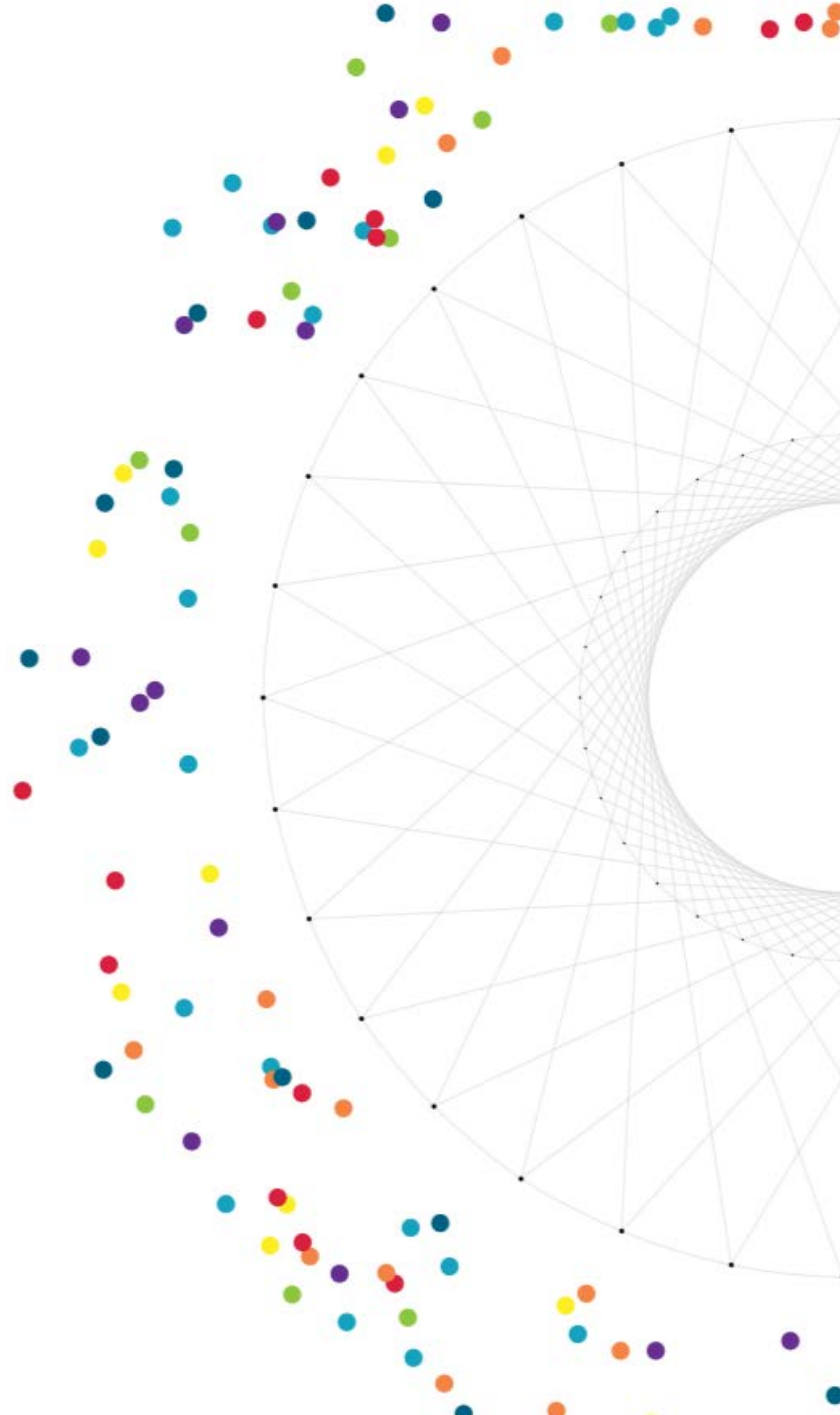
Creating Knowledge Map For Topic Search

Te Lin
5140219432

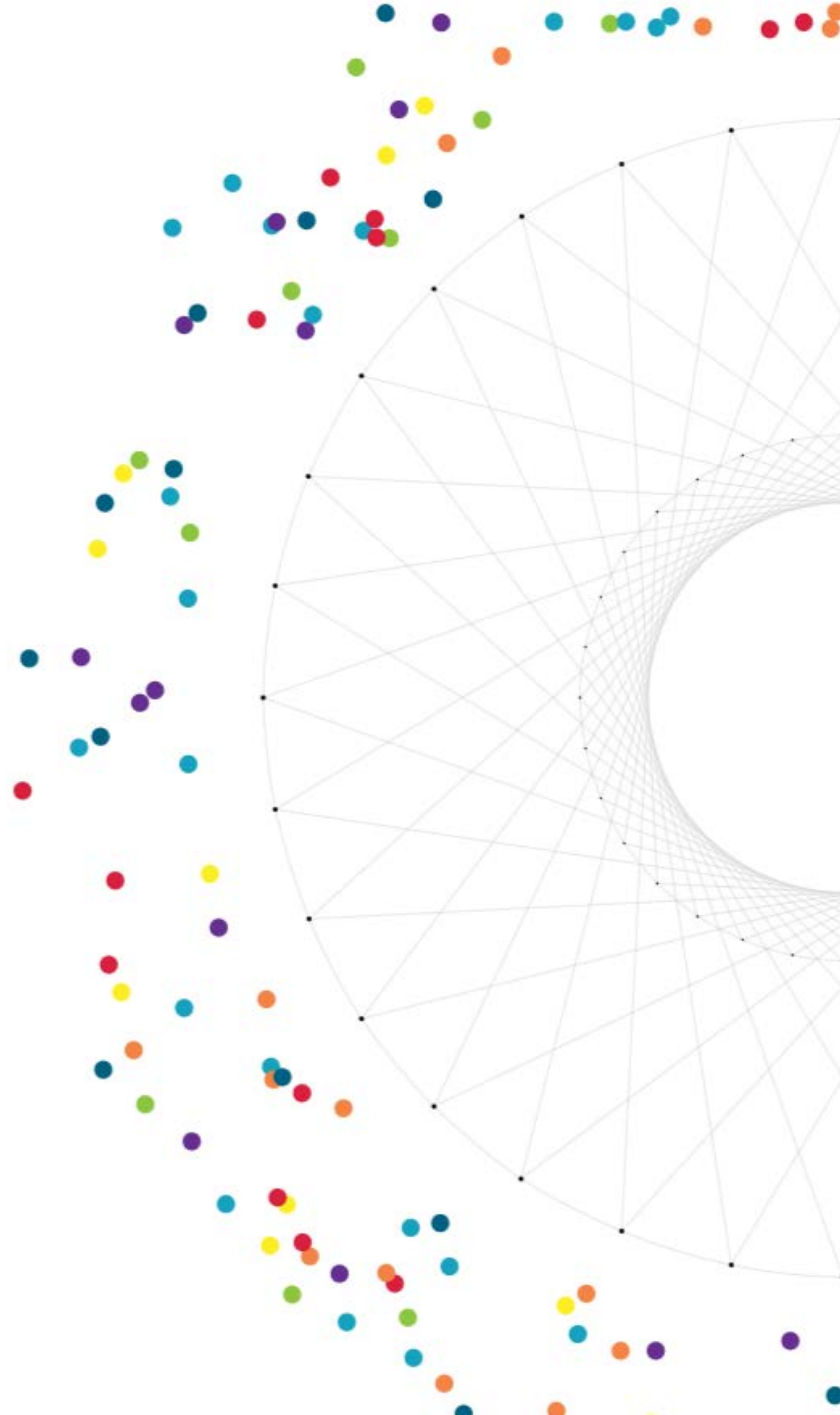
Outlines

- Motivation for Knowledge Map
- Construction
- Manipulating the Scale
- Visualization
- Future Work



