

# Name Disambiguation of Japanese Researchers: A Case Study with Statistics Research Community

Masao Takaku<sup>†</sup>, Yasumasa Baba<sup>††</sup>, Akiko Aizawa<sup>†††</sup>

<sup>†</sup> National Institute for Materials Science

<sup>††</sup> Research Organization of Information and Systems / The  
Institute of Statistical Mathematics

<sup>†††</sup> National Institute of Informatics

# Motivation

- There are many types of information on the Web
- Entity-based search
  - Categories for types of information
  - Events, place, person, etc.
- Searching for a person
  - Not sufficient to search Google with the name of the person
  - E.g. Expert search
  - Through the service of
    - SNS
    - Web search engine

**spock** BETA

Type in a name, keyword, or location to find people.

[Sign-up](#) to see where everyone you know is on the web

Steve Jobs | [Web \(71\)](#) | [Pictures \(21\)](#) | [Tags \(149\)](#) | [News \(250\)](#) | [Related People \(20\)](#)

**Steve Jobs** - male, 52 years old, [Cupertino, CA, United States](#)  
[Steven Paul Jobs](#) - [Steven P. Jobs](#) - [Steven P. Jobs](#) - [Add name](#)



[Add picture](#) | [Vote on pictures](#) | [All pictures](#)

Steven Paul Jobs (born February 24 1955) is the co-founder, chairman, and CEO of Apple and was the CEO of Pixar until its acquisition by Disney. He is currently the largest Disney shareholder and a member of the Board of Directors. He is considered a leading figure in the computer and entertainment industries and has contributed greatly to the development of the modern Silicon Valley entrepreneur.

Source: [Wikipedia](#)

**spock** BETA

Type in a name, keyword, or location to find people.

[Sign-up](#) to find people you know on the web

Yasuo Fukuda | [Web \(1\)](#) | [Pictures \(2\)](#) | [Tags \(4\)](#) | [News \(0\)](#)

**Yasuo Fukuda** - male, 72 years old  
[Add name](#)



[Add picture](#) | [Vote on picture](#)

Yasuo Fukuda is a Japanese politician who served as the 9th Prime Minister of Japan from September 16, 2006, to September 26, 2006. He was the first non-Liberal Democratic Party (LDP) member to become Prime Minister since 1955. Fukuda dropped out of the race for a second term.

Source: [Wikipedia](#)

[News](#) - [Add news](#) | [See all](#)

Be the first to add news to Yasuo Fukuda's search result.

[Websites](#) - [Add website](#) | [See all](#)

[Wikipedia](#)

**spock** BETA

Type in a name, keyword, or location to find people.

[Sign-up](#) to see where people you know are on the web

Bill Gates | [Web \(26\)](#) | [Pictures \(29\)](#) | [Tags \(107\)](#) | [News \(250\)](#) | [Related People \(20\)](#)

[Add trust](#) | [Send message](#) | [Widget](#)

**Bill Gates** - male, 52 years old, [Seattle, WA, United States](#)  
[William Henry Gates III](#) - [Add name](#)



[Add picture](#) | [Vote on pictures](#) | [All pictures](#)

William Henry Gates III (born October 28, 1955) is an American entrepreneur, philanthropist, and the chairman of Microsoft, the software company he founded with Paul Allen. During his career at Microsoft he has held the positions of CEO and chief software architect, and he remains the largest individual shareholder with more than 8% of the common stock. "Forbes" magazine's list of The World's Billionaires has ranked him as the richest person in the world since 1995, ...

