# Learning to Read Academic Paper

Yining Hong, Jialu Wang, Xueheng Zhang, Zheng Wu

May 26th, 2018

# Why

Why

# What

Why

What

# How

Why

What

How

What How

# Motivation

How

# Motivation

Service

# Motivation

Service

Pipeline

# Overview

# Motivation

# Service

# Pipeline

# 1st

# Motivation

# Why we need

# “Learning to Read Academic Paper”?

# arXiv

- Nowadays tens of thousands of papers released on arXiv every day
- Not all of them are of high quality

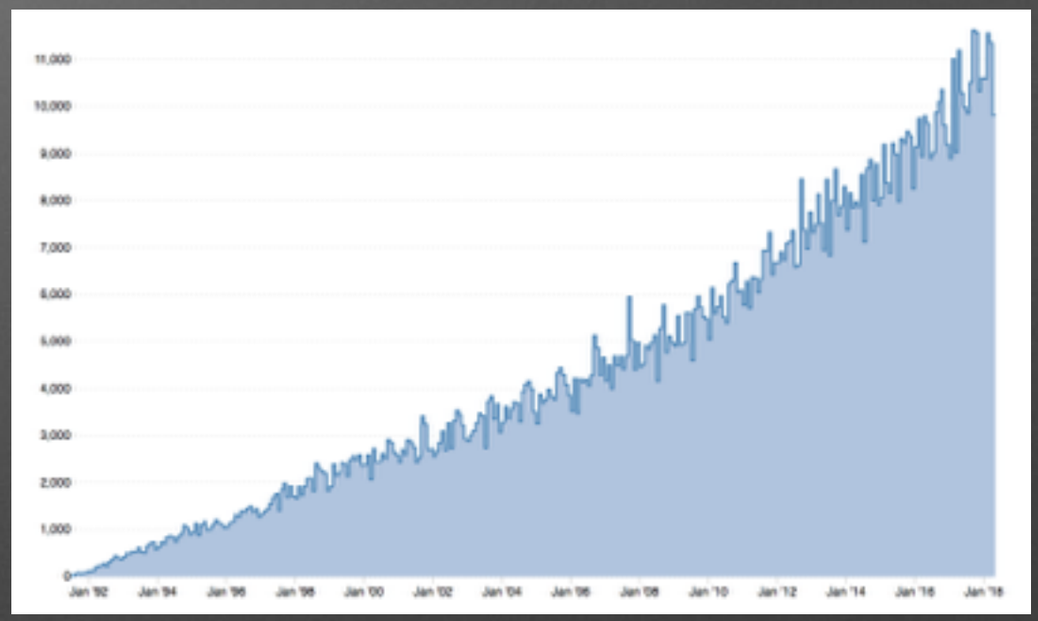



Fig 1. arXiv submission rate

# 2nd

# Service

# Service

Acemap

WeChat

# Acemap

# WeChat

3rd

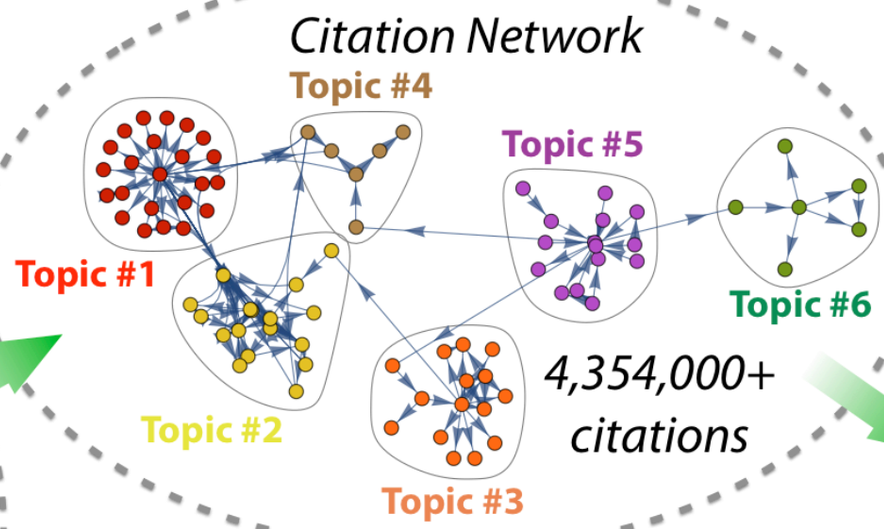