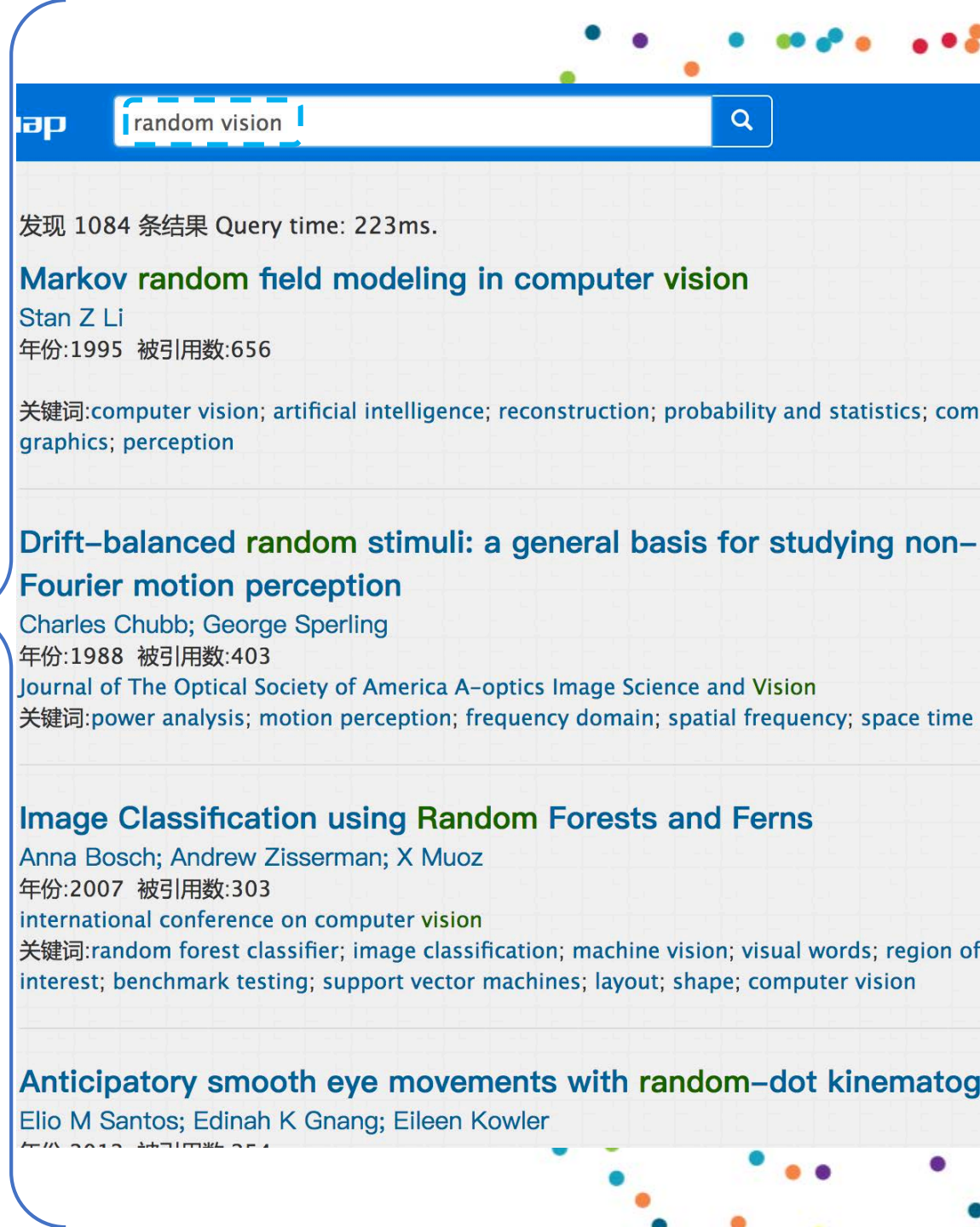
Guide

- Motivation for Knowledge Map
- Construction
- Manipulating the Scale
- Visualization
- Future Work



1. When users are searching the topic “random vision”

- No topic showing
- No information about topic related
- None clear idea of relation and hierarchy



2. When users are searching the topic “machine learning”

- Redundancy
- Vague hierarchy & relation
- None informative

父话题:

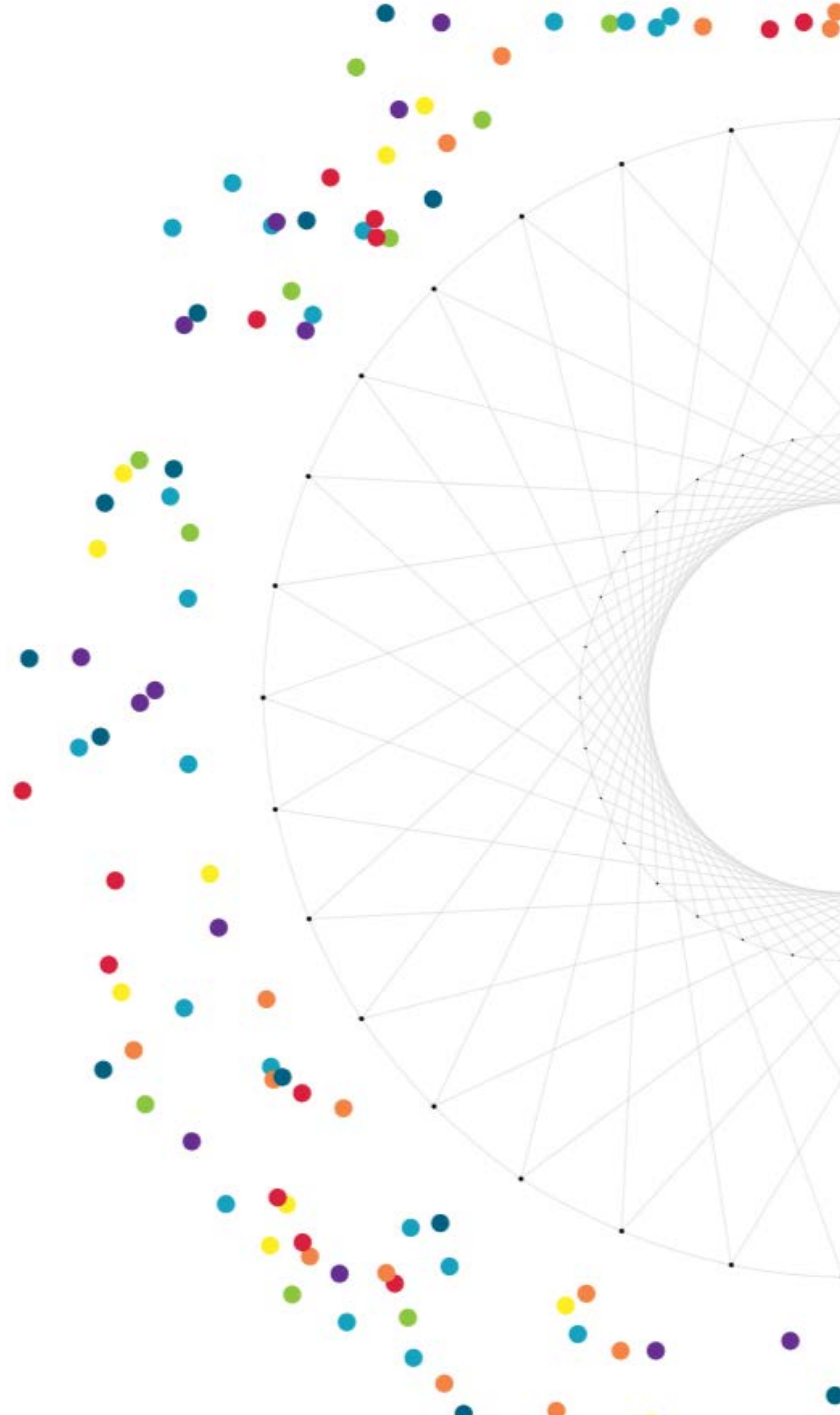
Computer Science

子话题:

Instance-based learning, Data stream mining, Feature selection, Heuristic, Gene expression programming, Swarm intelligence, Connectionism, Fuzzy control system, Recommender system, Unsupervised learning, Linear discriminant analysis, Intelligent word recognition, Feature learning, Probabilistic classification, Conceptual clustering, Convolutional neural network, Evolutionary computation, Intelligent character recognition, Semi-supervised learning, Computer-automated design, Algorithmic learning theory, Biological neural network, Computational intelligence, Computational learning theory, Anomaly detection, Learning to rank, Hidden Markov model, Artificial neural network, Computational neuroscience, Reinforcement learning, Deep learning, Soft computing, Automated

Guide

- Motivation for Knowledge Map
- Construction
- Manipulating the Scale
- Visualization
- Future Work



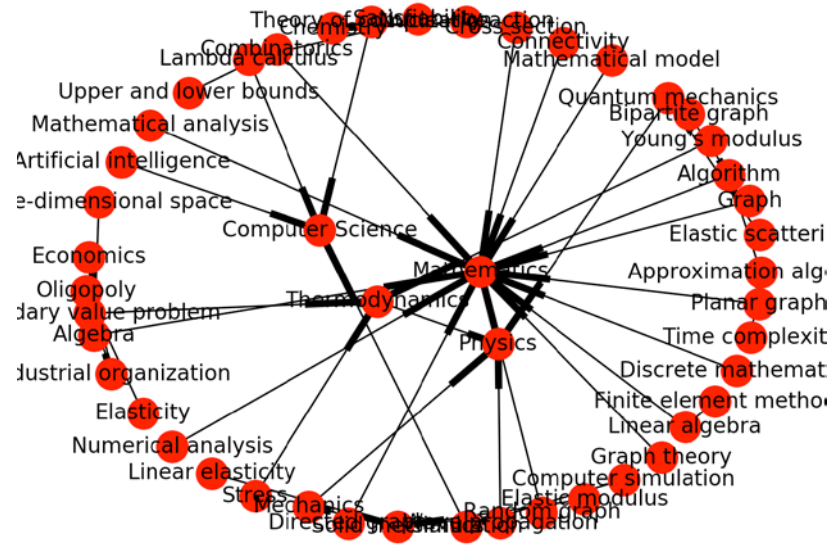
1.Extraction & Ranking

"Random vision"