Source: [Wikipedia](#)

[Tags](#) - [Add tag](#) | [Vote](#) | [See all](#)

[Microsoft](#) - [Microsoft founder](#) - [Bill & Melinda Gates Foundation](#) - [college dropout](#) - [geek](#) - [Windows](#) - [billionaire](#) - [philanthropist](#) - [Forbes World's Richest People](#) - [rich](#) - [software magnate](#) - [computer programmer](#) - [American billionaire](#) - [American](#) - [computer pioneer](#)

[Related People](#) - [Add people](#) | [See all](#)



[Sam Williams](#)  
friend



[Bill Clinton](#)  
friend



[Linus Torvalds](#)  
nemesis



[Bill Clinton](#)  
friend



[Bill Clinton](#)  
friend

[Linus Torvalds](#)  
nemesis

[Linus Torvalds](#)  
nemesis

[Login](#) | [Sign up](#)

**spock** BETA

Type in a name, keyword, or location to find people.

[Sign-up](#) to see where people you know are on the web

Shinzo Abe | [Web \(2\)](#) | [Pictures \(1\)](#) | [Tags \(14\)](#) | [News \(0\)](#) | [Related People \(1\)](#)

[Claim](#) | [Widget](#)

**Shinzo Abe** - male, 53 years old  
[Add name](#)



[Add picture](#) | [Vote on pictures](#) | [All pictures](#)

Shinzo Abe (born September 21 1954) is the current Prime Minister of Japan, elected by a special session of the National Diet on September 26 2006. He is Japan's youngest post-World War II prime minister and the first born after the war. Abe was born into a political family, and studied political science in Japan, and had studied in the United States. He worked in the private sector until 1982 when he began work in several government jobs.

Source: [Wikipedia](#)

[News](#) - [Add news](#) | [See all](#)

Be the first to add news to Shinzo Abe's search result.

[Websites](#) - [Add website](#) | [See all](#)

[Wikipedia](#)

The biggest multilingual free-content encyclopedia on the Internet. Over 7 million articles in over 200 languages, and still growing.

[Tags](#) - [Add tag](#) | [Vote](#) | [See all](#)

[Government minister of Japan](#) - [from Yamaguchi Prefecture](#) - [National Diet](#) - [Chief Cabinet Secretary](#) - [Liberal Democratic Party](#) - [Yamaguchi Prefecture](#) - [Diet of Japan](#) - [National Leader](#) - [Japanese](#) - [prime minister of Japan](#) - [north Korea](#) - [minister](#) - [political science](#) - [World War II](#)

[Related People](#) - [Add people](#) | [See all](#)



[Akie Abe](#)  
wife

[Quotes](#) - [Add quote](#) | [See all](#)

Be the first to add a quote to Shinzo Abe's search result.

[Tags](#) - [Add tag](#) | [Vote](#) | [See all](#)

[Wikipedia](#) - [Wikipedia founder](#) - [Wikia](#) - [Wikipedia person](#) - [Wikimedia Foundation](#) - [Internet personality](#) - [business owner](#) - [Search Wikia](#) - [co-founder of Wikipedia](#) - [entrepreneur](#) - [American entrepreneur](#) - [Business 2.0 The 50 Who Matter Now](#) - [Jimmy Donal "Jimbo" Wales](#) - [Jimbo Wales](#) - [Internet celebrity](#)

[Related People](#) - [Add people](#) | [See all](#)



