



中国科学院国家科学图书馆

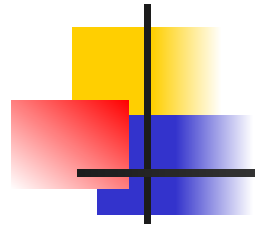
National Science Library, Chinese Academy of Sciences

Profiling Science and Innovation Policies of Obama Administration by Mining OSTP Web Resources



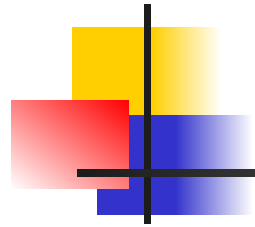
Zhang Zhixiong, Liu Jianhua, Qian Li, Guo Hongmei,
Zou Yimin, Xie Jing

National Science Library, Chinese Academy of Science, China
Sep 25, 2013



Outline

- 1. Introduction
- 2. Data sources
- 3. Methods
- 4. Results



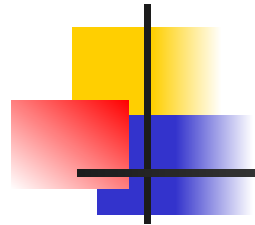
Outline

- **1. Introduction**
- **2. Data sources**
- **3. Methods**
- **4. Results**

1. Introduction

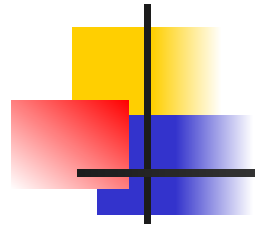


- **The Office of Science and Technology Policy (OSTP)**
 - one of the most important targeted institutes for information analyst in CAS
 - bears the responsibility to develop and implement science and technology policies for the United States
 - keep monitoring for about 4 years



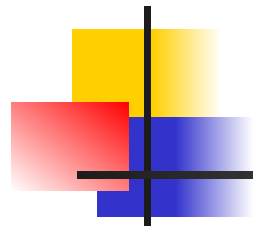
1. Introduction

- Aim of this paper
 - Can we deeply use the resources from OSTP web site ?
 - Try bring forth a sound method
 - mine the web resources from OSTP Web site to profile the science and innovation policy taken by Obama administration.



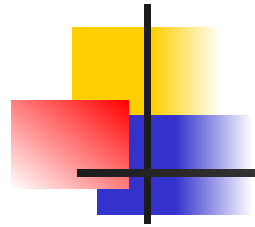
Outline

- 1. Introduction
- **2. Data Sources**
- 3. Methods
- 4. Results



2. Data Sources

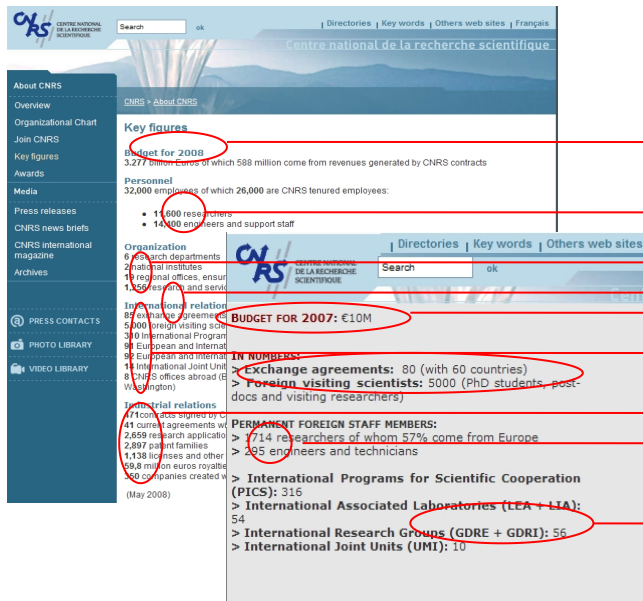
- **Web resource from OSTP Web site in our database**
 - **URL**
 - <http://www.whitehouse.gov/administration/eop/ostp>
 - **Crawled Directories**
 - **OSTP Pressroom**
 - **NSTC Documents & Reports**
 - **PCAST Documents & Reports**
 - **OSTP Blog**
 - **Documents, Speeches, & Reports**
 - **R&D Budgets**
 - **Homepage**
 - **582 web resources**
 - **150 PDF files and 432 HTML files**
 - **Published from Feb 17, 2009 to Apr. 12, 2013.**