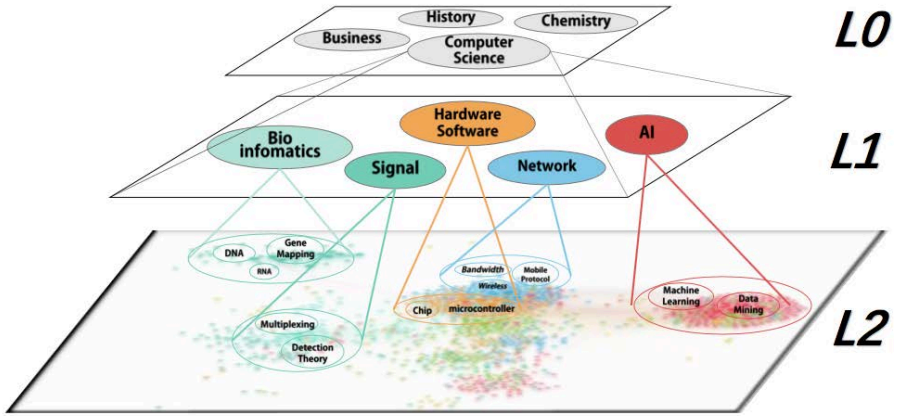
Searching engine

```
('0A0C148B', 'Randomized controlled trial', 'L3', 89959, 1377309)
('01E7DD16', 'Computer vision', 'L1', 63597, 519152)
('093C4716', 'Artificial intelligence', 'L1', 125469, 920411)
('007E3B49', 'Stochastic process', 'L2', 111274, 658167)
('00517AB5', 'Machine vision', 'L2', 19394, 112117)
('073947DC', 'Random walk', 'L3', 21130, 157513)
('076BC823', 'Image segmentation', 'L3', 56771, 471470)
('07477385', 'Random variable', 'L3', 38051, 230955)
('0322F49A', 'Feature extraction', 'L3', 93472, 786318)
('008FB4C8', 'Image processing', 'L2', 102494, 566598)
('07FC5C84', 'Vision', 'L3', 30238, 247463)
```