[Larry Sanger](#)  
co-founder of Wikipedia



[Amber](#)  
Myspace friend



[Ted](#)  
Myspace friend



[Irene McGee](#)  
Myspace friend



[Ted](#)  
Myspace friend



[Irene McGee](#)  
Myspace friend

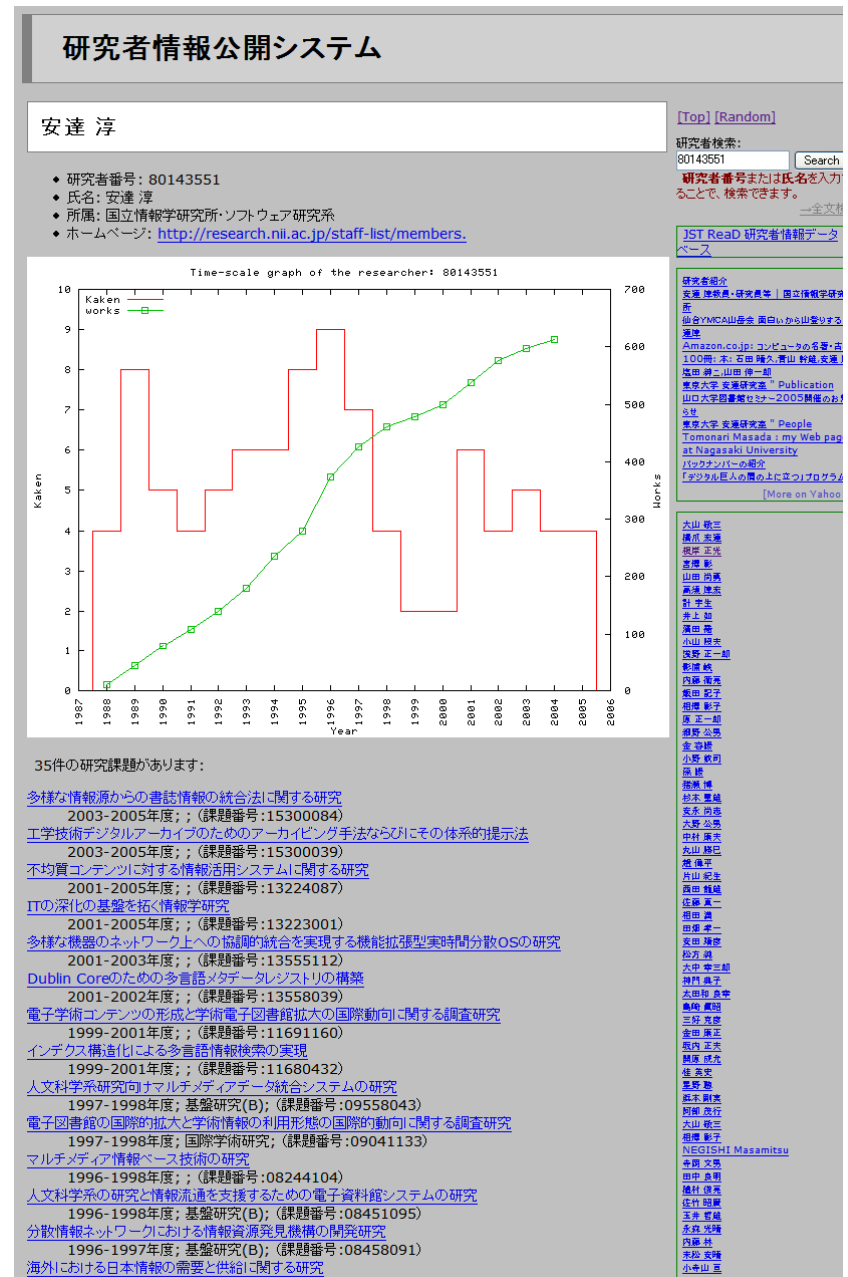
http://www.spock.com

# Objective

- Name identification
  - We integrated persons from different resources.
  - For identifying a different person with the same name, it is necessary to use several resources.
  - Estimating the costs
- Researcher's information:
  - KAKEN Researcher ID (for researchers in Japan)  
(→ We developed “Researcher Information Server”).)
- We have made a dataset for statistics related researchers.
  - We intend to help evaluating and developing identification data and search system for researchers.
  - Data: Member lists of three academic associates for statistics related field in Japan

# Researcher Information Server

- Data taken from KAKEN DB
  - # of projects (Funded by KAKEN): 247,745
    - Annual reports of the projects from 1989 to 2004
  - # of researchers (project leader and co-researchers): 133,067
    - Individual researcher information
- Basic information for researcher
  - Researcher name, affiliation name, and position (job title).
- Visualization
  - Timeline: # of KAKEN projects and publications
  - Co-researcher network:
    - Past KAKEN project's co-researcher
    - Using Google Maps



# Statistics related field researchers (in three academic societies)

- We used member lists of academic societies (total: 3,031)
  - The Japan Statistical Society (2003): 1,545
  - The Behaviormetric Society of Japan (2005): 1,070
  - The Japanese Society of Computational Statistics (2002): 416
- Registered information: name, affiliation name, job title, address, zip-code
  - Only name and affiliation name were used.