Outline

- 1. Introduction
- 2. Data sources
- **3. Methods**
- 4. Results

3. Methods

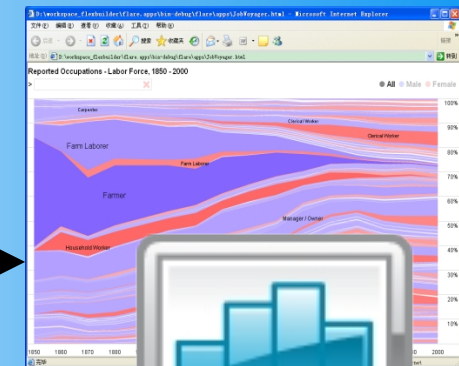


basic idea

Automatic Extraction

	Budget	Researchers	Engineers
2007	€ 1 B	1417	295
2008	€ 3.277B	11,600	14,400
		...	

Data Mining And In-depth Analysis



Research Profiling



3. Methods

- Knowledge Extraction

- analyze the PDF files and web pages
- extract
 - key terms
 - objects
 - named entities such as **important strategies, major projects, significant research reports, R&D spending** etc
- turn the free text into a computable knowledge base

3. Methods

科技政策与战略动态监测 - Windows Internet Explorer

http://policymonitor.las.ac.cn/PolicyMonitor/home/semanticannotation.htm?id=151319@parentPageId=

文件(F) 编辑(E) 查看(V) 收藏夹(A) 工具(T) 帮助(H)

收藏夹 科技政策与战略动态监测

您当前的位置： 首页 > 网页快照

资源原文

Remarks by President Obama on **Space Exploration** in the 21st Century

The White House
Office of the Press Secretary
For Immediate Release
April 15, 2010
Remarks by the President on **Space Exploration** in the 21st Century.
John F. Kennedy Space Center Merritt Island, Florida
2:55 P.M. EDT

THE PRESIDENT: Thank you, everybody. Thank you. (Applause.) Thank you so much. Thank you, everybody. Please have a seat. Thank you.

I want to thank **Senator Bill Nelson** and NASA **Administrator Charlie Bolden** for their extraordinary leadership. I want to recognize **Dr. Buzz Aldrin** as well, who's in the house. (Applause.) Four decades ago, Buzz became a legend. But in the four decades since he's also been one of America's leading visionaries and authorities on human space flight.

Few people -- **present company** excluded -- can claim the expertise of Buzz and Bill and Charlie when it comes to space exploration. I have to say that few people are as singularly unimpressed by Air Force One as those three. (Laughter.) Sure, it's comfortable, but it can't even reach low Earth orbit. And that obviously is in striking contrast to the Falcon 9 rocket we just saw on the launch pad, which will be tested for the very first time in the coming weeks.

A couple of other acknowledgments I want to make. We've got Congresswoman **Sheila Jackson**

原始页面快照

涉及对象

科研机构

Kennedy Space Center (3)

人物

John F. Kennedy (2)
Senator Bill Nelson (1)
Administrator Charlie Bolden (1)
Dr. Buzz Aldrin (1)
Sheila Jackson Lee (1)
John P. Holdren (1)
Suzanne Kosmas (1)
President Eisenhower (1)
John Glenn (1)

涉及术语

Space Exploration (6)
space shuttle (2)
previous program (2)
larger purpose (2)
essential part (2)
greatest achievement (2)
reach space (2)

