. Clicking on a given DrugCard button brings up the full data content for the corresponding drug. A complete explanation of all the DrugCard fields and sources is given here. The PharmaBrowse button allows users to browse through drugs as grouped by their indication. This is particularly useful for pharmacists and physicians, but also for pharmaceutical researchers looking for potential drug leads. The ChemQuery button allows users to draw (using MarvinSketch applet or a ChemSketch applet) or write (SMILES string) a chemical compound and to search DrugBank for chemicals similar or identical to the query compound. The TextQuery button supports a more sophisticated text search (partial word matches, case sensitive, misspellings, etc.) of the text portion of DrugBank. The SeqSearch button allows users to conduct BLASTP (protein) sequence searches of the 18,000 sequences contained in DrugBank. Both single and multiple sequence (i.e. whole proteome) BLAST queries are supported. The Advanced Search button opens an easy-to-use relational query search tool that allows users to select or search over various combinations of subfields. The Data Extractor is the most sophisticated search tool for DrugBank. Users may download selected text components and sequence data from DrugBank and track the latest DrugBank statistics by clicking on the Download button.

SCIENTIFIC REPORTS

OPEN

Flavor network and the principles of food pairing

SUBJECT AREAS:
STATISTICAL PHYSICS, THERMODYNAMICS AND NONLINEAR DYNAMICS
APPLIED PHYSICS
SYSTEMS BIOLOGY
STATISTICS

Yong-Yeol Ahn[1,2,3*], Sebastian E. Ahnert[1,4*], James P. Bagrow[1,2] & Albert-László Barabási[1,2]

[1]Center for Complex Network Research, Department of Physics Northeastern University, Boston, MA 02115, [2]Center for Cancer Systems Biology Dana-Farber Cancer Institute, Harvard University, Boston, MA 02115, [3]School of Informatics and Computing Indiana University, Bloomington, IN 47408, [4]Theory of Condensed Matter, Cavendish Laboratory, University of Cambridge, Cambridge CB3 0HE, UK.

Received
18 October 2011

Accepted
24 N[illegible] 2011

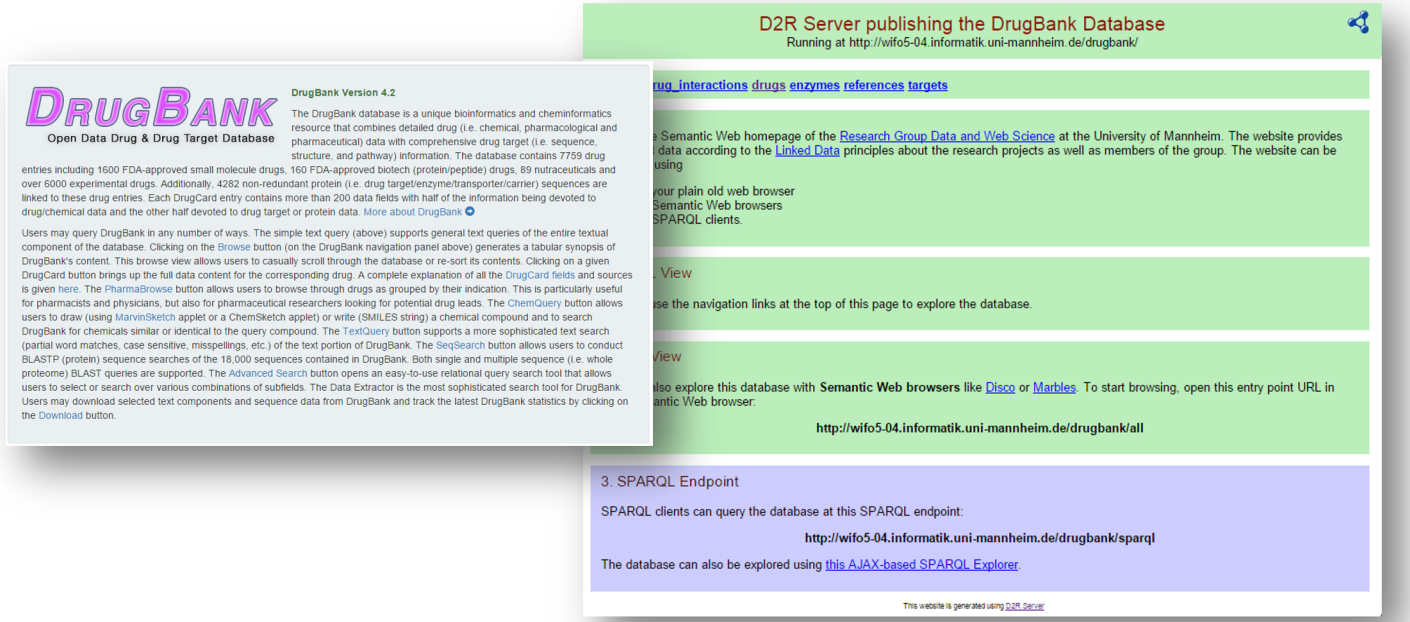
Publish
15 D[illegible]ember 20[illegible]

The cultural diversity of culinary practice, as illustrated by the variety of regional cuisines, raises the question of whether there are any general patterns that determine the ingredient combinations used in food today or principles that transcend individual tastes and recipes. We introduce a flavor network that captures the flavor compounds shared by culinary ingredients. Western cuisines show a tendency to use ingredient pairs that share many flavor compounds, supporting the so-called food pairing hypothesis. By contrast, East Asian cuisines tend to avoid compound sharing ingredients. Given the increasing availability of information on food preparation, our data-driven investigation opens new avenues towards a systematic understanding of culinary practice.

Drug Data

Recipe Data

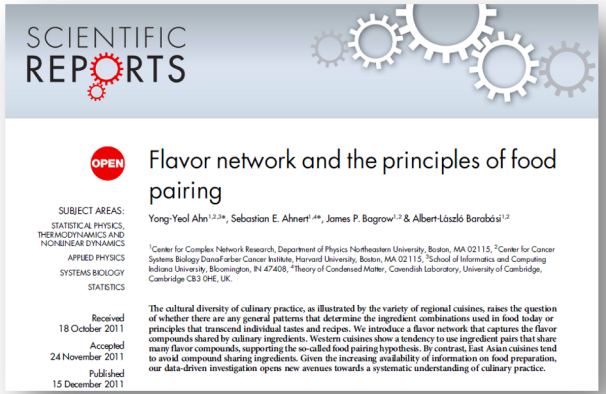
The Drugs Dataset



2 star data ➜ 5 star data

Data and Web Science RG, University of Mannheim

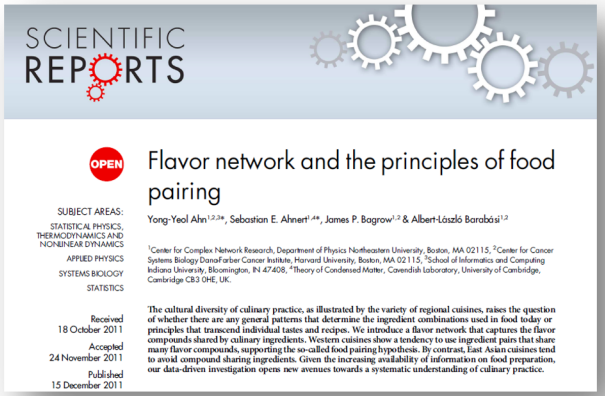
The Recipes Dataset



1 star data → 3 star data

Yong-Yeol Ahn et al., “Flavor Network and the Principles of Food Pairing”