# Pipeline

# Pipeline

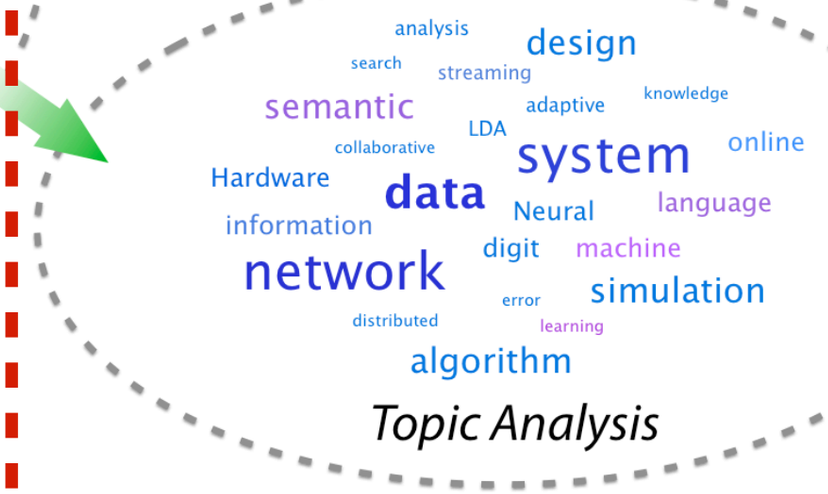
## Data

2,244,000+ papers

## Model




4,354,000+ citations




Topic Analysis

## Visualization


User Interface

# Data

Acemap Dataset


Python Crawler

# Model

- FastText— View each paper individually

# Model

Hofmann
Probabilistic Latent Semantic Indexing

Green et al.
Modelling heterogeneity with and without the Dirichlet
...

Jordan et al.
An Introduction to Variational Methods for Graphical ...

Collins et al.
A Generalization of Principal Component Analysis to ...

Roweis
EM Algorithms for PCA and SPCA

Blei et al.
Latent Dirichlet Allocation

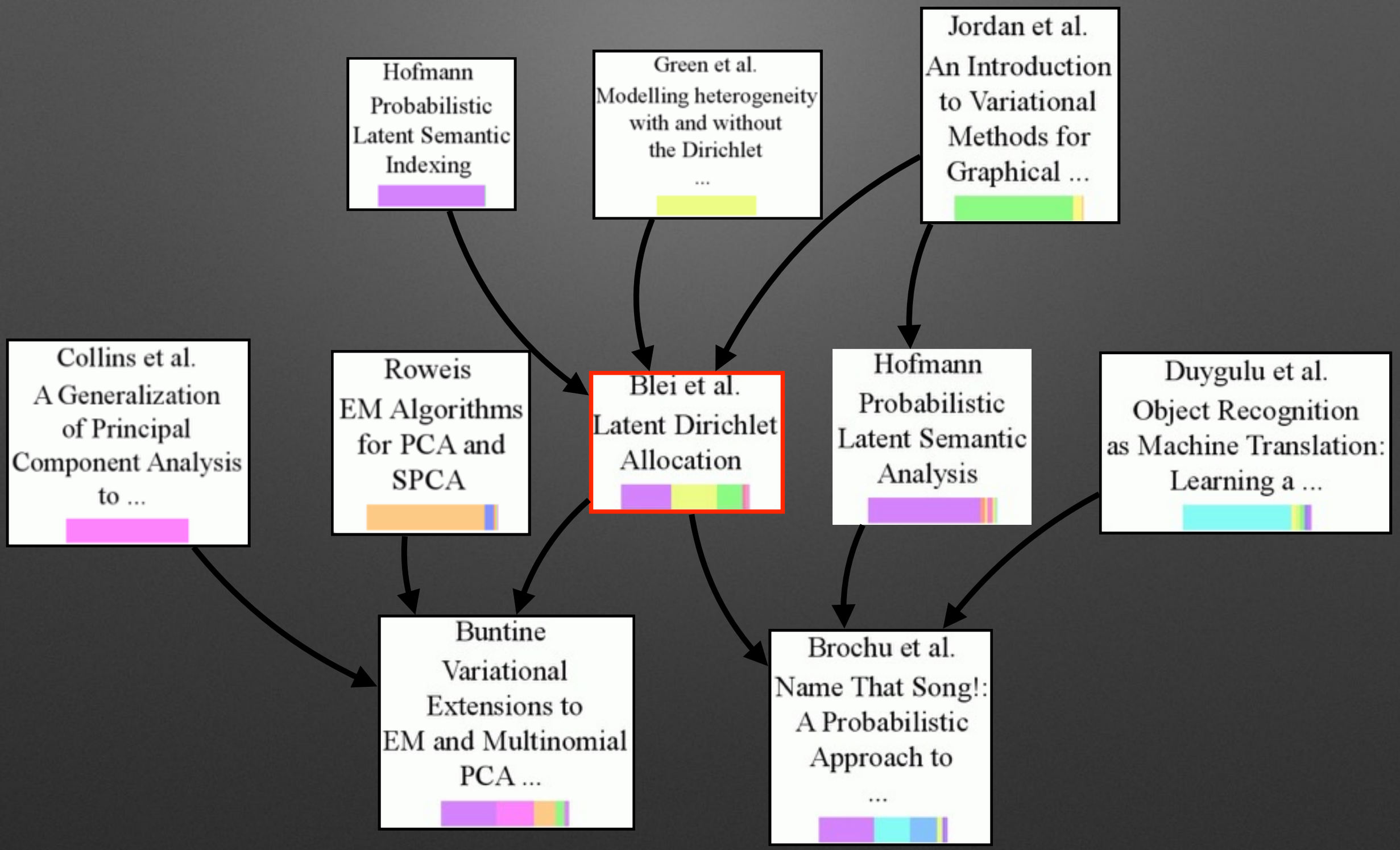
Hofmann
Probabilistic Latent Semantic Analysis