Institutes

Persons

Terms



3.Methods

- **Count & Clustering**
 - Count and identify major players (Terms & Objects) in the resources
 - Backlink count to identify the most influential reports released from the OSTP Web site
 - Perform clustering based on extracted terms and objects



3.Method

■ Network Analysis

- Base on the co-occurrence of terms and objects in a file, a paragraph or sentence
- constructure graphs between files, terms and objects
- perform network analysis such as
 - Terms relationship in the reports
 - term, term, same report
 - Objects relationship in those reports
 - object, object, same report
 - Relationship between those reports
 - report, report, same term or same objects
- dipicture the relationship between major players



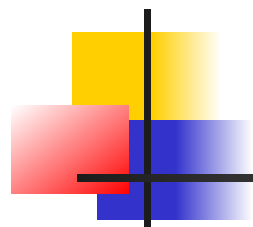
3. Methods

- Visualization & Profiling
 - Easy charts (line, bar, area chart)
 - Graph visualization
 - Timeline
 - Geomapping



3. Methods

- Tools
 - Knowledge extraction
 - GATE, Stanford parser...
 - Link analysis
 - Majestic SEO
 - Network analysis
 - Gephi, Pajek
 - Visualization
 - D3, Sigma.js, Flex chart, Simile Widgets Timeline component
 -



Outline

- 1. Introduction
- 2. Data sources
- 3. Methods
- 4. Results



4. Results

- 4.1 The Key Reports Analysis
- 4.2 The Policy Focuses Analysis
- 4.3 Inter-relationship Analysis



4.1 The Key Reports Analysis

- Identify the most important and influential reports released from the OSTP Web site
- Analyze the topics of those reports
- Perform clustering analysis to identify the relationship between those reports.