1.Frequency
2.Citation
3.Time
4.Conference

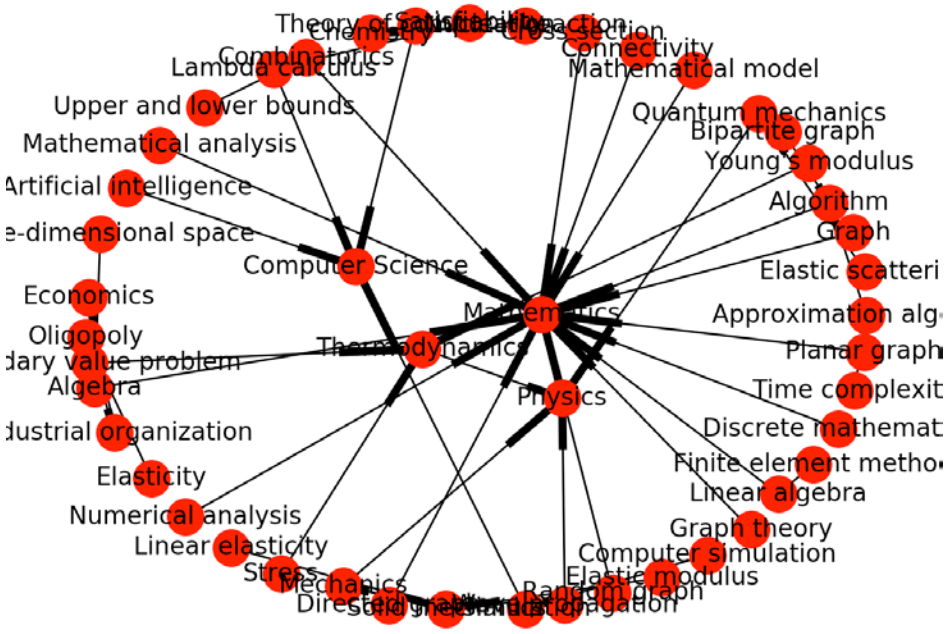


Hierarchy



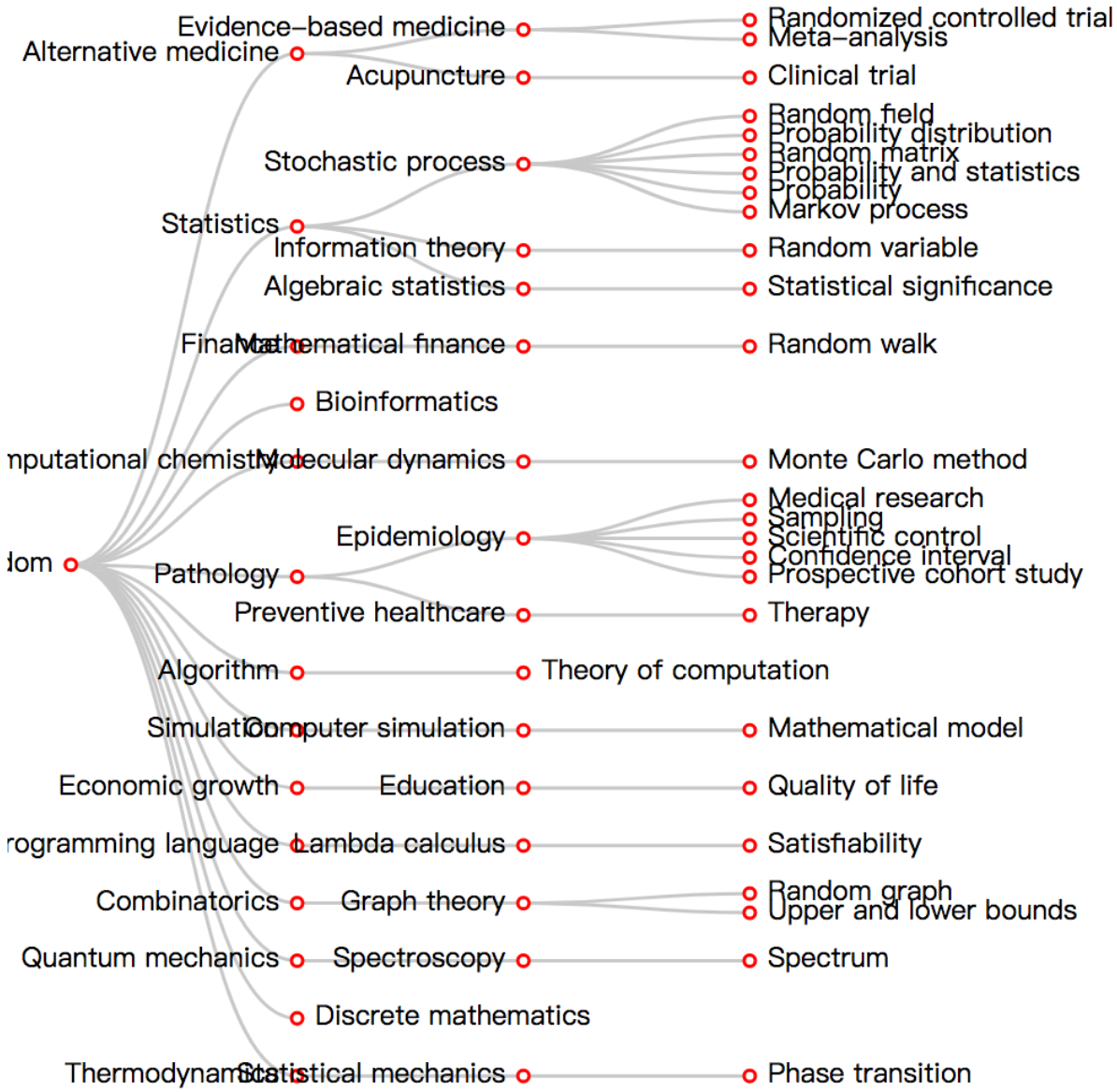
2. Construction

Topic Graph



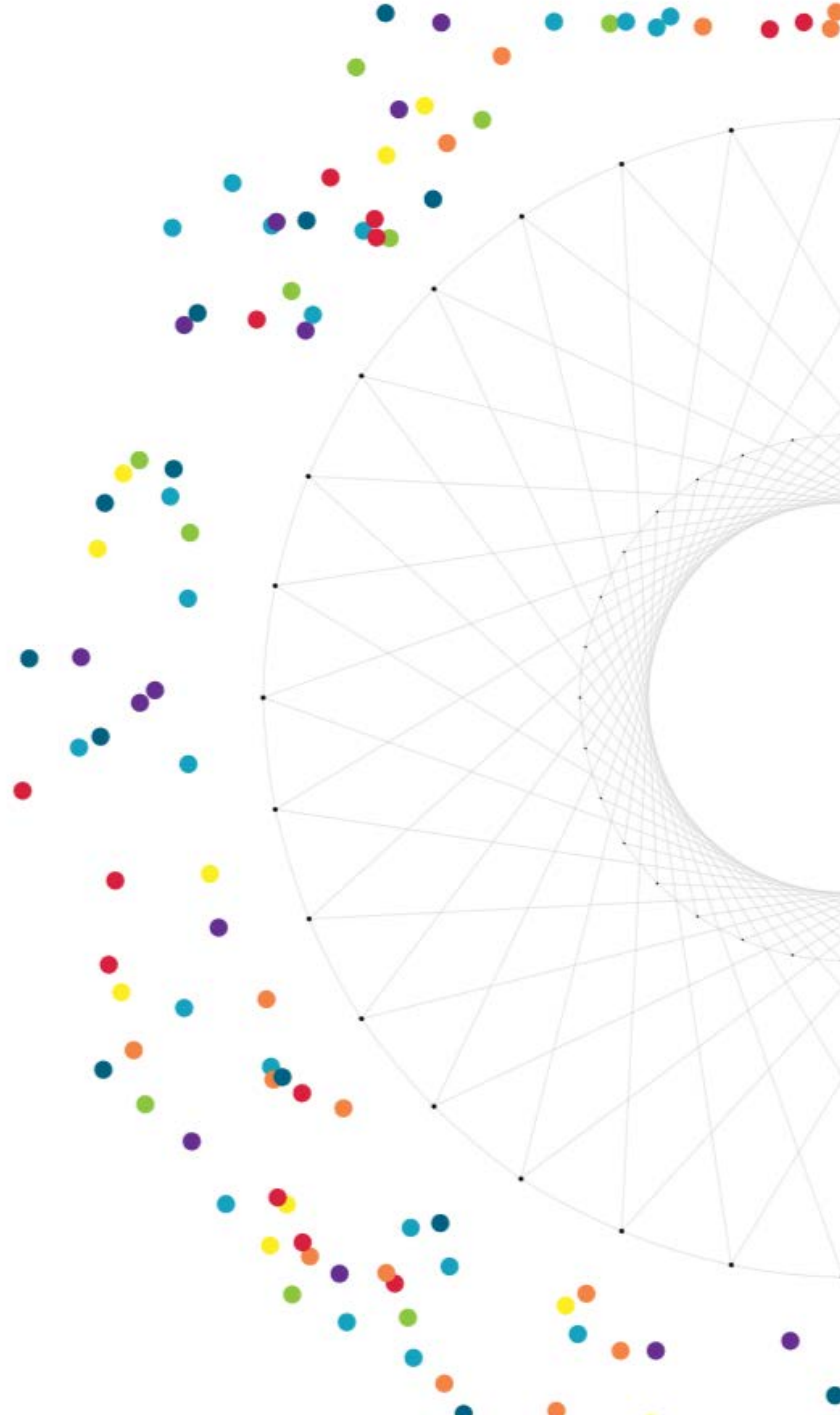
Transform

Topic Tree



Guide

- Motivation for Knowledge Map
- Construction
- Manipulating the Scale
- Visualization
- Future Work

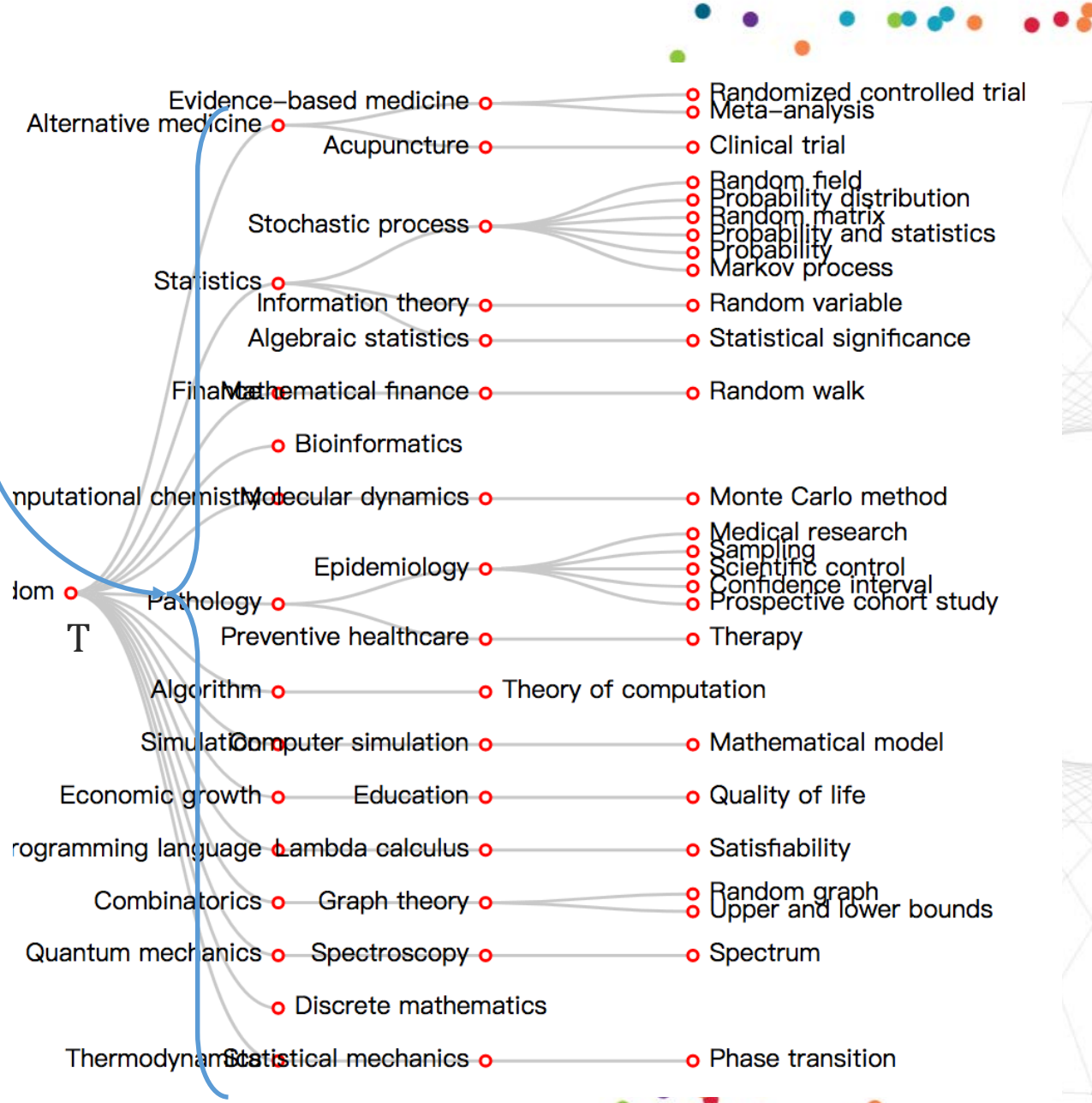


1.Filtering the topic preliminarily

For a knowledge tree T, if we delete the root node, we can obtain tree sequences $T_1, T_2, \cdots T_M$

- For root node in each subtree, the T_i will be delete the hierarchy of the root i is $'L_2', 'L_3'$
- Whose scale is too small and its hierarchy can't represent the knowledge.

New tree sequences $T_1, T_2, \cdots T_N$



2.Computing the scale

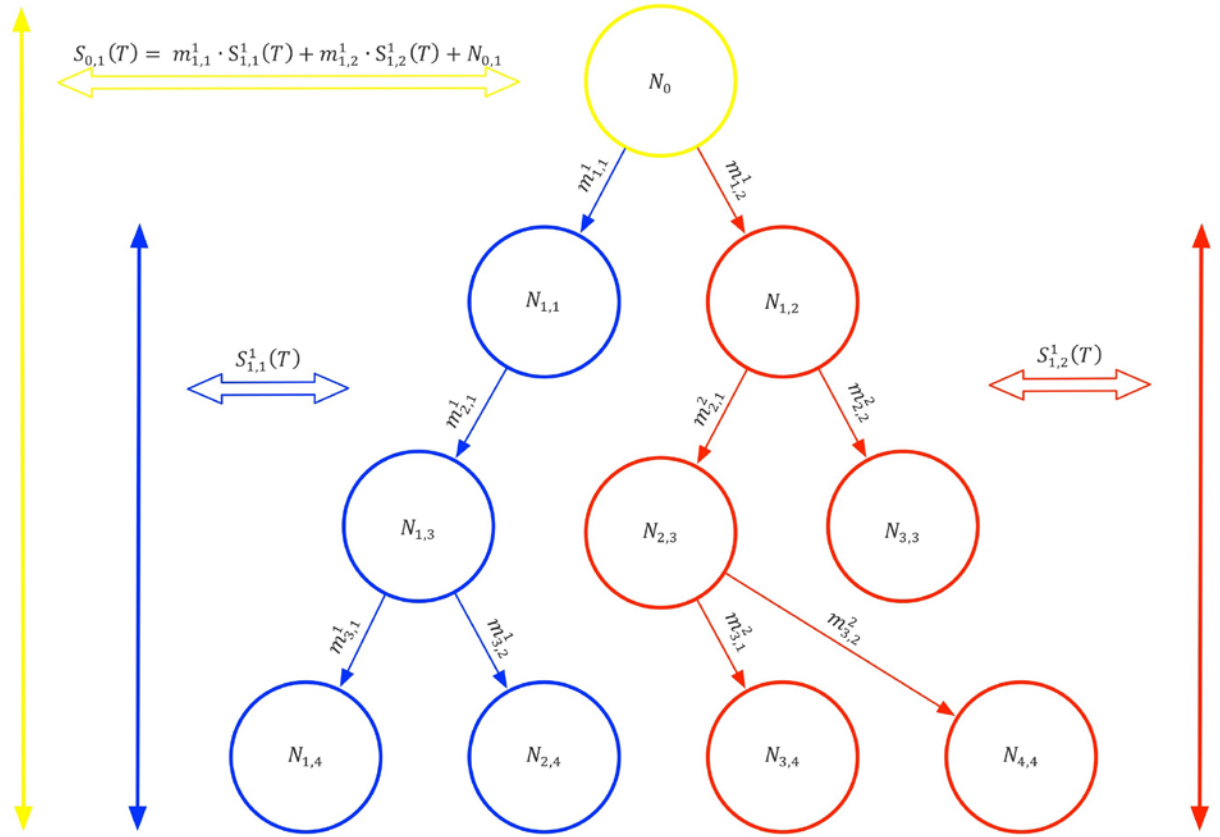
Recursive property
define S(T), M(T),N

$$S^{Q}_{j-1,M(j-1)}(T) = \sum_{i=1}^{n} m_{j,i}^{M'(j)} \cdot \mathrm{S}_{j,i}^{M'(j)}(T) + N_{j-1,M(j-1)}$$

$$N_{j,M(j)} = \left(\alpha_1 L_{j,M(j)} + \alpha_2 \frac{1}{D_{j,M(j)}}\right) \cdot \frac{F_j}{\sum_{i=1}^{U} F_i}$$

$$F_j = \frac{f_j - f_{min}}{f_{max} - f_{min}} \qquad m_{j,i}^{M'(j)} = \frac{\delta}{\delta_{i,j}}$$

Where $S^{Q}_{j-1,M(j-1)}(T)$ is the sum scale of node M in level $j-1$, node Q is the predecessor of node M, $N_{j-1,M(j-1)}$ is the scale of Node M in level $j-1$.