Linking the Datasets



SCIENTIFIC REPORTS

OPEN

Flavor network and the principles of food pairing

SUBJECT AREAS:
STATISTICAL PHYSICS, THERMODYNAMICS AND NONLINEAR DYNAMICS
APPLIED PHYSICS
SYSTEMS BIOLOGY
STATISTICS

Received
18 October 2011
Accepted
24 November 2011
Published
15 December 2011

Yong-Yeol Ahn[1,2,3]*, Sebastian E. Ahnert[1,4]*, James P. Bagrow[1,3] & Albert-László Barabási[1,2]

[1]Center for Complex Network Research, Department of Physics Northeastern University, Boston, MA 02115, [2]Center for Cancer Systems Biology DanaFarber Cancer Institute, Harvard University, Boston, MA 02115, [3]School of Informatics and Computing Indiana University, Bloomington, IN 47408, [4]Theory of Condensed Matter, Cavendish Laboratory, University of Cambridge, Cambridge CB3 0HE, UK.

The cultural diversity of culinary practice, as illustrated by the variety of regional cuisines, raises the question of whether there are any general patterns that determine the ingredient combinations used in food today or principles that transcend individual tastes and recipes. We introduce a flavor network that captures the flavor compounds shared by culinary ingredients. Western cuisines show a tendency to use ingredient pairs that share many flavor compounds, supporting the so-called food pairing hypothesis. By contrast, East Asian cuisines tend to avoid compound sharing ingredients. Given the increasing availability of information on food preparation, our data-driven investigation opens new avenues towards a systematic understanding of culinary practice.

D2R Server publishing the DrugBank Database
Running at http://wifo5-04.informatik.uni-mannheim.de/drugbank/

Home | drug_interactions drugs enzymes references targets

This is the Semantic Web homepage of the Research Group Data and Web Science at the University of Mannheim. The website provides structured data according to the Linked Data principles about the research projects as well as members of the group. The website can be accessed using

1. your plain old web browser
2. Semantic Web browsers
3. SPARQL clients.

1. HTML View

You can use the navigation links at the top of this page to explore the database.

2. RDF View

You can also explore this database with **Semantic Web browsers** like Disco or Marbles. To start browsing, open this entry point URL in your Semantic Web browser:

http://wifo5-04.informatik.uni-mannheim.de/drugbank/all

3. SPARQL Endpoint

SPARQL clients can query the database at this SPARQL endpoint:

http://wifo5-04.informatik.uni-mannheim.de/drugbank/sparql

The database can also be explored using this AJAX-based SPARQL Explorer.