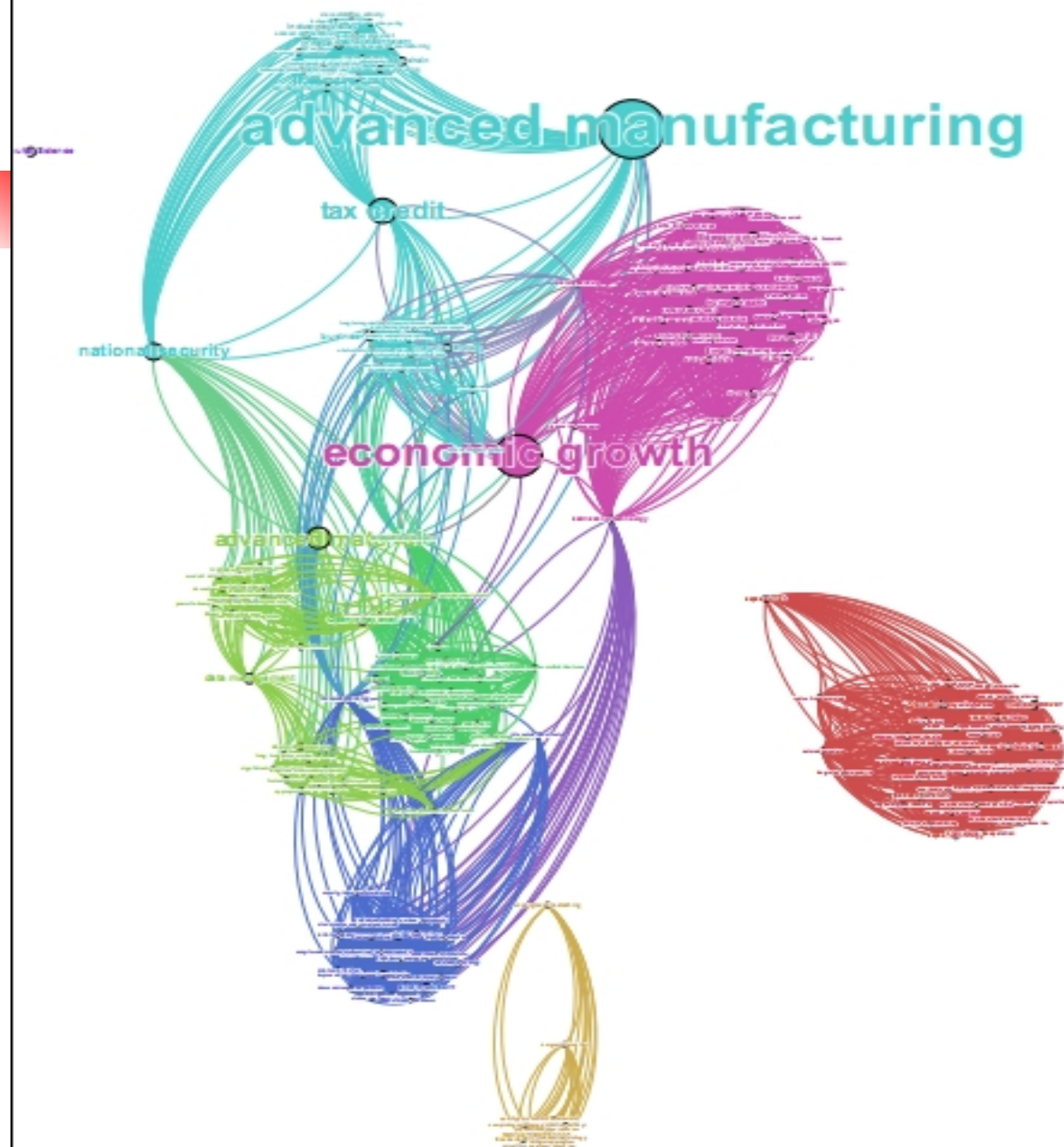


4.1 The Key Reports Analysis

- The Key Reports Analysis
 - Majestic SEO
 - <http://www.majesticseo.com/reports/site-explorer>
 - 15 most influential reports from the OSTP website