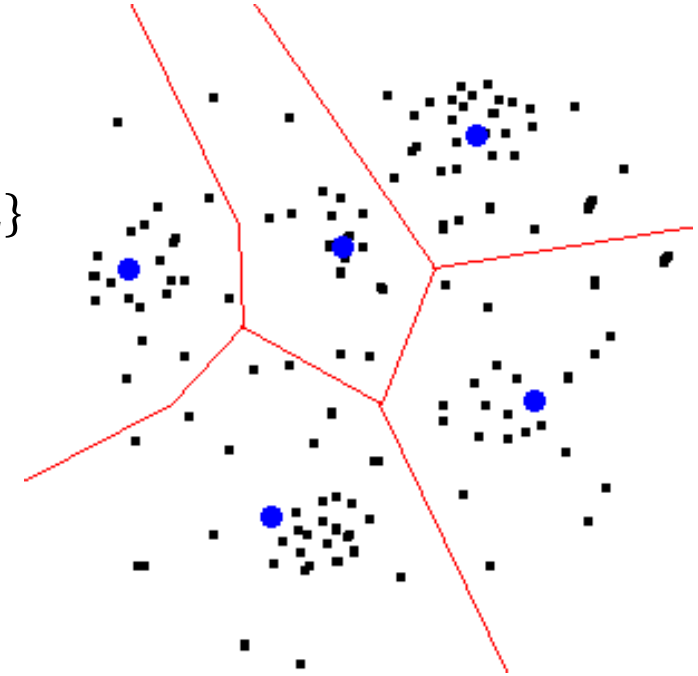
3.Clustering subtrees

K-means Algorithm:

1. Initialize assignment variables $\{r_{ik}\}$ and cluster centers $\{\boldsymbol{m}_1, \boldsymbol{m}_2, \dots, \boldsymbol{m}_c\}$
2. Repeat
   - **E-step**: with fixed $\{\boldsymbol{m}_1, \boldsymbol{m}_2, \dots, \boldsymbol{m}_c\}$, assign each sample to its nearest cluster center, i.e., $r_{ik} = 1, if\ k = arg\min_l ||\boldsymbol{x}_i - \boldsymbol{m}_l||$
   - **M-step**: with fixed $\{r_{ik}\}$, recalculate each cluster center, i.e.,

$$\boldsymbol{m}_k = \frac{\sum_i \boldsymbol{x}_i r_{ik}}{\sum_i r_{ik}}$$

3. Until Convergence



$$J_e \triangleq \sum_{i=1}^{c} \sum_{\mathbf{x} \in H_i} \|\mathbf{x} - \mathbf{m}_i\|^2$$

$[S_1^1, S_2^1, \cdots S_N^1]_1, [S_1^2, S_2^2, \cdots S_N^2]_2, \cdots [S_1^k, S_2^k, \cdots S_N^k]_k$ → Largest sequence $S_1', S_2', \cdots S_w'$

mapping

New tree sequences $T_1, T_2, \cdots T_w$

3.Recursion

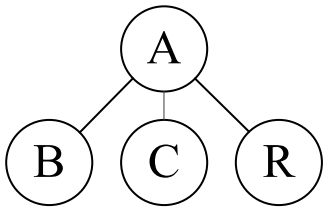
- Filtering the topic preliminarily
- Computing the scale
- Clustering subtrees

For the selected subtree sequences $T_1, T_2, \cdots T_w$, we delete each root of them, thus obtaining forest sequences.