This website is generated using D2R Server

3 star data → 5 star data
Using Food Ontology

5 star data

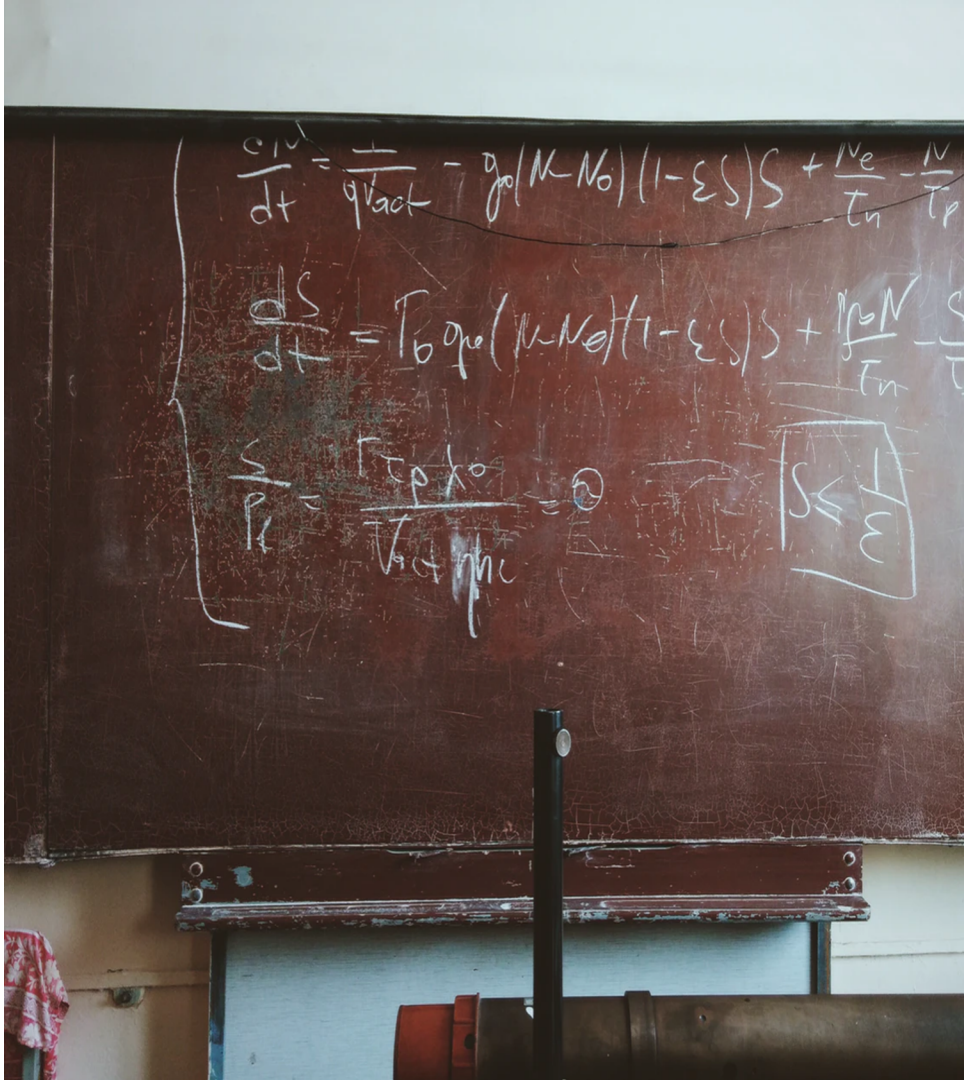
RDF Store
Virtuoso Universal Server

5 star data

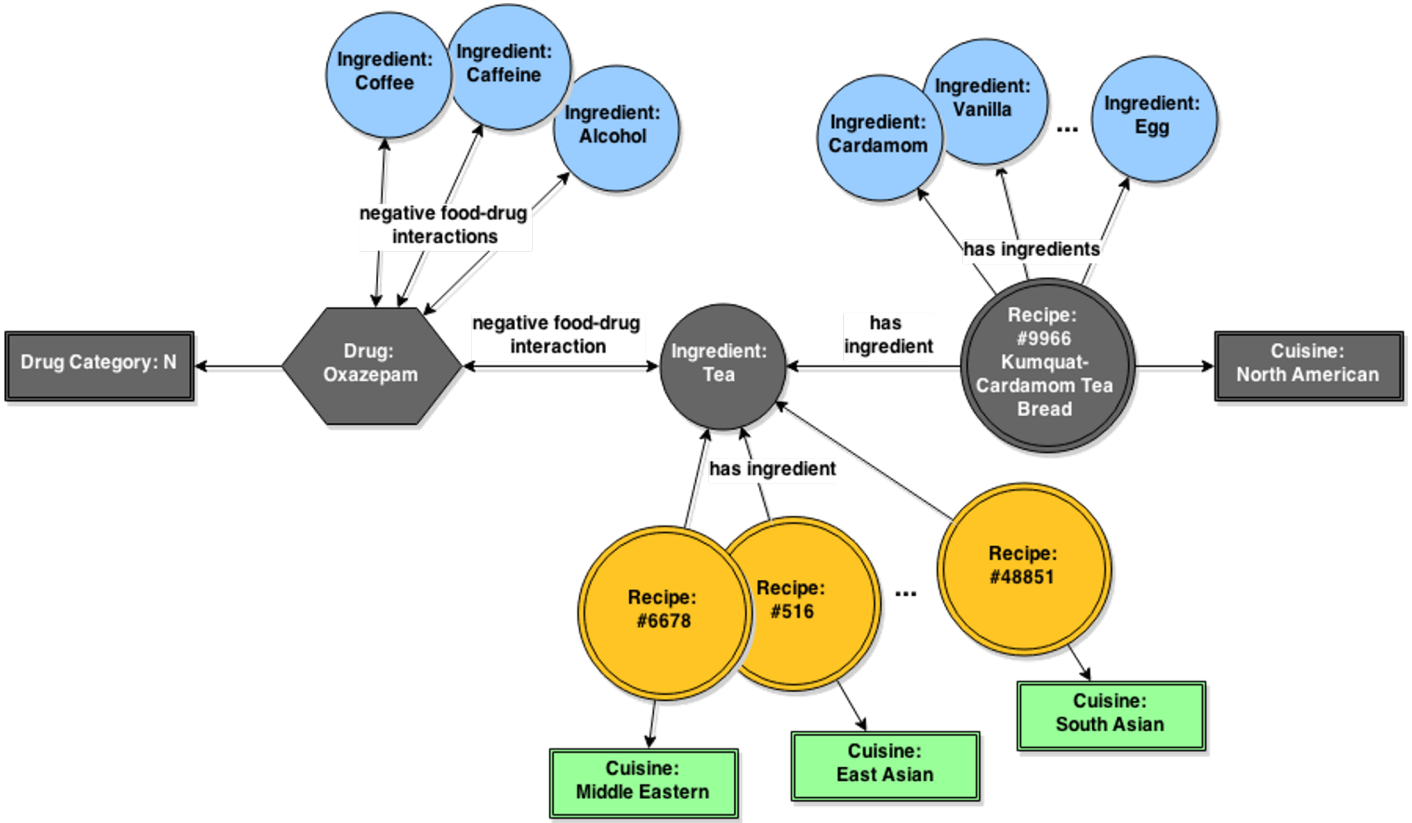
Ratio of Interactions

$$(E_I / P_I) * 1000$$

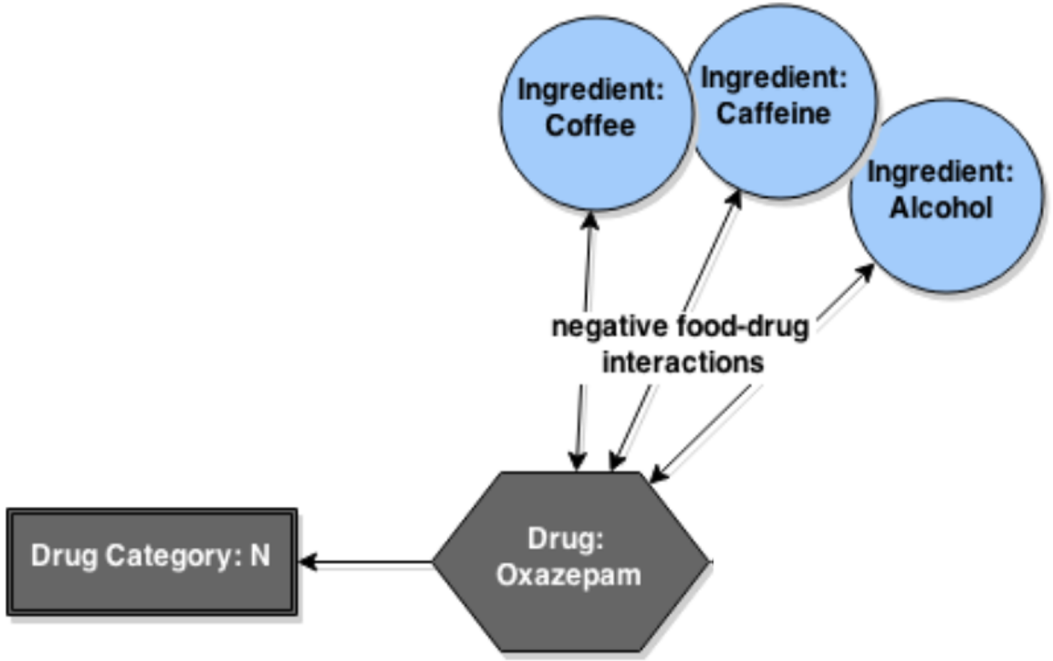
- E_I is the number of existing interactions in the dataset
- P_I is the number of possible interactions between the number of drugs in one category and the number of recipes in a cuisine