Duygulu et al.
Object Recognition as Machine Translation: Learning a ...

Buntine
Variational Extensions to EM and Multinomial PCA ...

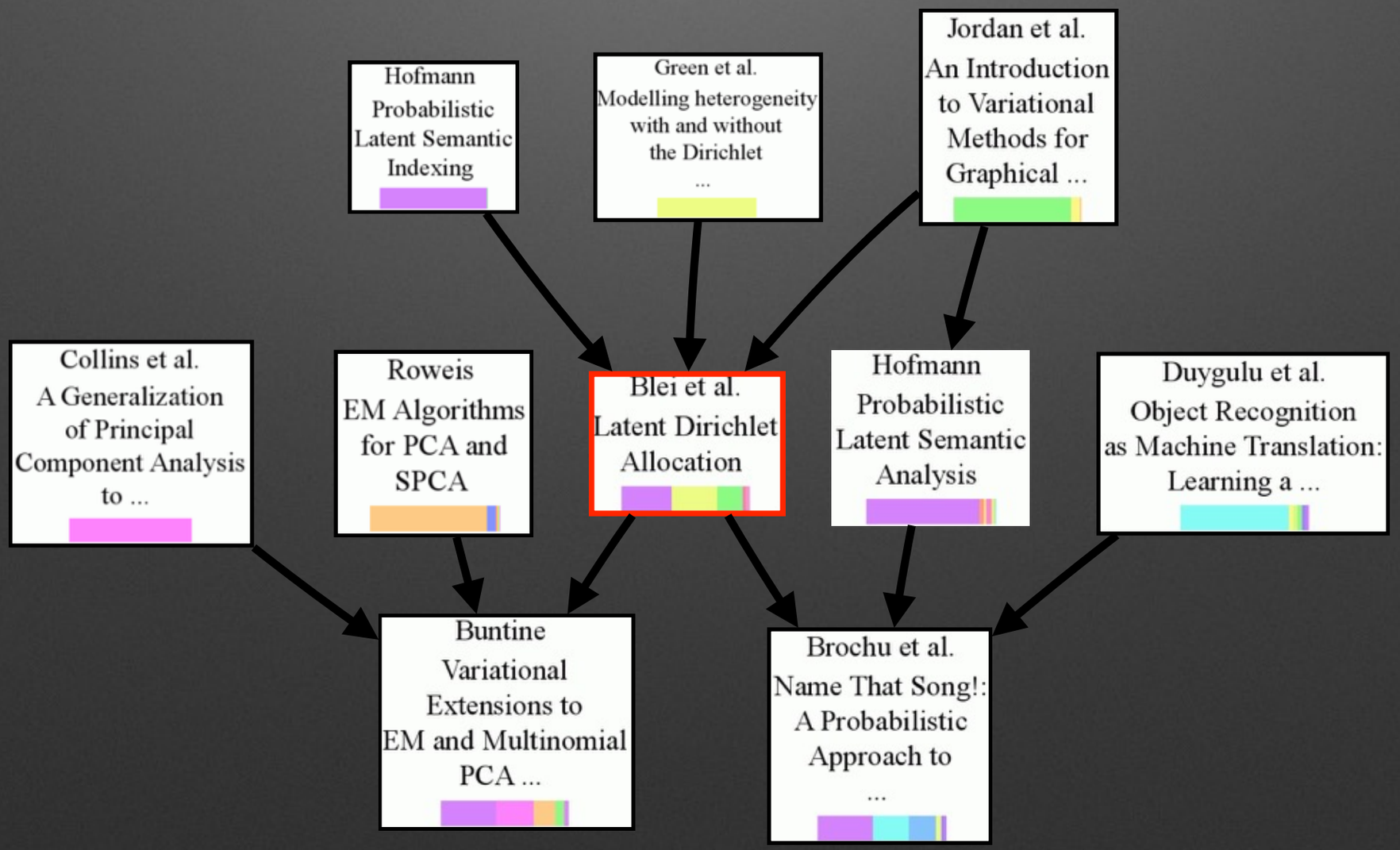
Brochu et al.
Name That Song!: A Probabilistic Approach to
...