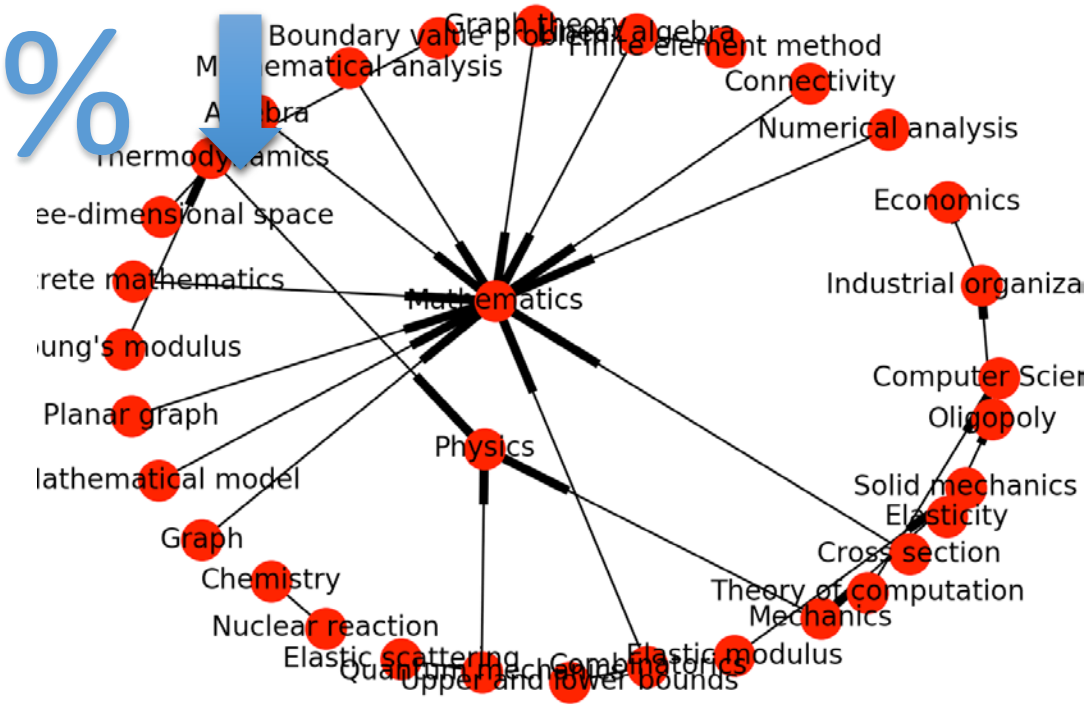
Repeating this proceed until the leaves

New tree sequences $T_1, T_2, \cdots T_w$

70%

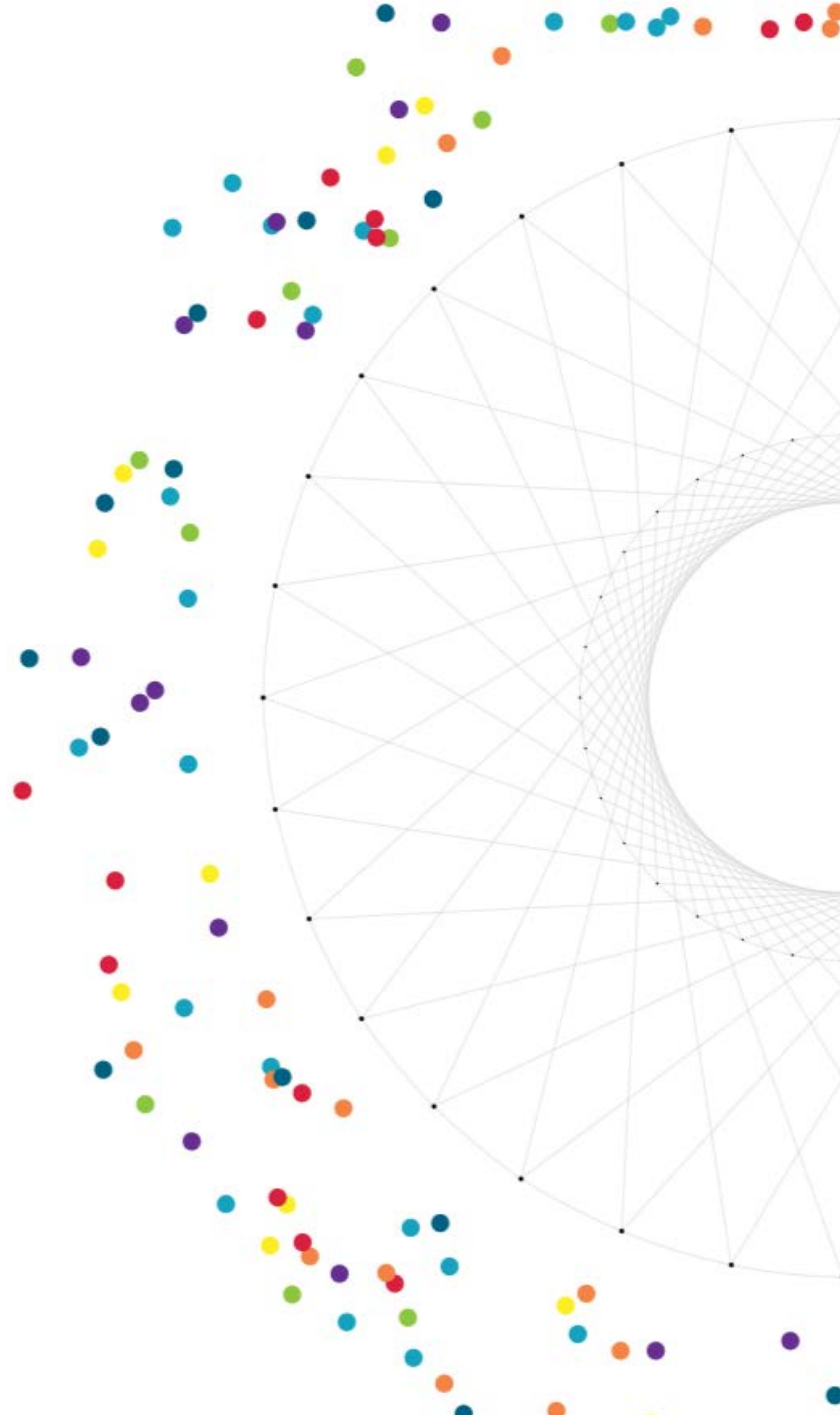


forest sequences F_i



Guide

- Motivation for Knowledge Map
- Construction
- Manipulating the Scale
- Visualization
- Future Work

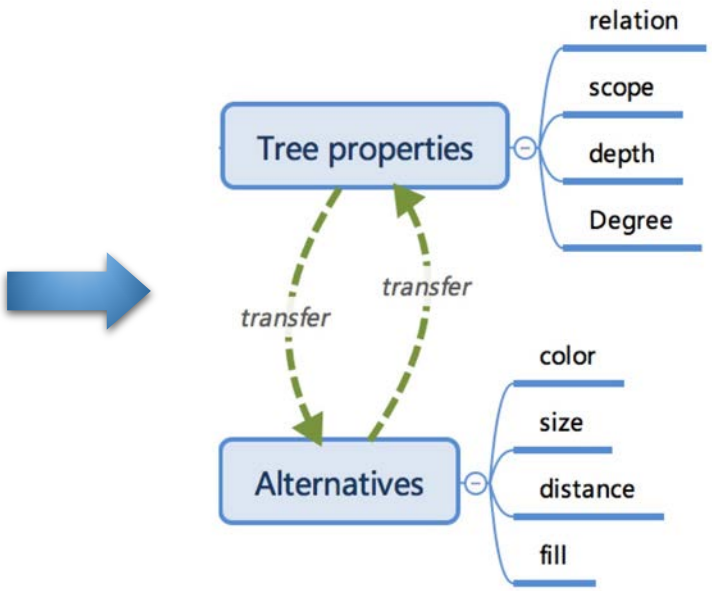


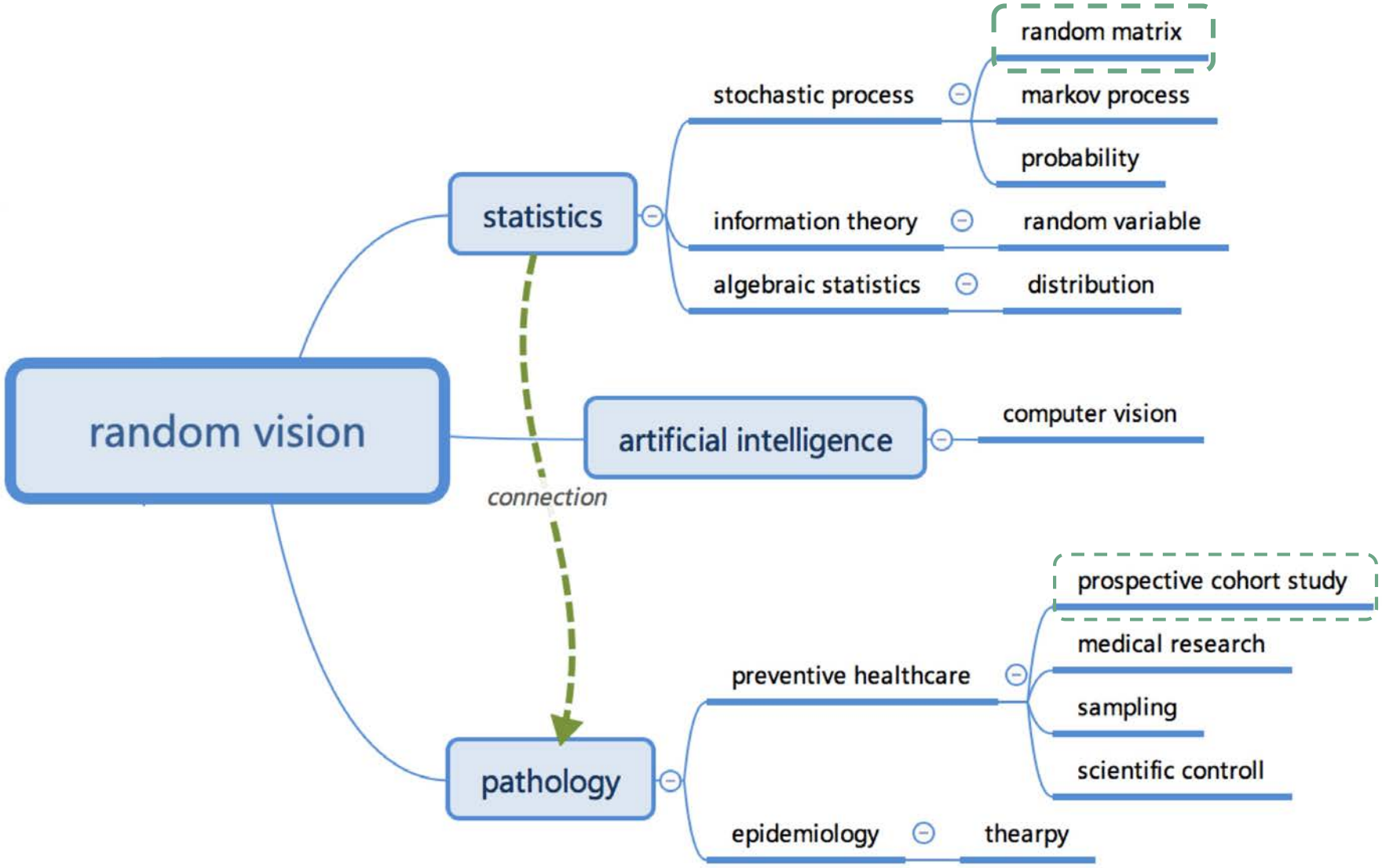
- Our goal

More informative

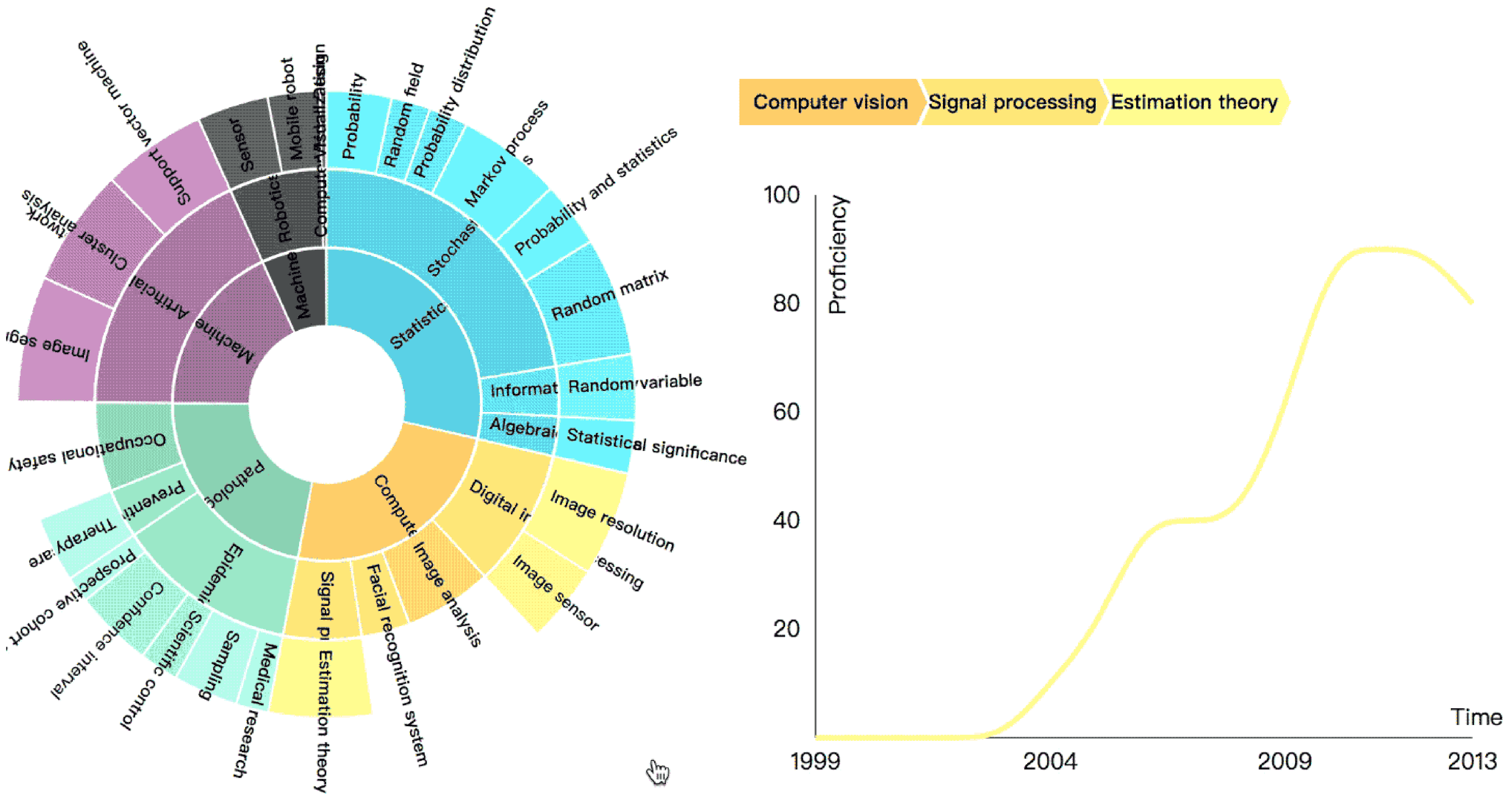
Let user know what topic he is searching