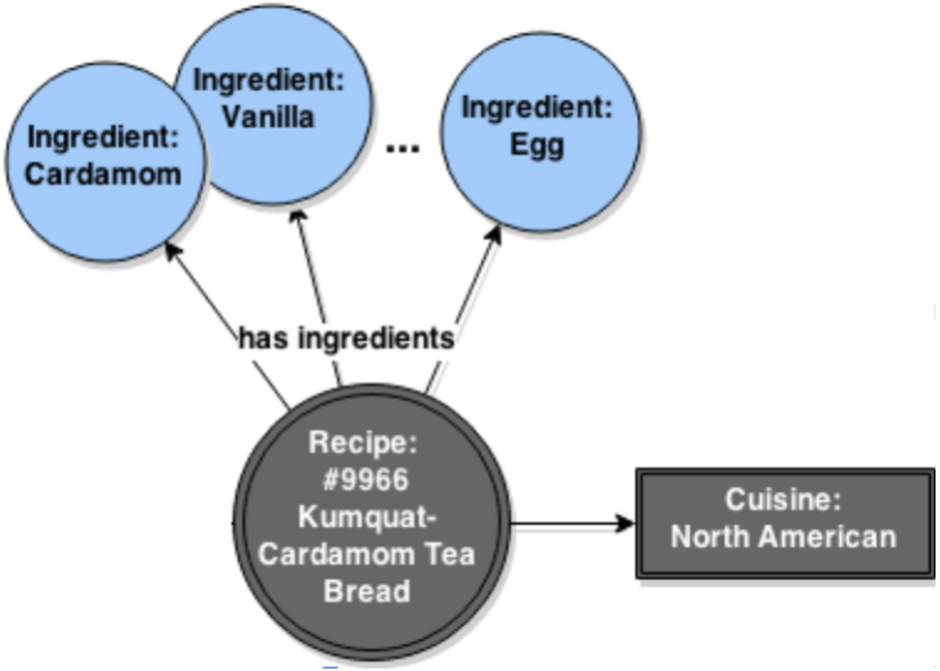
Cuisine-Drug Interaction Example



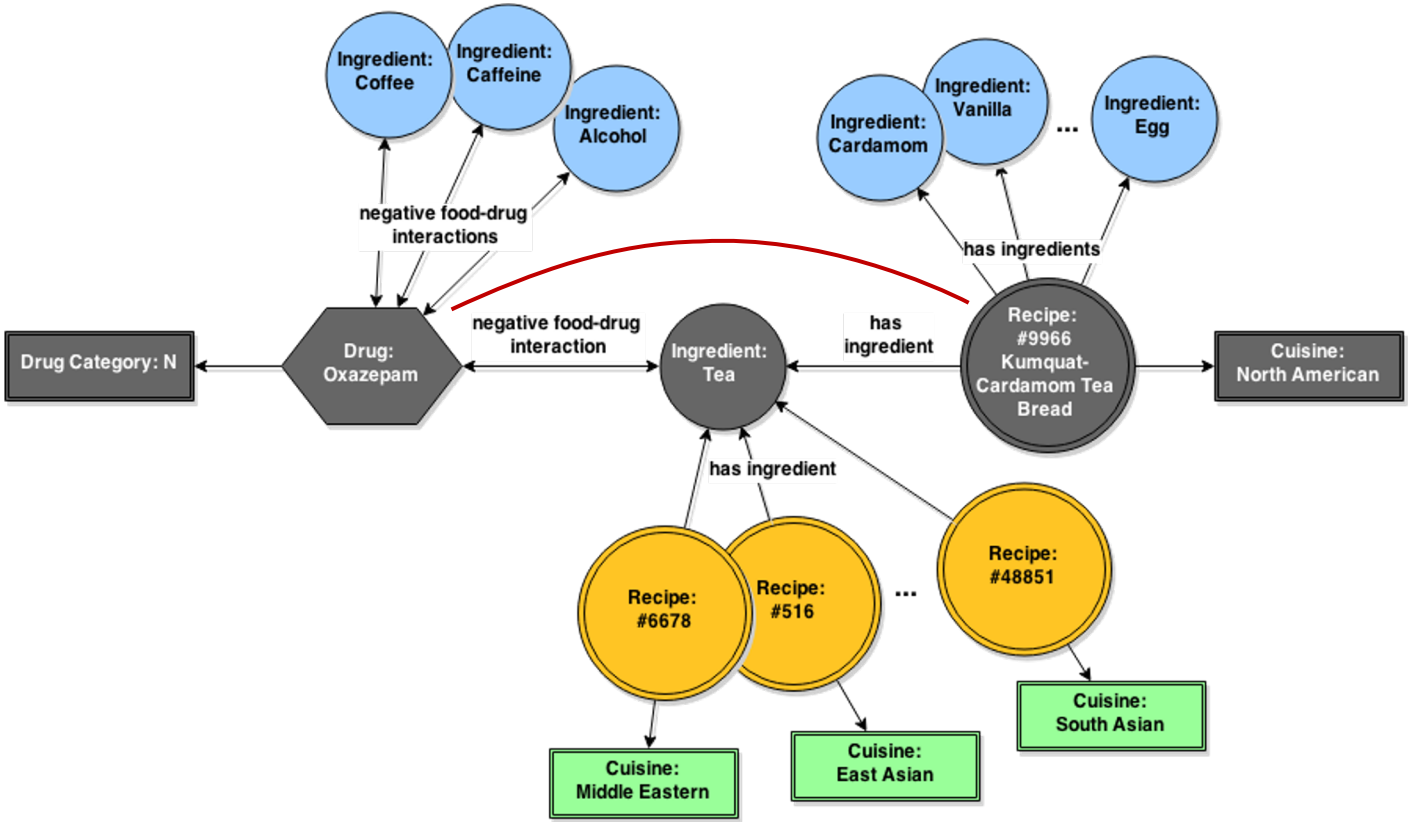
Cuisine-Drug Interaction Example



Cuisine-Drug Interaction Example



Cuisine-Drug Interaction Example



Results

Negative interactions between drugs from a given category and recipes from a given cuisine