# Identification (Step 1)

## Name matching

- Matching the researcher with the same name researcher in Research Information Server (KAKEN-DB)
  - A total of 1,400 candidates:
    - The Japan Statistical Society : 697 (45%)
    - The Behaviormetric Society of Japan : 519 (49%)
    - The Japanese Society of Computational Statistics: 184 (44%)
  - After removing duplicates: 1,307 candidates
    - Candidates of having exactly the same name and affiliation were considered as an identical researcher.
- We manually identify the following data to KAKEN researchers ID.

# Identification (Step 2)

## Manual judgments

- Adding a KAKEN ID for each member manually.

Candidates		Results	
Members of societies	KAKEN IDs	Members of societies	KAKEN IDs
BSJ0004	20024581	BSJ0004	Different 20024581
BSJ0005	50305313	BSJ0005	Identical 50305313
JSS0012	70303047 09246528	JSS0012	Different 70303047 09246528
BSJ0007	90184332	BSJ0007	Identical 90184332
JSS0014	90184332	JSS0014	Identical 90184332
JSS0015	90132696	JSS0015	Different 90132696

A member of multiple societies

Judged by manual



# Judgment system for identification



- ◆ [redacted] 京都学園大学人間文化学部人間関係学科 (BSJ0226)

[redacted] 60240628 ☒ 同一人物 ☐ 別人

同定確定

- ◆ [redacted] 一橋大学経済学部 (JSS0400)

[redacted] 70154838 ☒ 同一人物 ☐ 別人

同定確定

- ◆ [redacted] 豊田工業高等専門学校 (BSJ0229)

[redacted] 40112833 ☐ 同一人物 ☒ 別人

[redacted] 90216705 ☒ 同一人物 ☐ 別人

同定確定

- ◆ [redacted] 東京工業大学大学院情報理工学研究科数学・計算科学専攻 (JSS0402)