4.1 The Key Reports Analysis

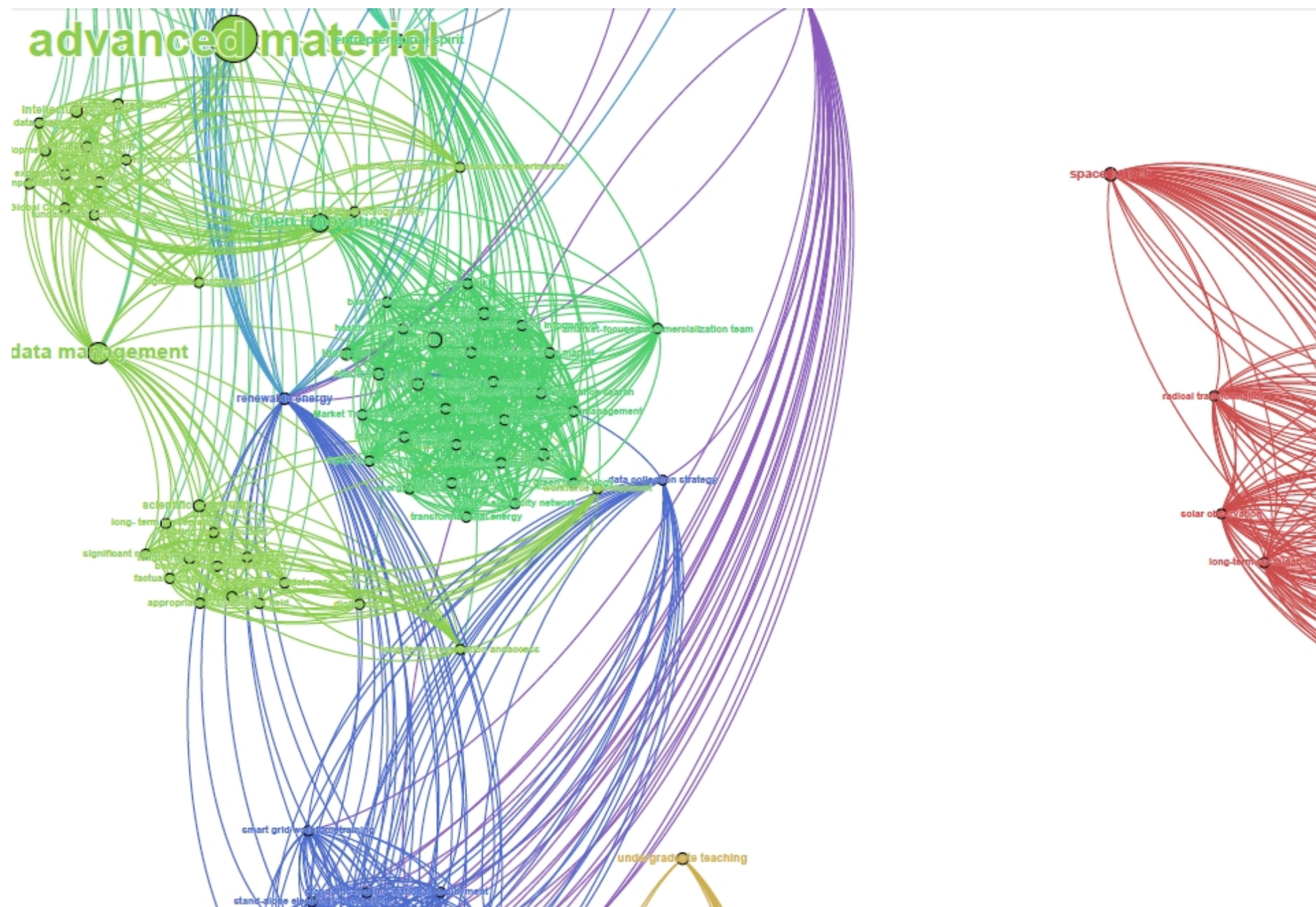
Rank	Report Title	Published Time	Backlinks
1	Memo: Increasing Access to the Results of Federally Funded Research	2013/2/22	6647
2	NATIONAL SPACE POLICY of the UNITED STATES of AMERICA	2010/06/28	2837
3	A POLICY FRAMEWORK FOR THE 21 st CENTURY GRID : Enabling Our Secure Energy Future	2011/06	1150
4	NATIONAL BIOECONOMY BLUEPRINT	2012/04	877
5	Preparing a 21 st Century Workforce Science, Technology, Engineering, and Mathematics (STEM) Education in the 2014 Budget	2013/02	854
6	Materials Genome Initiative for Global Competitiveness	2011/06	749
7	Report to the President on Ensuring American Leadership in Advanced Manufacturing	2011/06	654
8	Executive Office of the President President's Council of Advisors on Science and Technology	2012/07	577
9	A POLICY FRAMEWORK FOR THE 21 ST CENTURY GRID: A PROGRESS REPORT	2013/01	354
10	THE FEDERAL SCIENCE, TECHNOLOGY, ENGINEERING, AND MATHEMATICS (STEM) EDUCATION PORTFOLIO	2011/12	318
11	SUSTAINING ENVIRONMENTAL CAPITAL: PROTECTING SOCIETY AND THE ECONOMY	2011/07	268
12	TRUSTWORTHY CYBERSPACE: STRATEGIC PLAN FOR THE FEDERAL CYBERSECURITY RESEARCH AND DEVELOPMENT PROGRAM	2011/12	242
13	A STRATEGY FOR AMERICAN INNOVATION: Securing Our Economic Growth and Prosperity	2011/02	241
14	Report to The President Engage to Excel: Producing One Million Additional College Graduates With Degrees In Science, Technology, Engineering, and Mathematics	2012/02	206
15	Winning the Future through Open Innovation - a Progress Report on Our Open Government Initiative	2011/06/08	162