# Model


Hofmann
Probabilistic Latent Semantic Indexing
Green et al.
Modelling heterogeneity with and without the Dirichlet ...
Jordan et al.
An Introduction to Variational Methods for Graphical ...
Collins et al.
A Generalization of Principal Component Analysis to ...
Roweis
EM Algorithms for PCA and SPCA
Blei et al.
Latent Dirichlet Allocation
Hofmann
Probabilistic Latent Semantic Analysis
Duygulu et al.
Object Recognition as Machine Translation: Learning a ...
Buntine
Variational Extensions to EM and Multinomial PCA ...
Brochu et al.
Name That Song!: A Probabilistic Approach to ...

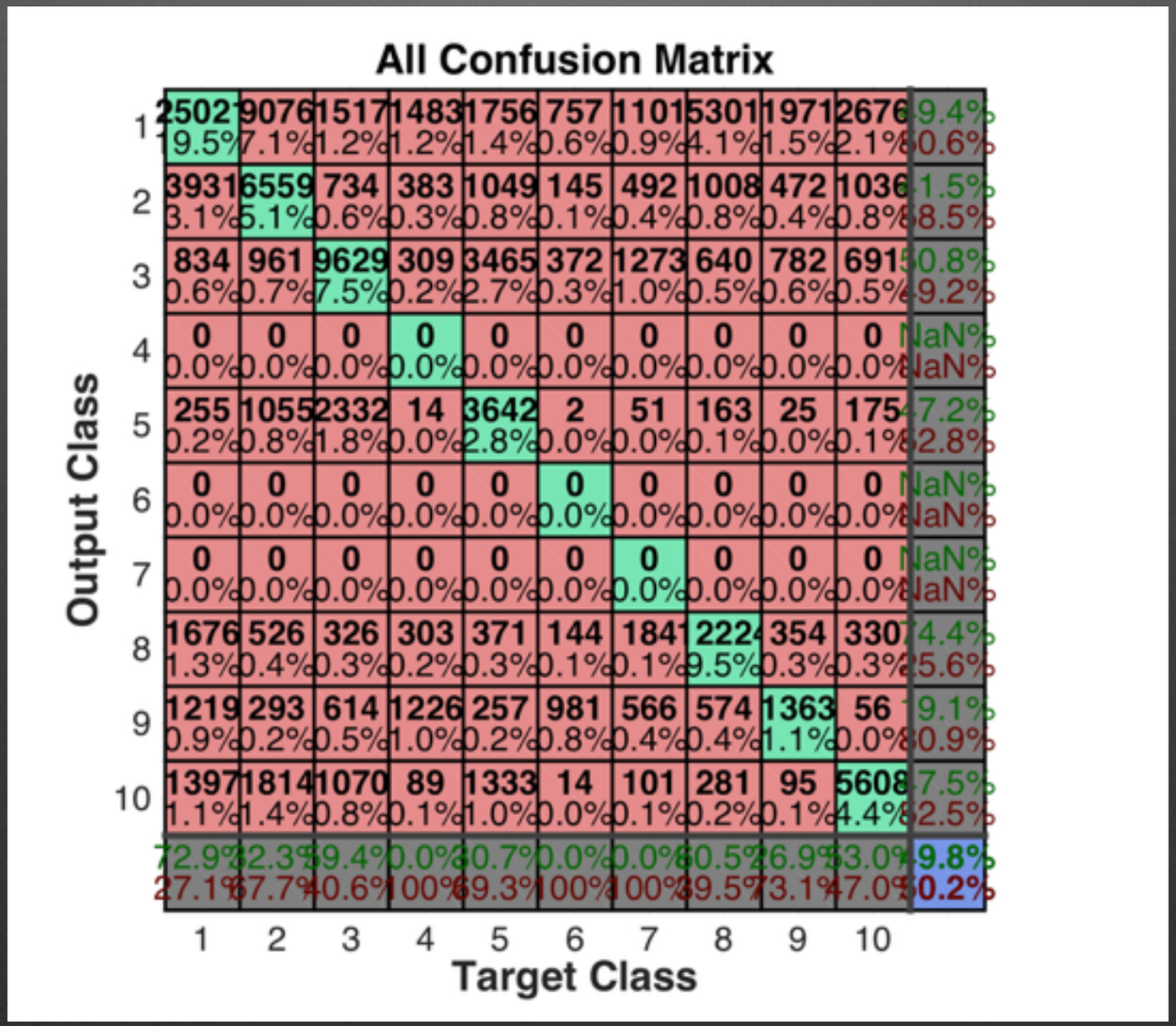
# Model

- Citation Network— Analyze relationship among papers



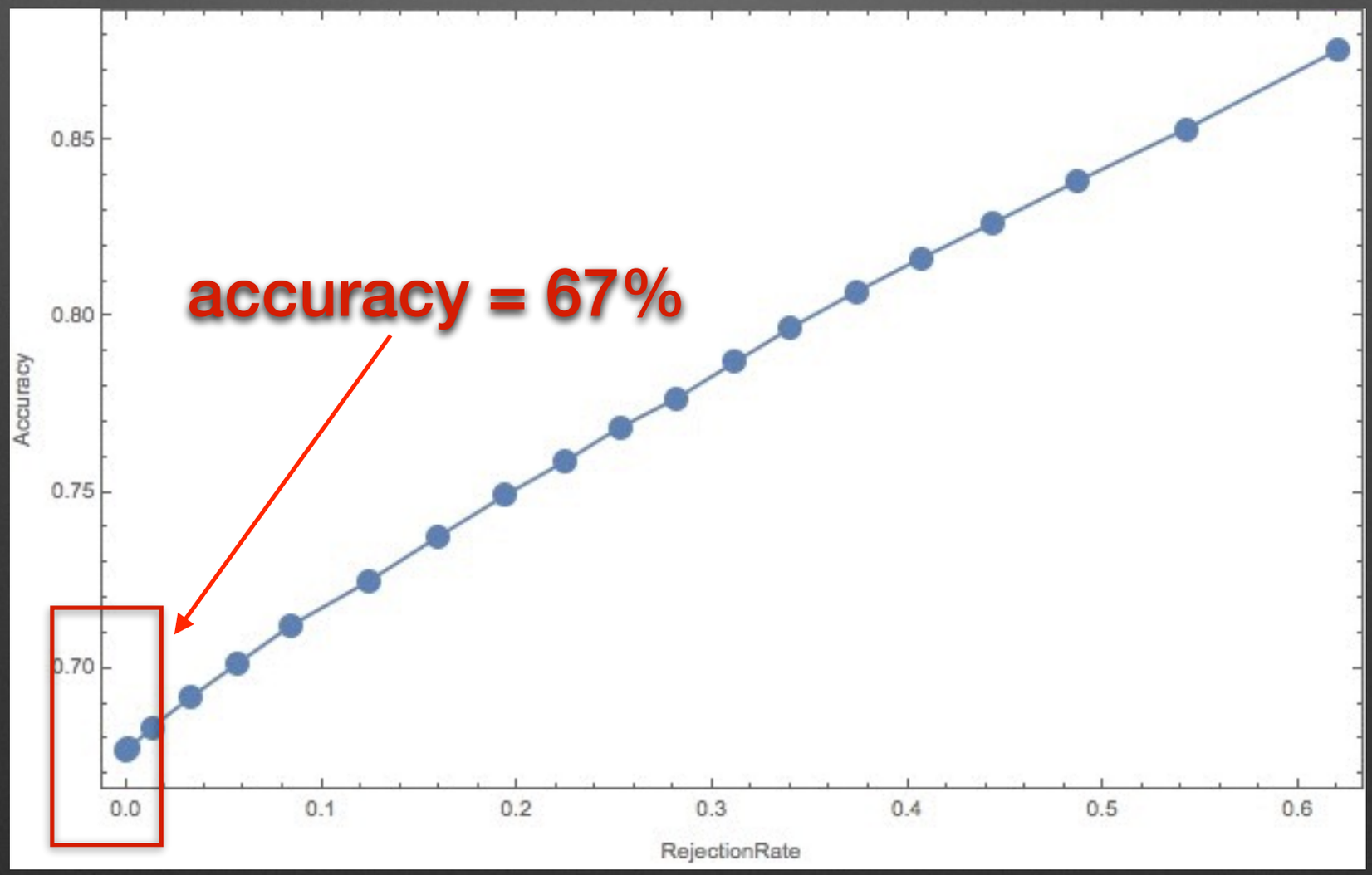
# Model

- LDA + neural network



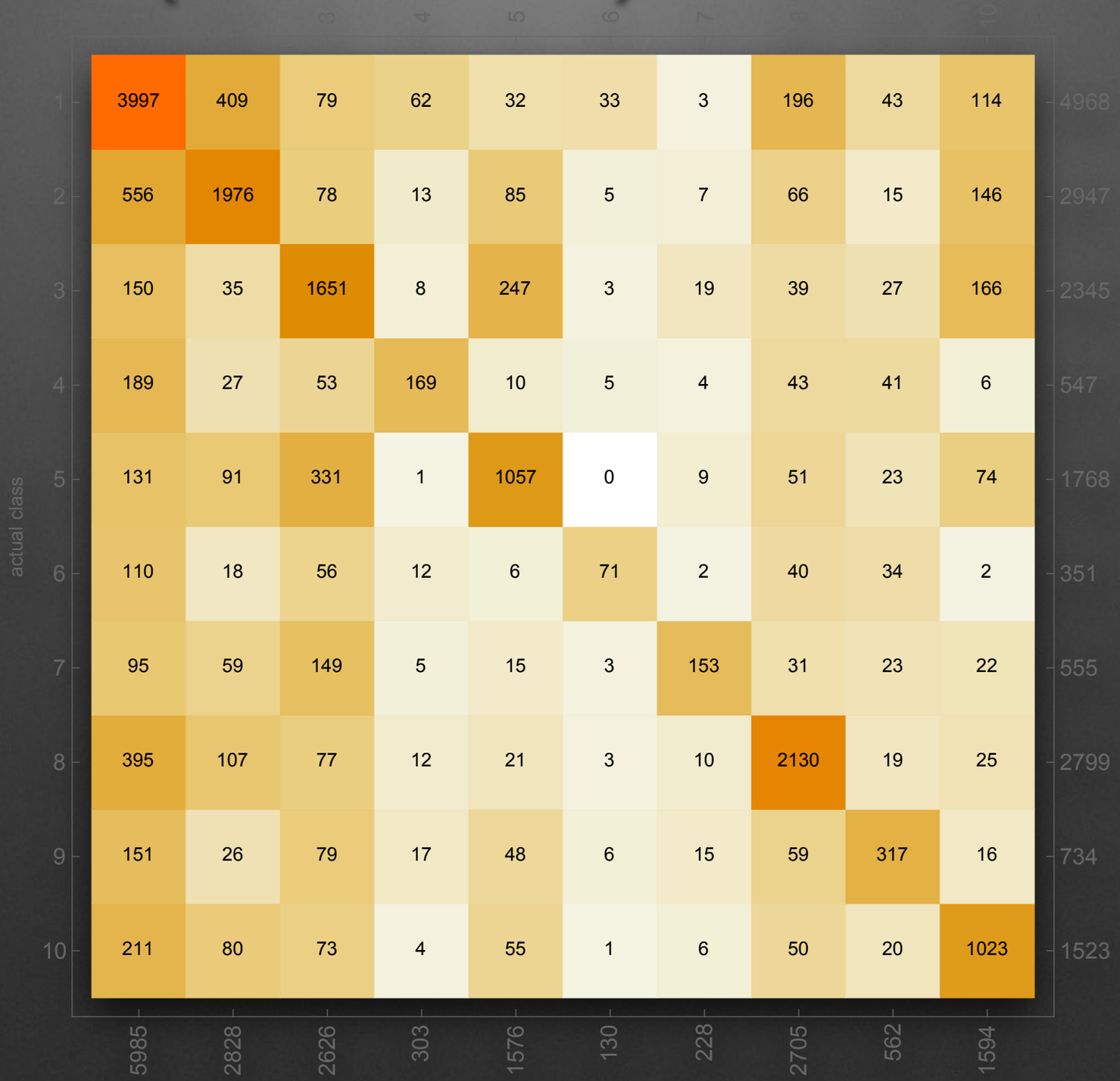
# Model

- Markov Chain ( on title words )
    - Mathematica built-in classifier



# Model

- Markov Chain ( on title words )



# Model

- IEEE taxonomy ( TXT Format )

IEEE_Taxonomy_clean.json × IEEE_Taxonomy_clean.txt ×

```
1  Aerospace and electronic systems
2  ....Aerospace control
3  ........Air traffic control
4  ........Attitude control
5  ........Ground support
6  ....Aerospace engineering