[redacted] 30154876 ☐ 同一人物 ☒ 別人

[redacted] 40312300 ☐ 同一人物 ☒ 別人

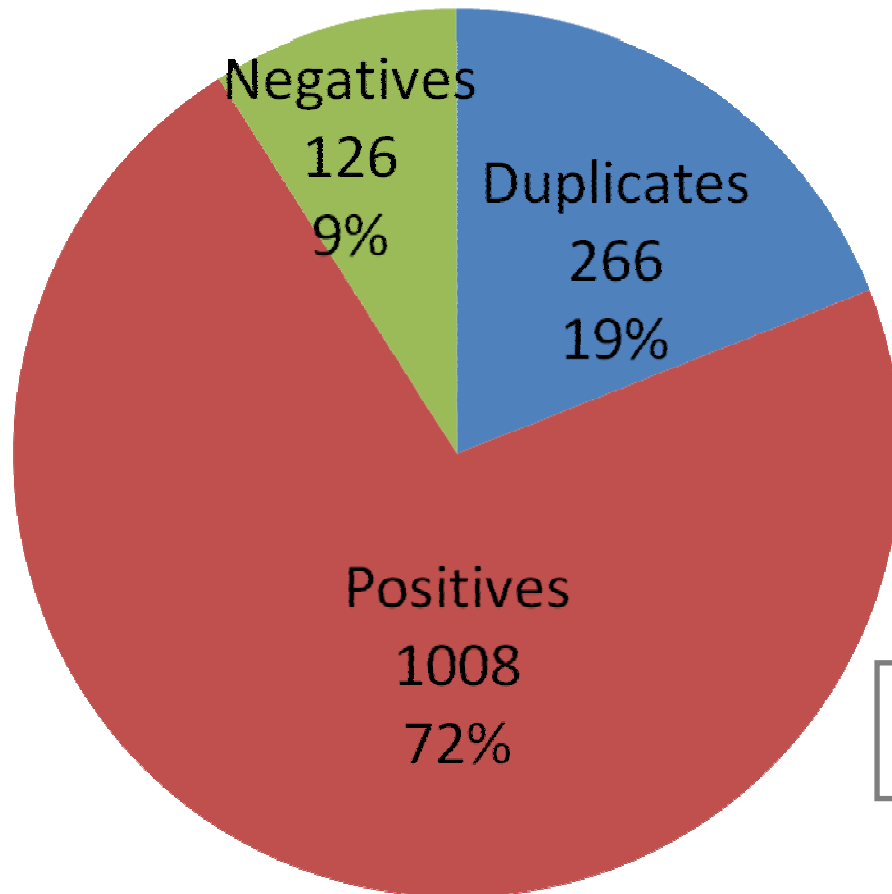
[redacted] 70016153 ☒ 同一人物 ☐ 別人

同定確定

Links to the  
"Researcher  
Information Server"

# Results

- 266 candidates were marked as “duplicates” (members of multiple societies)
- 1008 candidates were identified with a researcher within KAKEN-DB.



Marked as  
duplicates

Members of societies		KAKEN IDs
BSJ0004	Different	20024581