Inferring Cuisine - Drug Interactions Using the Linked Data Approach

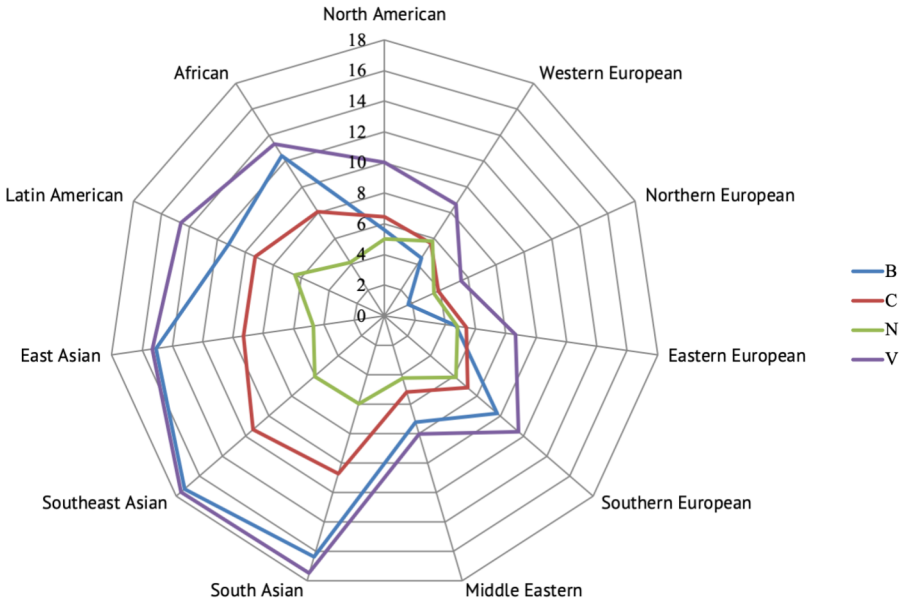
http://go.nature.com/rnPO91

Anatomical Therapeutic Chemical Classification System of Drugs

ATC - Code List

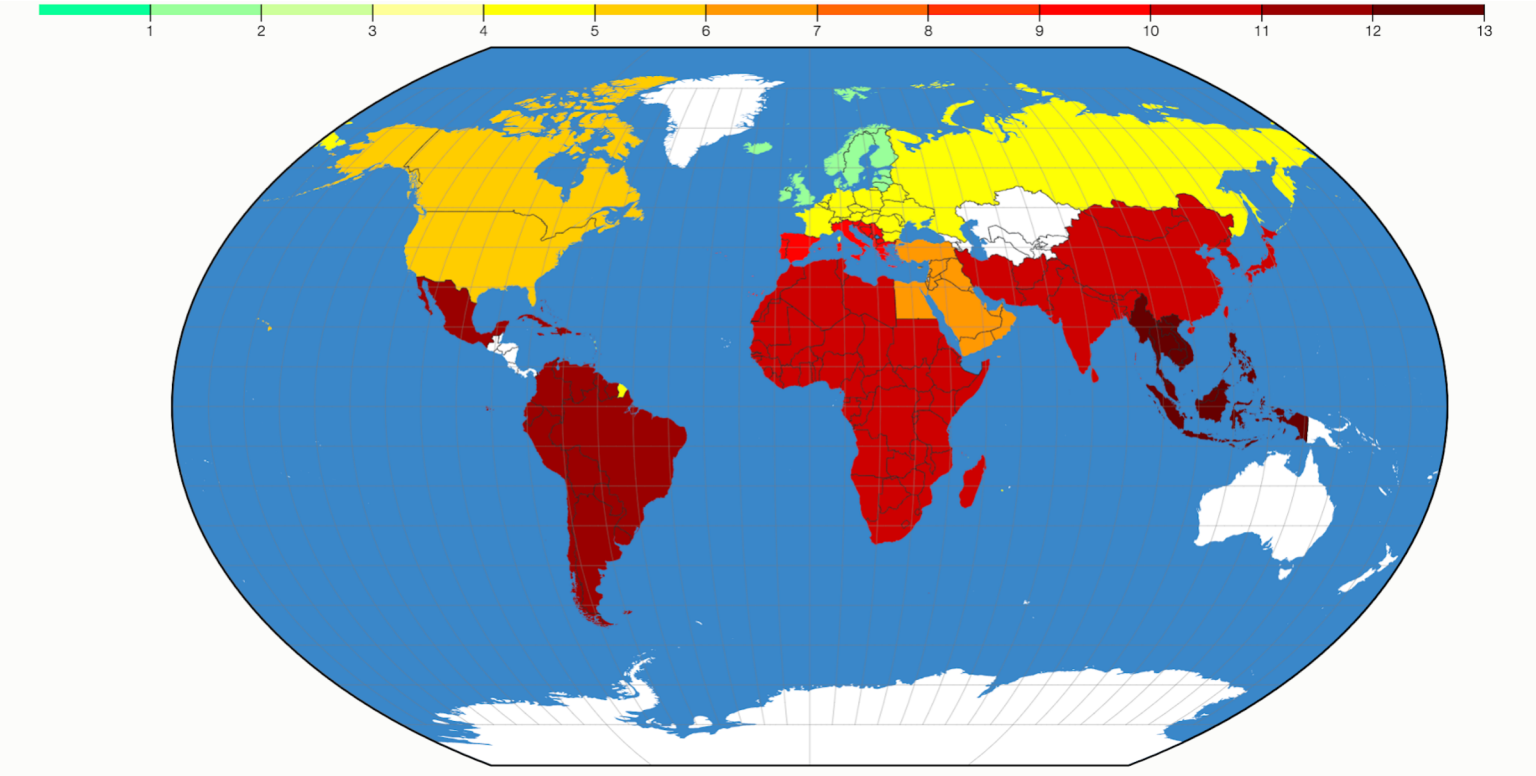
Code	Contents
A	Alimentary tract and metabolism
B	Blood and blood forming organs
C	Cardiovascular system
D	Dermatologicals
G	Genito-urinary system and sex hormones
H	Systemic hormonal preparations
J	Antiinfectives for systematic use
L	Antineoplastic and immunomodulating agents
M	Musculo-skeletal system
N	Nervous system
P	Antiparasitic products. Insecticides and repellents
R	Respiratory system
S	Sensory agents
V	Various