7  ........Aerospace biophysics
8  ........Aerospace electronics
9  ........Aerospace safety
10 ............Air safety
11 ........Aerospace simulation
12 ........Aerospace testing
13 ........Satellites
14 ............Artificial satellites
15 ............Earth Observing System
16 ............Low earth orbit satellites
17 ............Moon
18 ............Space stations
19 ........Space technology
20 ............Space exploration
21 ....Aerospace materials
```

# Model

- IEEE taxonomy ( XML Format )

IEEE_Taxonomy_clean.json

IEEE_Taxonomy_clean.json × IEEE_Taxonomy_clean.txt ×

```
{
    "id": "Node0",
    "name": "root",
    "data": [],
    "children": [
        {
            "id": "Node1",
            "name": "Aerospace and electronic systems",
            "data": [],
            "children": [
                {
                    "id": "Node2",
                    "name": "Aerospace control",
                    "data": [],
                    "children": [
                        {
                            "id": "Node3",
                            "name": "Air traffic control",
                            "data": [],
                            "children": []
                        },
```

# Model

- IEEE keywords

test.xml

data_1_1.xml.txt × test.xml ×

```
1 <doc>
2 <field name="id">89</field>
3 <field name="title">Analysis of semiconductor laser arrays with high-intensity
  uniformity</field>
4 <field name="conference">Quantum Electronics, IEEE Journal of</field>
5 <field name="abstract">It is shown that an array with a highly uniform
  intensity distribution can be constructed. This decreases the spatial hole
  burning and makes stable, high-power operation easier. An estimate of the
  remaining spatial hole burning is derived and it is shown how its effects can
  be compensated for. It is considered that laser arrays which operate in the
  fundamental mode even at high-power levels can be designed</field>
6 <field name="ieeeTerms">Charge carrier density  Chirp  Laser beams Laser modes
   Laser theory  Optical arrays  Optical coupling  Optical design  Phased
  arrays  Semiconductor laser arrays</field>
7 <field name="citation">4247404  4256348 4852568 4853092 4853296 4251277 4250612
   4256577 4855377 4256639 4855456 1072927 1073433 1073204 4251507 4854469
  4251707 1074291 1073042 4257109</field>
8 </doc>
```

# Visualization

# Visualization

- Solr — Open source search engine