BSJ0005	Identical	50305313
JSS0012	Different	70303047 09246528
BSJ0007		
JSS0014	Identical	90184332
JSS0015	Identical	90132696

# Researchers having the same name

- (a) Within KAKEN-DB (for positives)
    - A member's name matched with KAKEN ID, and a different researcher having the same names within KAKEN-DB.
- 105 candidates (10%)

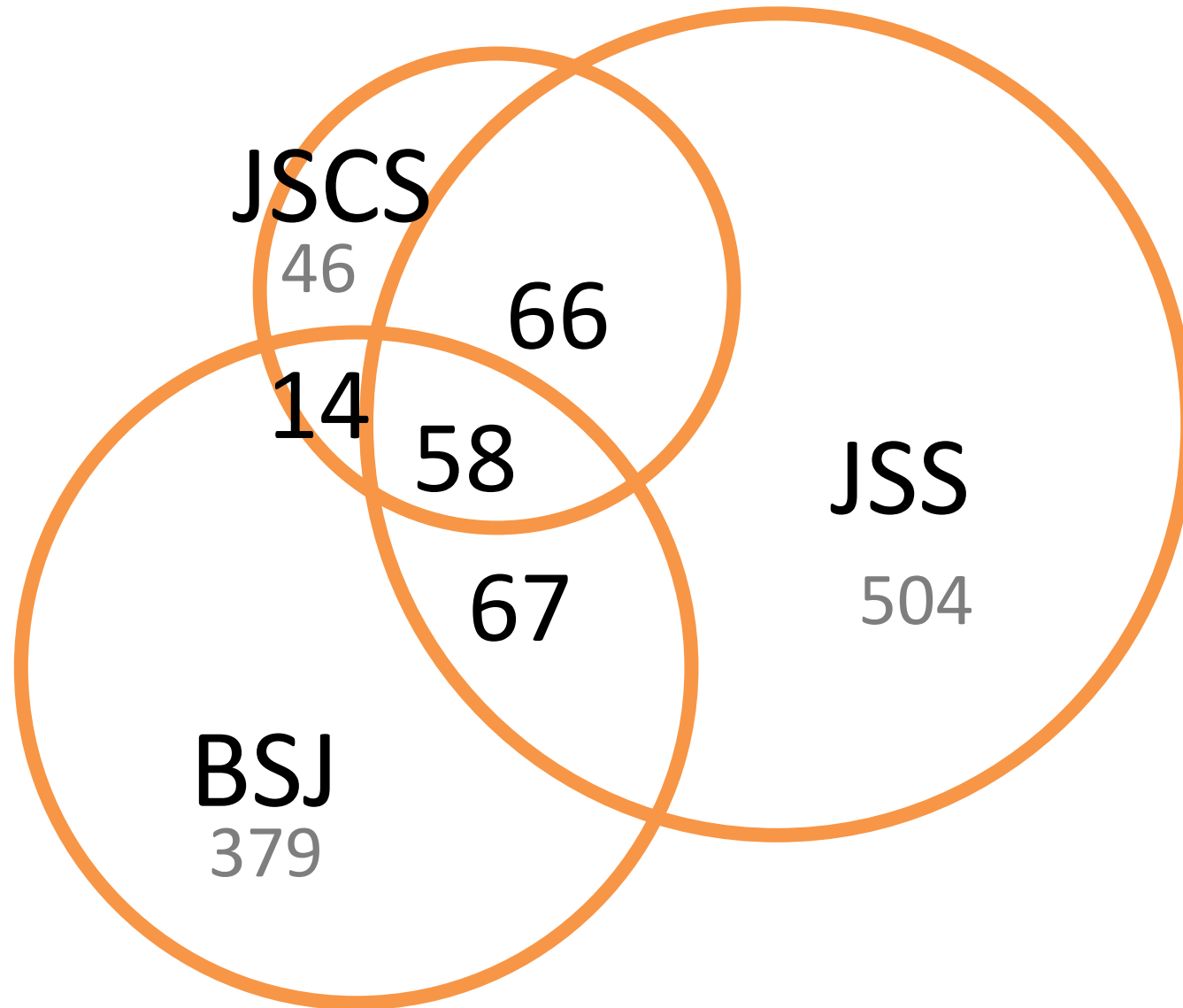
Members of societies	KAKEN IDs
JSS0035	Identical 0308679
	Different 241595