- Different variables
- Depth
- Size
- Color
- fill of the stripe and frame

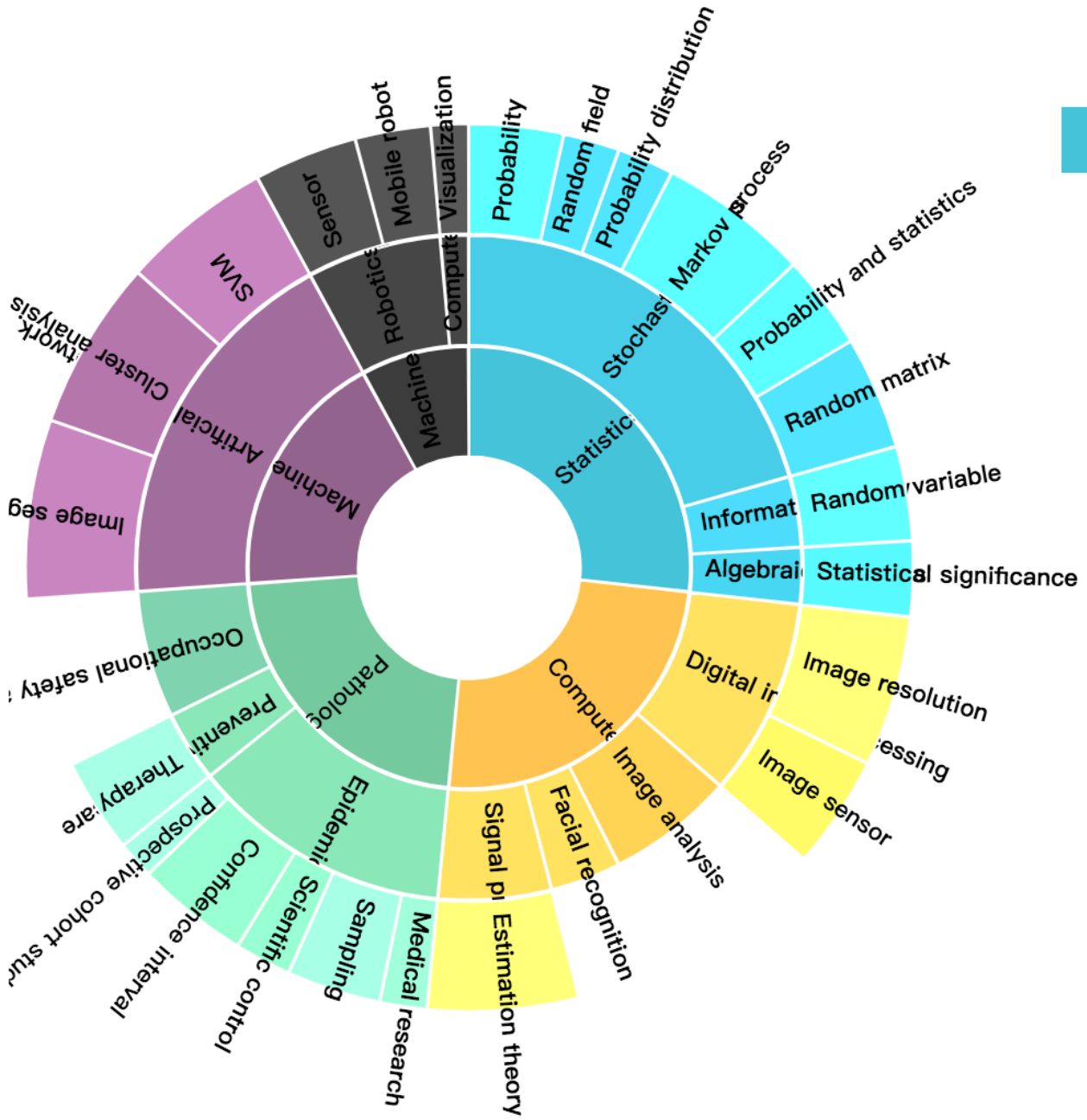


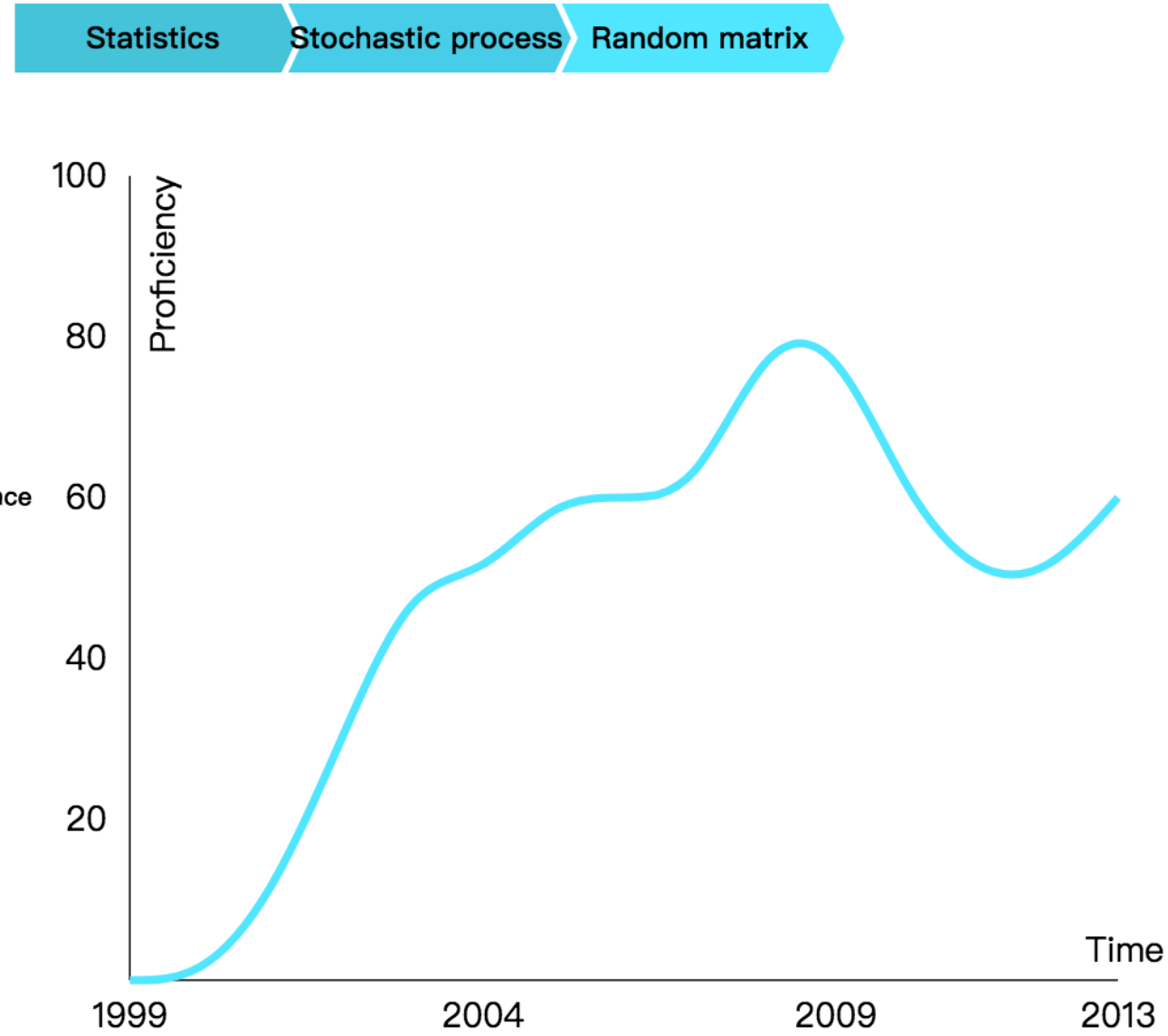

random vision
statistics
stochastic process
random matrix
markov process
probability
information theory
random variable
algebraic statistics
distribution
artificial intelligence
computer vision
connection
pathology
preventive healthcare
prospective cohort study
medical research
sampling
scientific controll
epidemiology
thearpy