- Amaze UI — HTML5 Web page framework
- JIT — JavaScript InfoVis ToolKit

# Visualization

- Deploy Apache on our server
- Install and configure Solr
- Develop web front-end pages ( Amaze UI )
- Transform all dataset to required formats ( json, xml )
- Run Solr and return search results to specific webpage
- Add animation with JIT

# Visualization

AceRec | Get Started | Topic List | Result | Help | Login

## Academic Paper Recommendation

请输入你想找的话题

Search | More info

**By Auther:**

A B C D E F G H I J K L M N O P Q R S T U V W X Y Z

**By Topic:**

A B C D E F G H I J K L M N O P Q R S T U V W X Y Z

**By Published Year:**

Before 1950 | 1950s | 1960s | 1970s | 1980s | 1990s | 2000-2005 | 2006 | 2007 | 2008 | 2009 | 2010 | 2011 | 2012 | 2013 | 2014

# Visualization

**By Topic:**

A B C D E F G H I J K L M N O P Q R S T U V W X Y Z

**By Published Year:**

Before 1950 | 1950s | 1960s | 1970s | 1980s | 1990s | 2000-2005 | 2006 | 2007 | 2008 | 2009 | 2010 | 2011 | 2012 | 2013 | 2014

## Statistics

# of publications: 2,909,788
# of authors: 1,537,934
# of conferences: 3,759
# of journals: 1,409

## Topic Display

### Feature 1

The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog.