Patterns of Cuisine – Drug Interactions

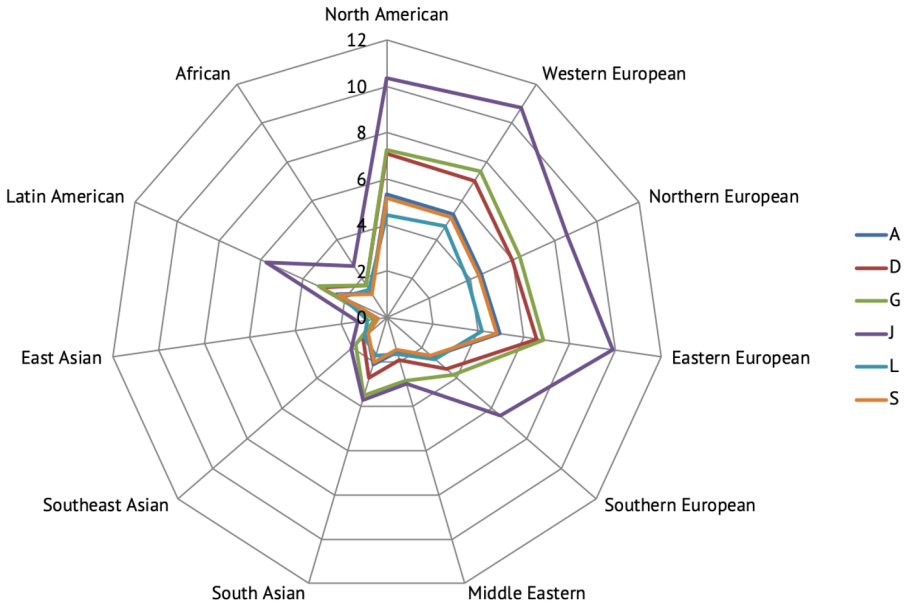


Code	Contents
A	Alimentary tract and metabolism
B	Blood and blood forming organs
C	Cardiovascular system
D	Dermatologicals
G	Genito-urinary system and sex hormones
H	Systemic hormonal preparations
J	Antiinfectives for systematic use
L	Antineoplastic and immunomodulating agents
M	Musculo-skeletal system
N	Nervous system
P	Antiparasitic products. Insecticides and repellents
R	Respiratory system
S	Sensory agents
V	Various

Food Interactions of Blood Drugs in ‰

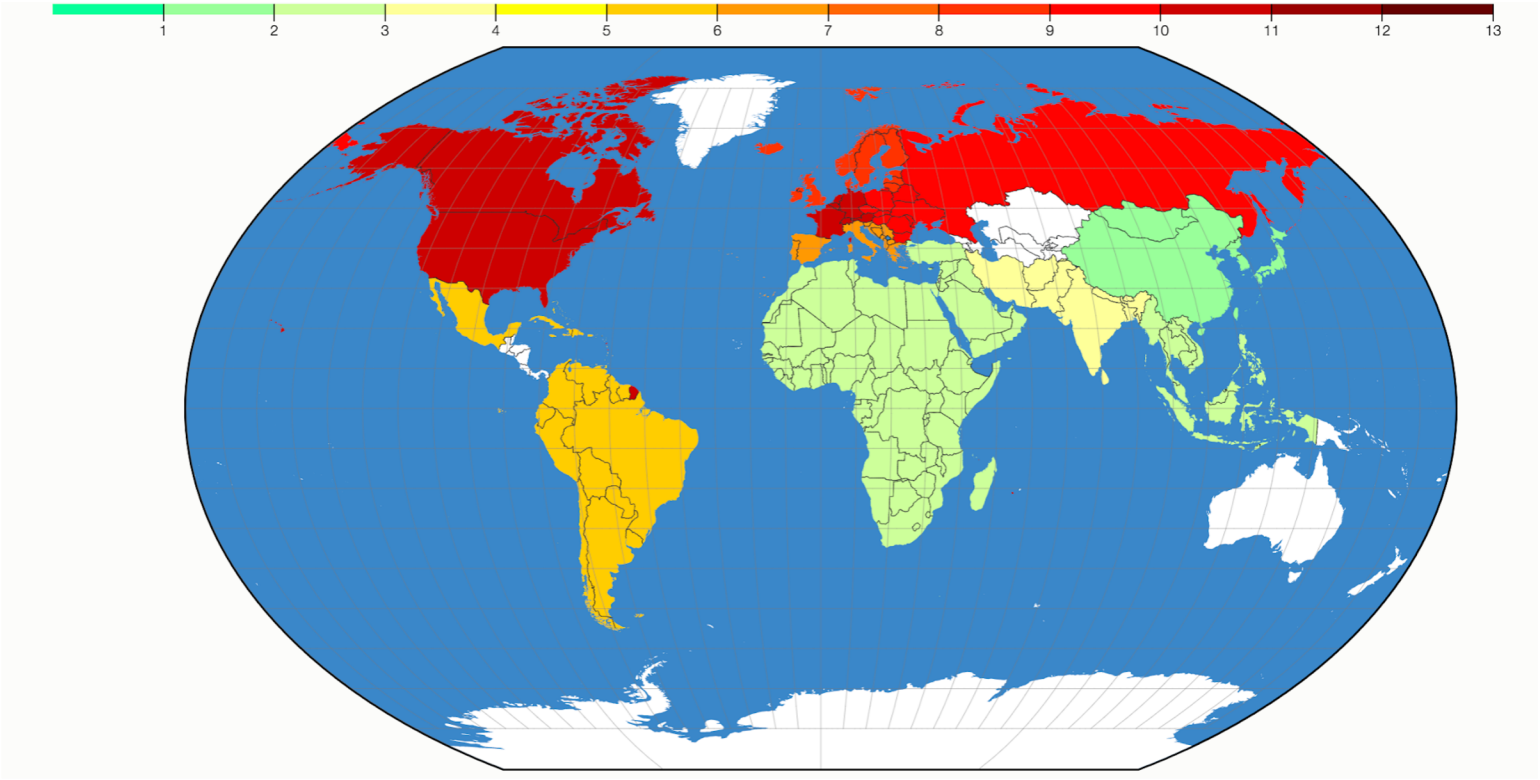


Patterns of Cuisine – Drug Interactions



Code	Contents
A	Alimentary tract and metabolism
B	Blood and blood forming organs
C	Cardiovascular system
D	Dermatologicals
G	Genito-urinary system and sex hormones
H	Systemic hormonal preparations
J	Antiinfectives for systematic use
L	Antineoplastic and immunomodulating agents
M	Musculo-skeletal system
N	Nervous system
P	Antiparasitic products. Insecticides and repellents
R	Respiratory system
S	Sensory agents
V	Various

Food Interactions of Antibiotics in ‰



Results

Ingredients' impact on the negative food - drug interactions in different parts of the world

Inferring Cuisine - Drug Interactions Using the Linked Data Approach

http://go.nature.com/rnPO91

Percentage of Ingredients Responsible for the Negative Food – Drug Interactions in the World

Ingredient	% of interactions involving the Ingredient
milk	56.110%
garlic	22.617%
coffee	8.388%
ginger	5.109%
cheese	2.197%
bacon	2.165%
red wine	1.865%
grapefruit	1.684%
ham	1.296%
wine	1.174%
tea	1.149%
avocado	0.869%
beer	0.304%
licorice	0.120%