- (b) Concerning outside of KAKEN-DB (for negatives)
    - A member's name matched with KAKEN ID, but that represented a different researchers.
- 126 candidates

Members of societies	KAKEN IDs
BSJ0004	Different 024581
JSS0012	Different 70303047
	09246528

- (a) + (b) → 231 candidates (20%)

# Duplicates (overlaps among the Japanese academic societies in Statistics-related field)



# Discussion: Criteria for identification

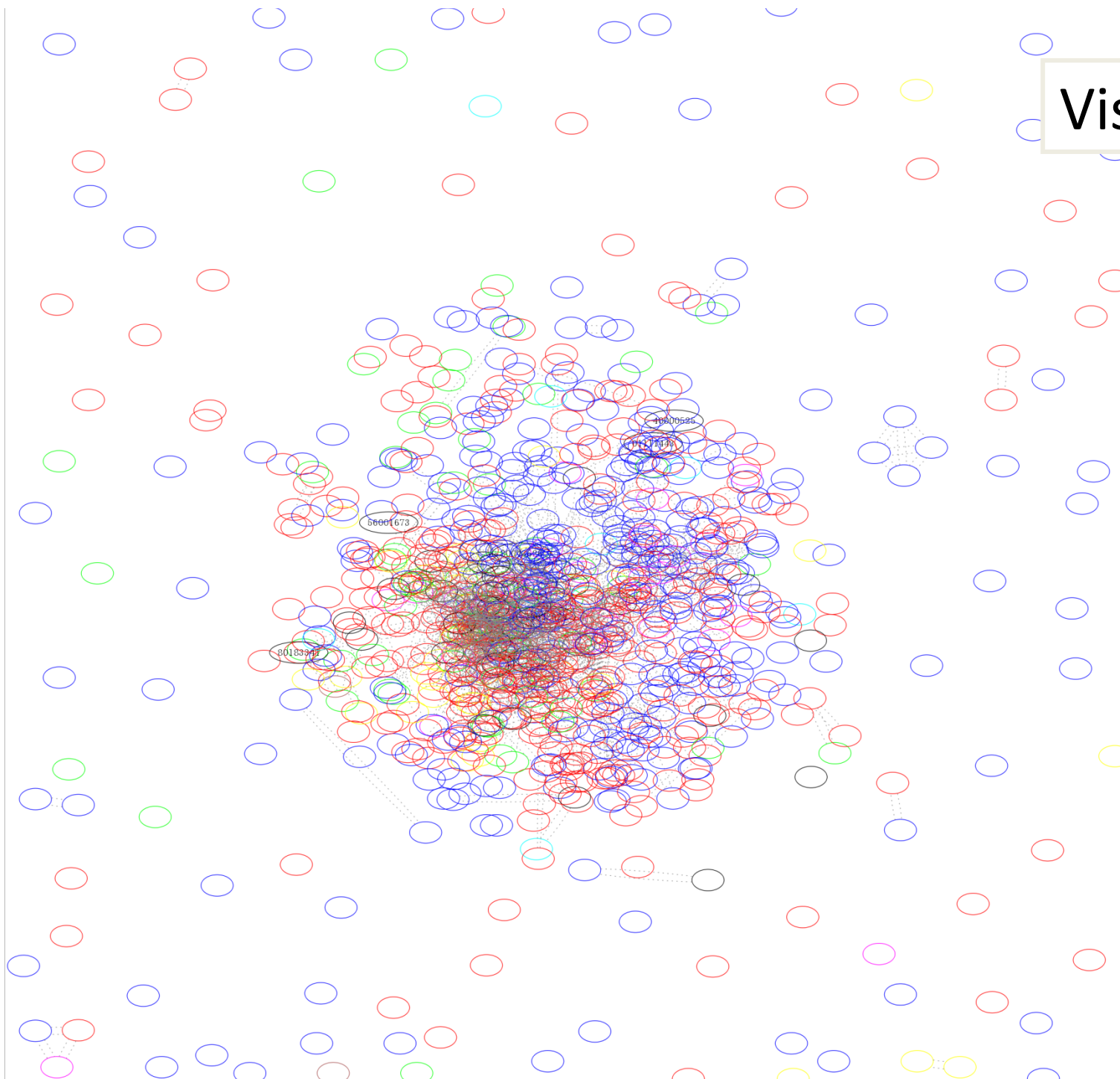
- How to determine two researchers are identical or not?
  - Researcher Information Server
    - Research area
    - Affiliations
    - Timeline based info.
  - Other resources:
    - JST ReaD (researcher directory)
    - On the Web
      - Organization website the researcher affiliates
      - Homepage provided by the researcher him/herself
  - Biography information provided as author introduction in his/her publications
- For statistics related field, left-side info are useful.
  - (Of course) His/her research fields are related to Statistics. It is a main hints for
- But, Sometime there is a difficult case to determine
  - By lack of information
  - It depends on the available resources and costs .
    - Need to clarify to utilize after the identification
- (Another possibility) Saving criteria and process for identification for the further analysis (afterwards).

# Discussion: Visualizations

- Visualization of co-researchers network
  - Provides summary of researcher's community.
  - Connects researchers with having joint-research in the past.
  - Shows one more hop into the other researchers.
- Applications (It will be helpful for ...)
  - Judgment of Identification
  - Summarization of researchers network (community)

# Visualization

- JSS
- BSJ
- JSCS





# Conclusions

- Identification of Japanese researchers in the field of Statistics
  - Identification between academic societies and KAKEN-DB
  - Identification of about 3000 Statistics-related researchers with about 1,300,000 database leads to about 1000 identified persons.
- Future works
  - Semi-automatic process for identification (by using machine learning techniques)
  - Another dataset and resources
    - Other databases, transcriptions, and so on.
  - Comparison with other scientific fields
  - Represent a domain knowledge from researcher community
  - Entity-based search engine by identification results