D3.js Sunburst



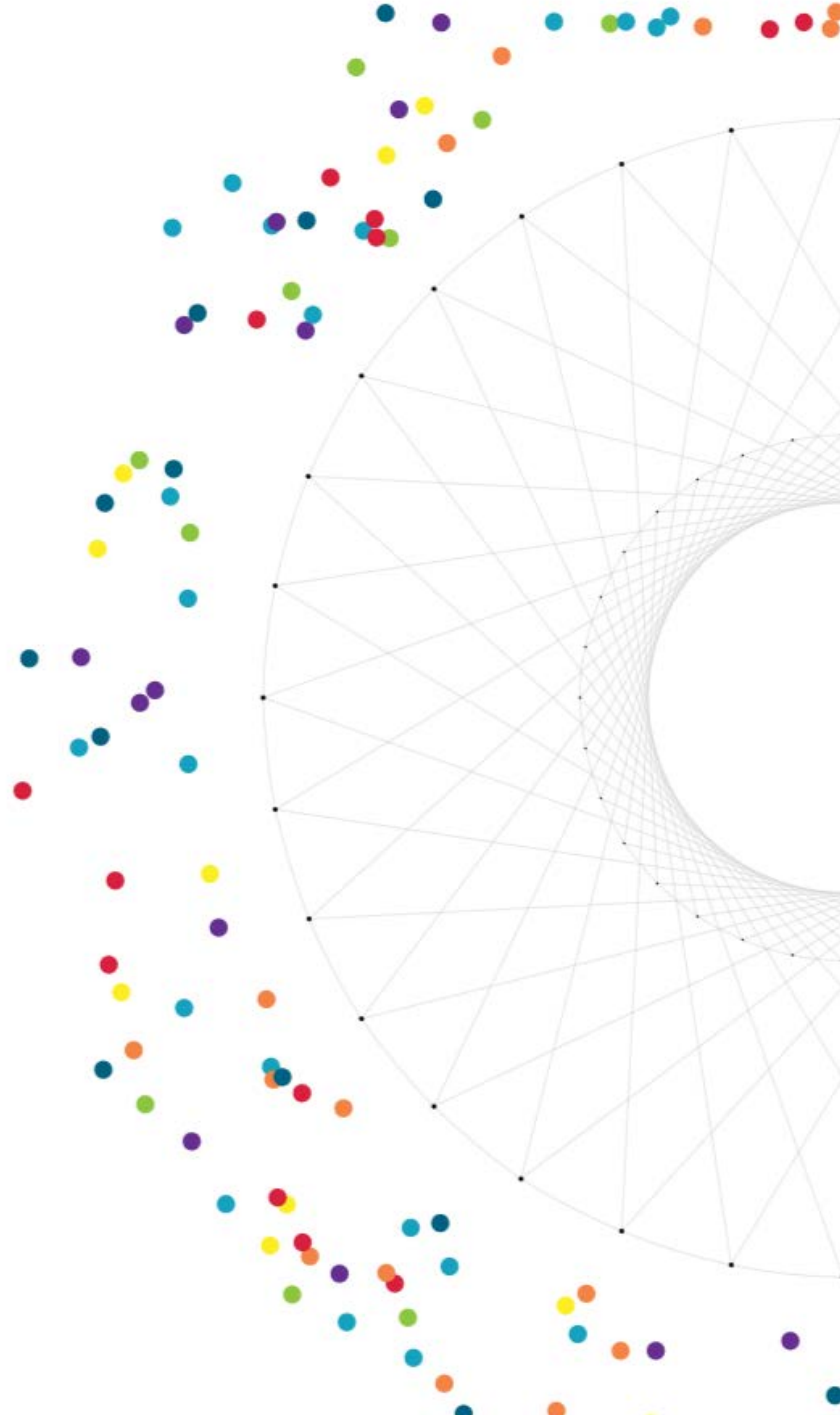
Computer vision
Signal processing
Estimation theory
Proficiency
Time
100
80
60
40
20
1999
2004
2009
2013
Statistic
Stochas
Informat
Algebrai
Probability
Random field
Probability distribution
Markov process
Probability and statistics
Random matrix
Random variable
Statistical significance
Compute
Digital i
Image resolution
Image sensor
Image analysis
Facial recognition system
Signal p
Estimation theory
Patholog
Epidemi
Preventi
Occupational safety
Therapy
Prospective cohort
Confidence interval
Scientific control
Sampling
Medical research
Machine
Artificial
Support vector machine
Cluster analysis
Image seg
Robotics
Sensor
Mobile robot
Comput


Probability
Random field
Probability distribution
Markov process
Probability and statistics
Random matrix
Random variable
Statistical significance
Image resolution
Image sensor
Image analysis
Facial recognition
Estimation theory
Medical research
Sampling
Scientific control
Confidence interval
Prospective cohort stur
Therapy are
Occupational safety
Image seg
Cluster analysis
SVM
Sensor
Mobile robot
Visualization
Statistic
Stochast
Machine
Robotics
Machine
Artificial
Pathology
Epidemi
Compute
Digital ir
Algebrai
Informat


Statistics
Stochastic process
Random matrix
Proficiency
100
80
60
40
20
Time
1999
2004
2009
2013

Guide

- Motivation for Knowledge Map
- Construction
- Manipulating the Scale
- Visualization
- Future Work

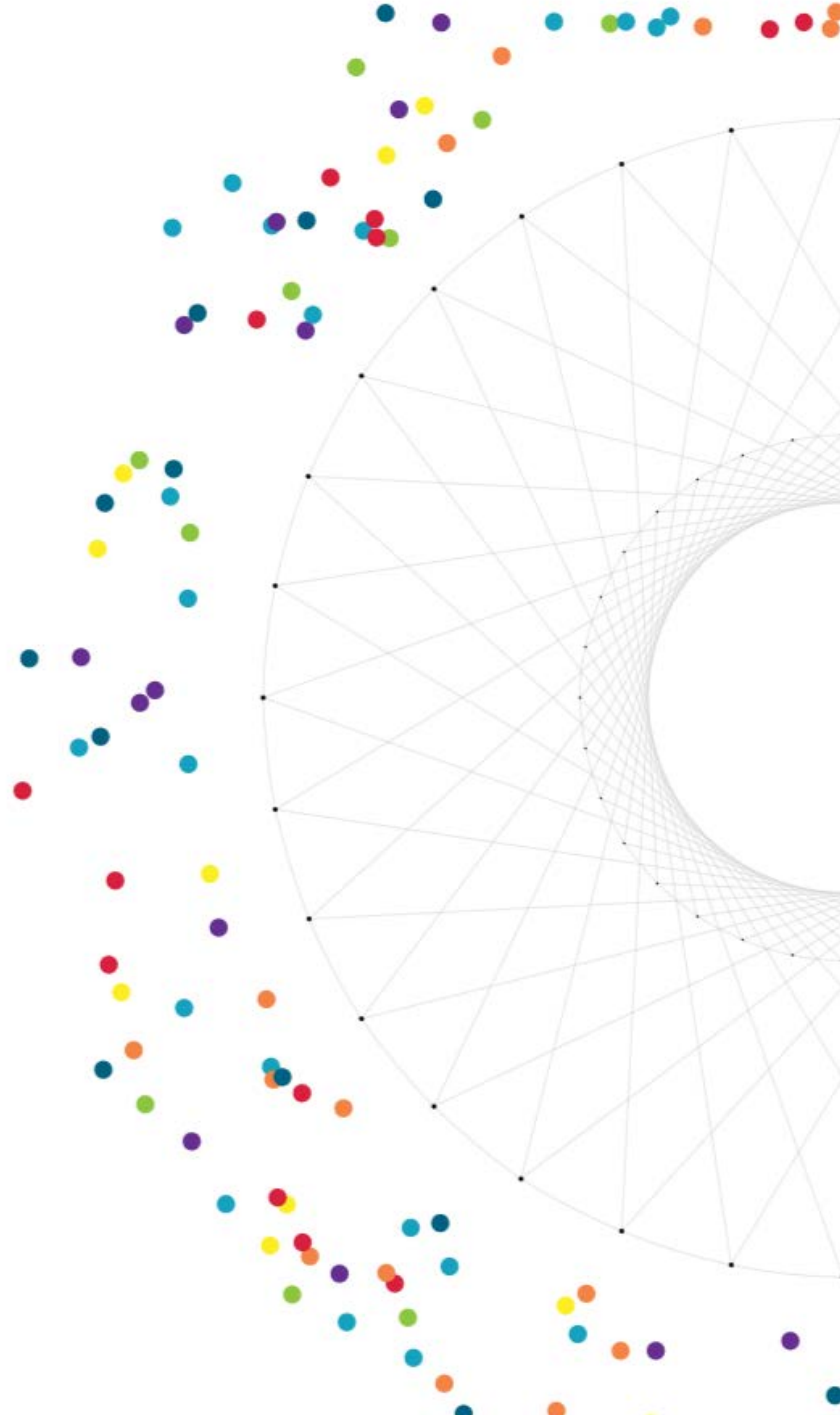


1. Theoretical Part

(i) More Clustering method to be tested
(ii) Seeking useful features to be added into the feature vector to accomplish the clustering.

2. Visualization

(i) Changing the chart form
(ii) More dynamic effect
(iii) More topics information




Thank you!