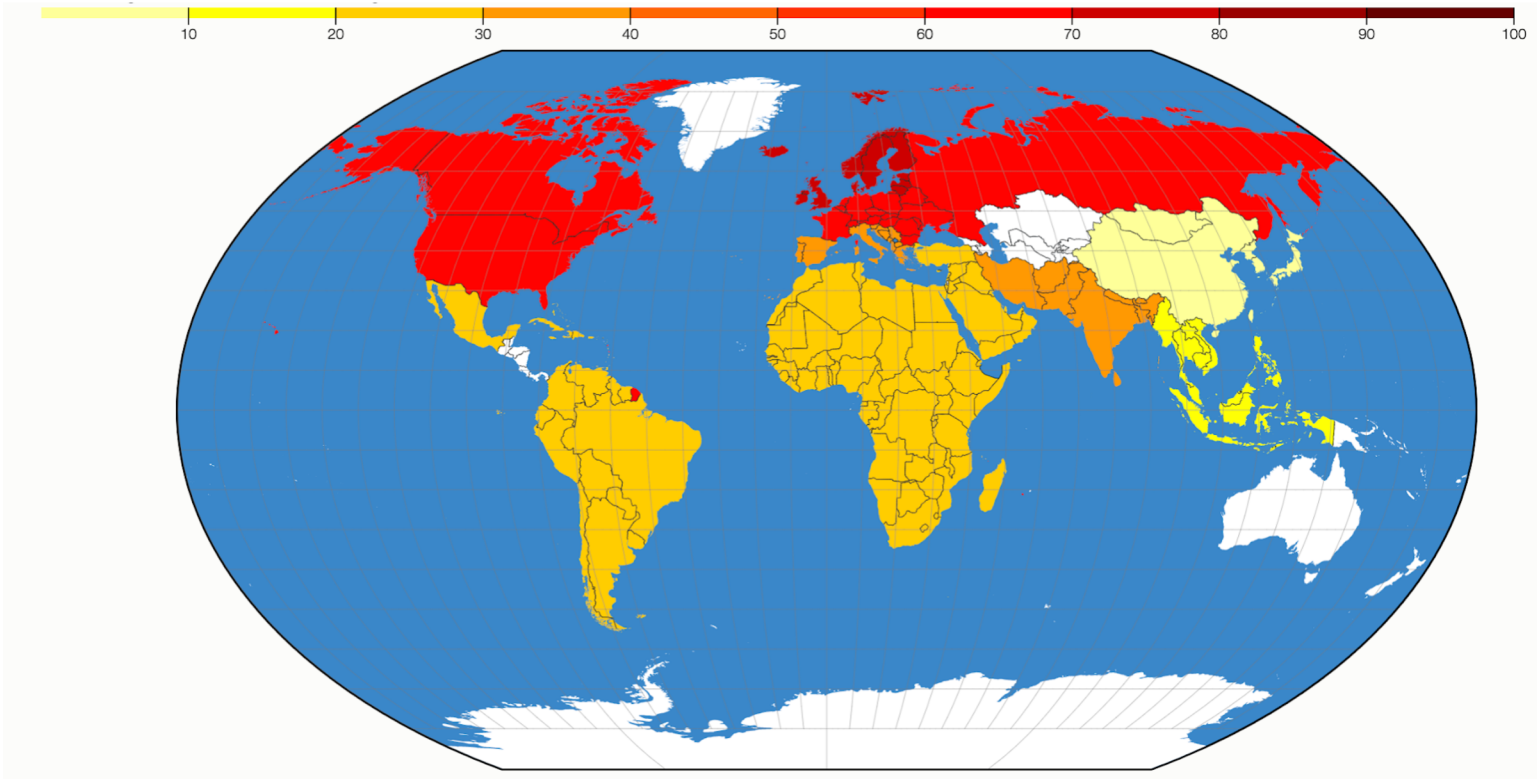
Percentage of interactions involving the Ingredient, for all Cuisines

Milk is the 'Problem' in the Western World, Garlic in South Europe, Asia, Africa, Latin America

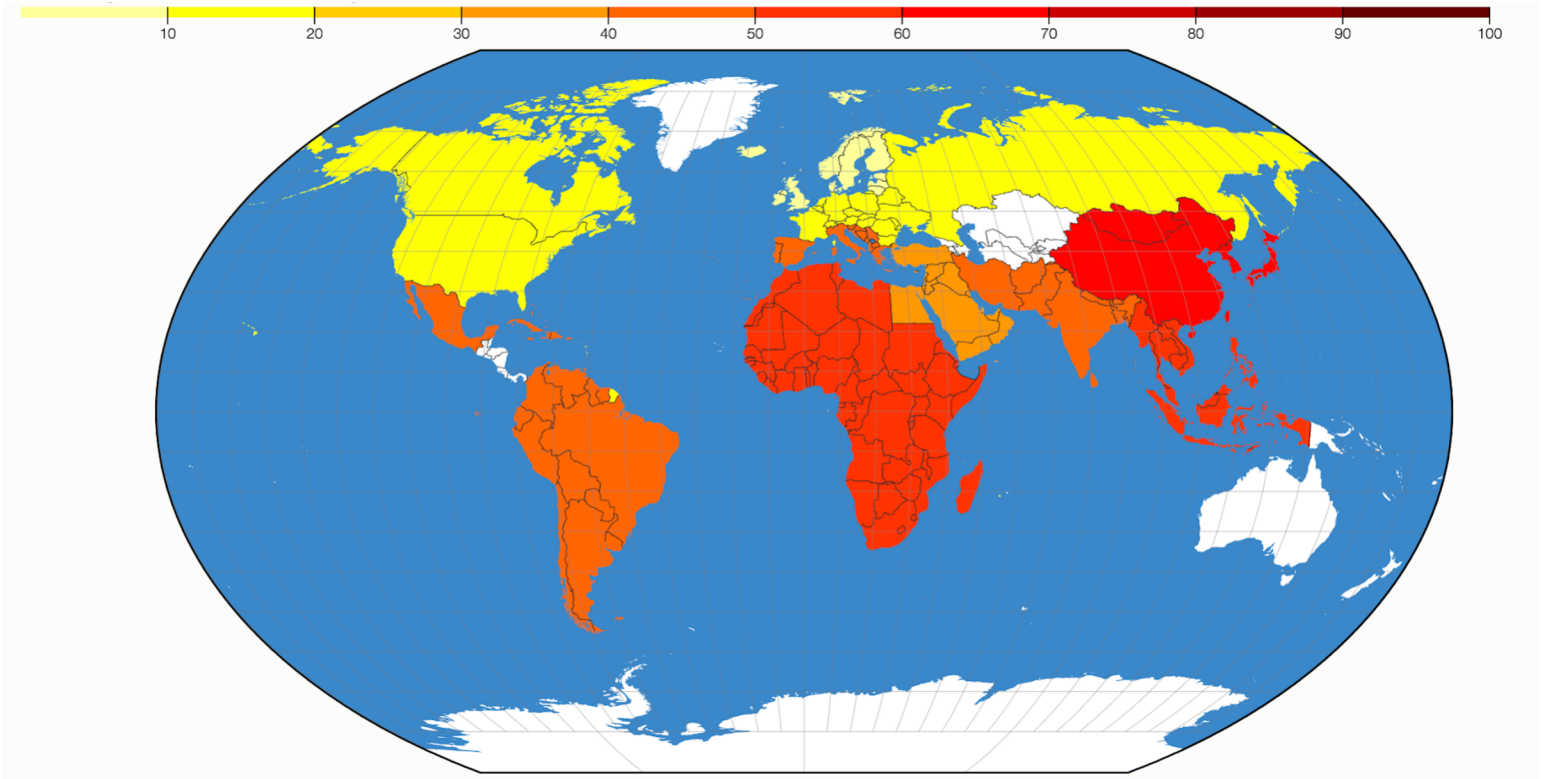
Cuisine	Top 3 interacting Ingredients
North America	Milk, garlic, coffee
Western European	Milk, garlic, coffee
Eastern European	Milk, garlic, coffee
Northern European	Milk, coffee, ginger
Southern European	Garlic, milk, coffee
Middle Asian	Garlic, milk, coffee
South Asian	Garlic, ginger, milk
Southeast Asian	Garlic, ginger, milk
East Asian	Garlic, ginger, milk
African	Garlic, ginger, milk
Latin American	Garlic, milk, avocado