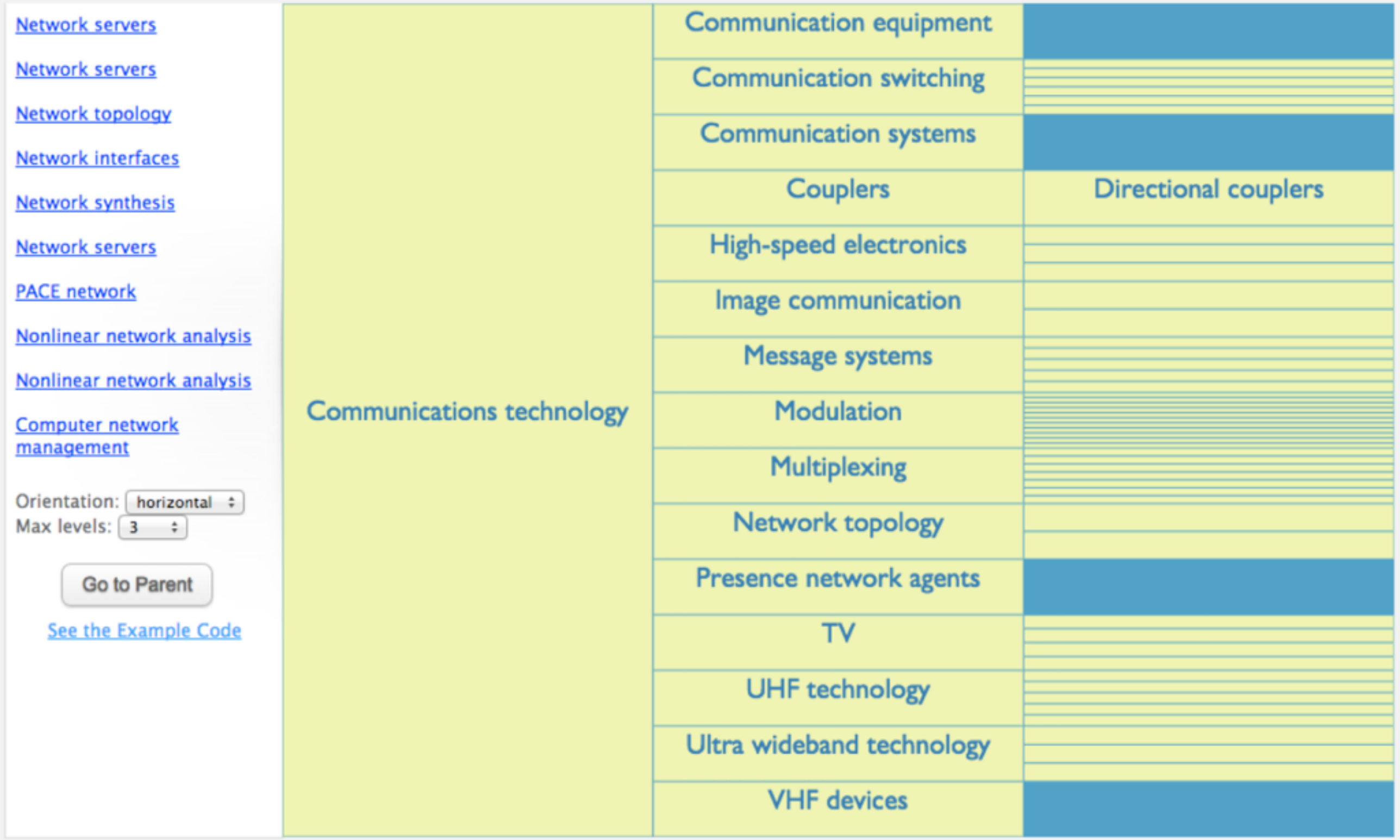
### Feature 2

The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog.

### Feature 3

The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog.

# Visualization



# Review

- Motivation
- Service
    - Acemap, WeChat
- Pipeline
    - Data / Model / Visualization

# Thank you

# Q & A