Top 3 interacting Ingredients per Cuisine

The Global Impact of Milk in %



The Global Impact of Garlic in %



Some Useful Links

- https://datahub.io/dataset/drug-dataset
- https://datahub.io/dataset/recipe-dataset
- http://linkeddata.finki.ukim.mk/sparql
- ttp://viz.linkeddata.finki.ukim.mk/

Discussion

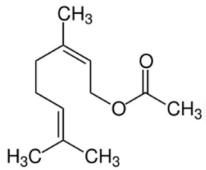
- Transformation and connection of the two datasets using Linked Data
- Basic Cuisine - Drug interaction analysis
- Discovery of two patterns of Cuisine - Drug category interactions
- The impact of milk and garlic

What Else Can We Discover Using Linked Data

How We Can Connect Food and Health Data

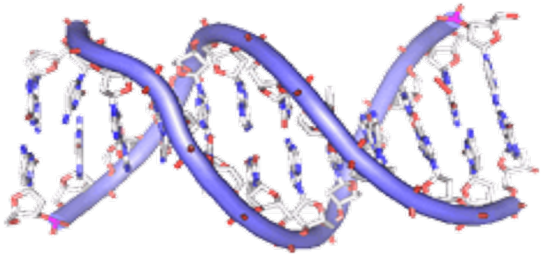
- Food Data - online recipes
  - Recipes
  - Cuisines
  - Ingredients
  - Serving sizes
  - Nutrition
  - Chemical information
- Health Data - Bio2RDF
  - Diseases
  - Drugs
  - Genomic information
  - Chemical information
  - And much more

How We Can Connect Food and Health Data

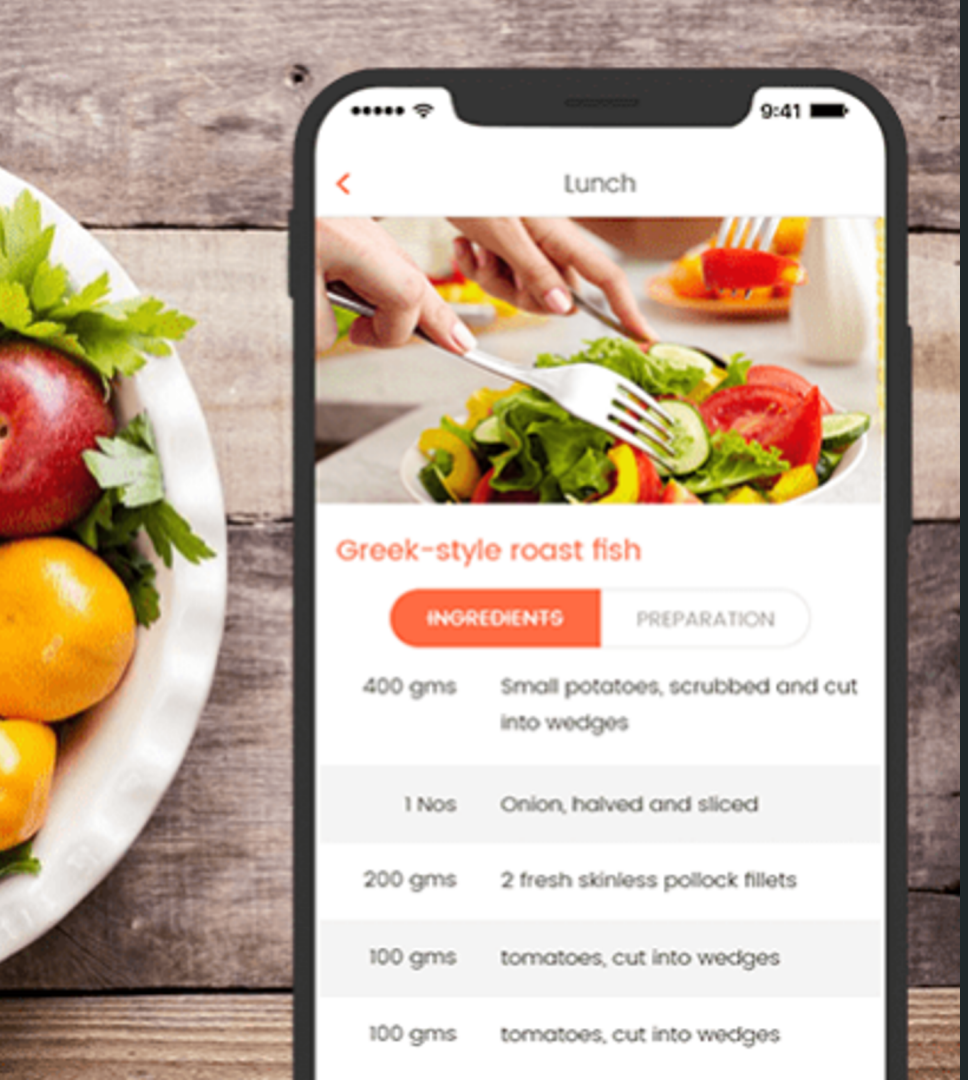


Ingredients - nutrients & chemical compounds

Genes, proteins, chemical compounds, drugs



GeneAge - genes related to ageing and longevity



Personalized Diet to Improve our Health

- Medical Conditions
- Prescription Drugs
- Nutrition
- Health goals
- Recipes recommendations
- Reminders, notifications ...

Conclusion

Data is the new oil.
But it is only useful when it is refined.

Clive Humby, UK Mathematician and architect, 2006

Questions?