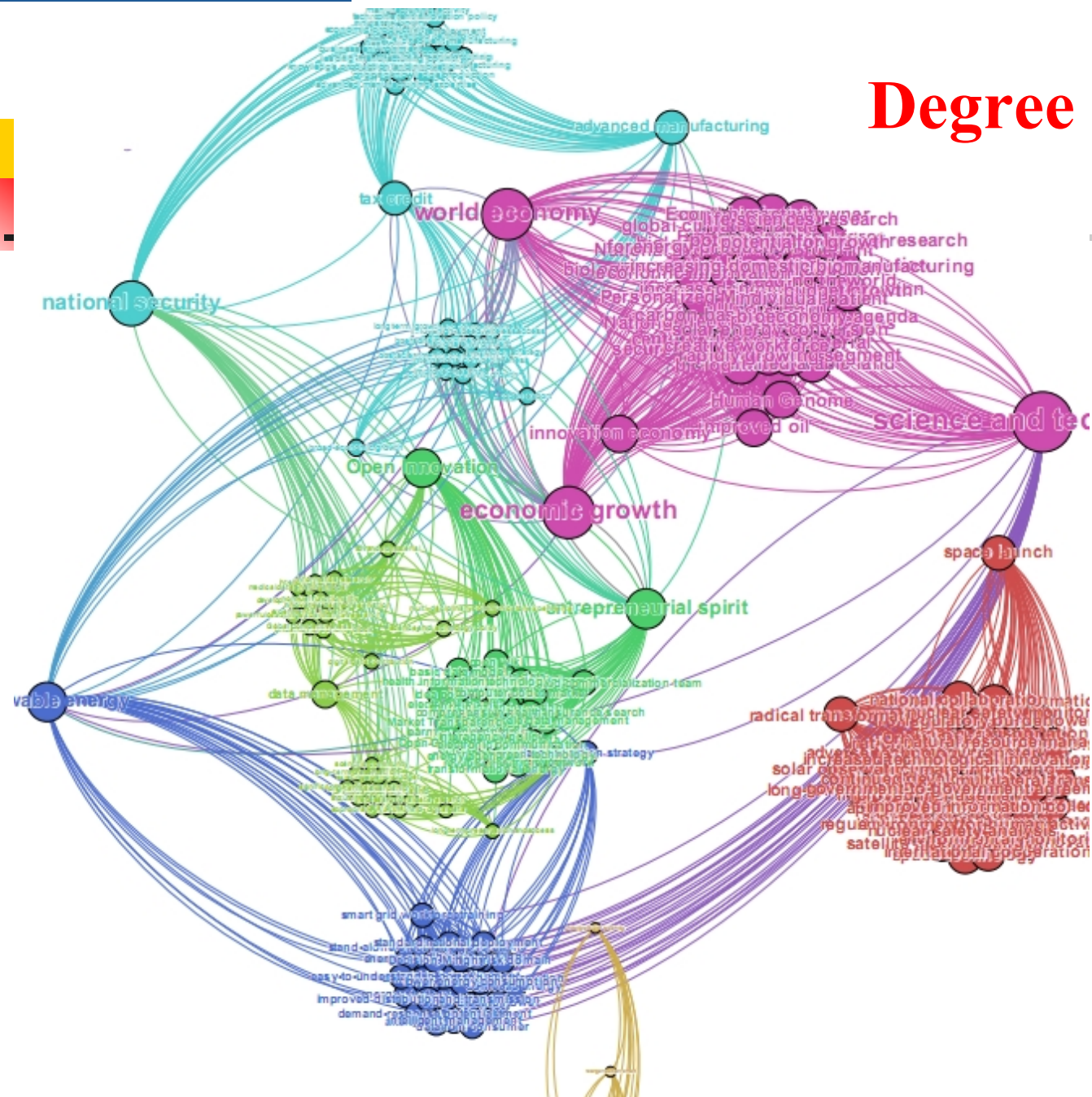


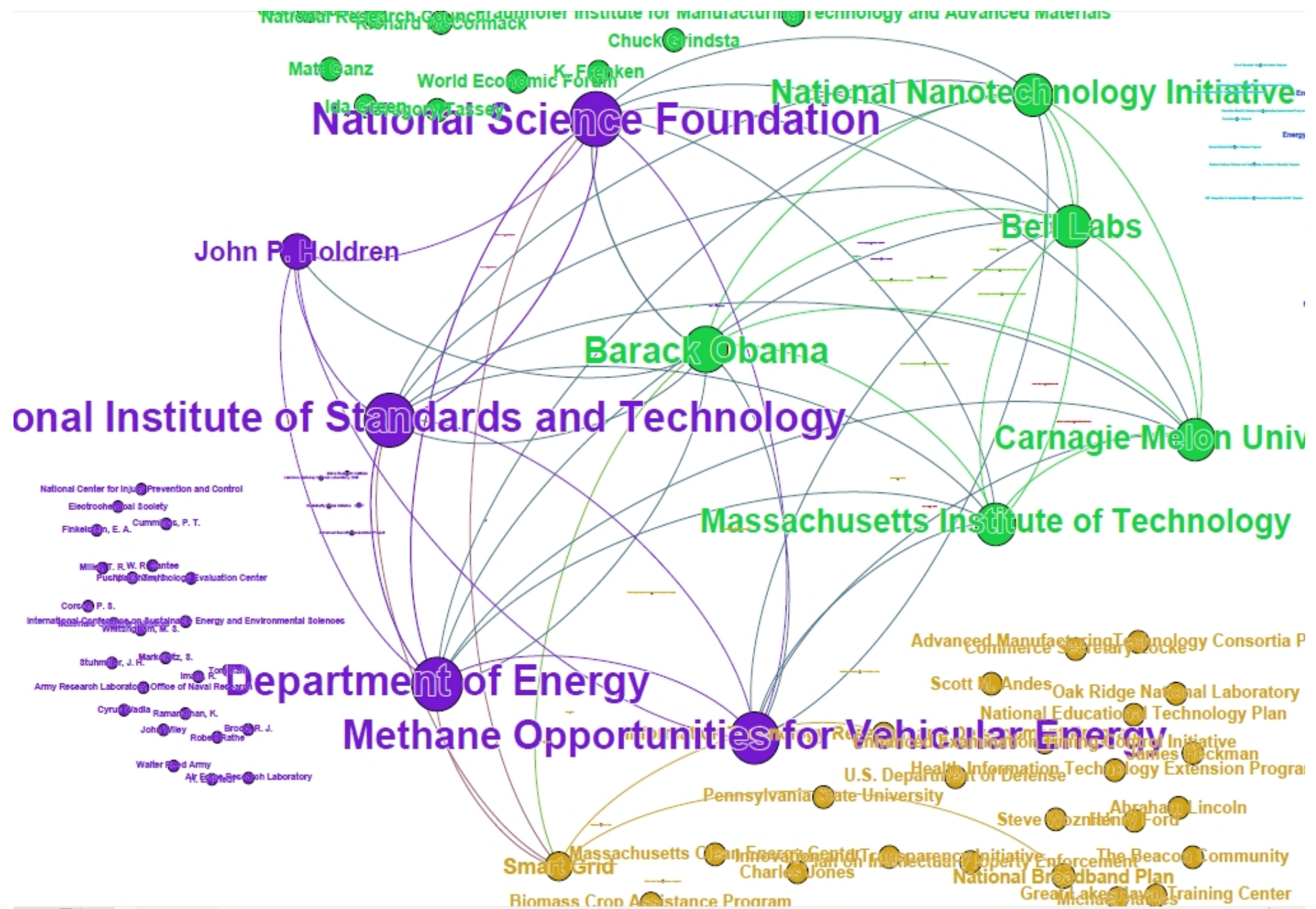
**Major topics of
those reports**

put Science













4.2 The Policy Focuses Analysis

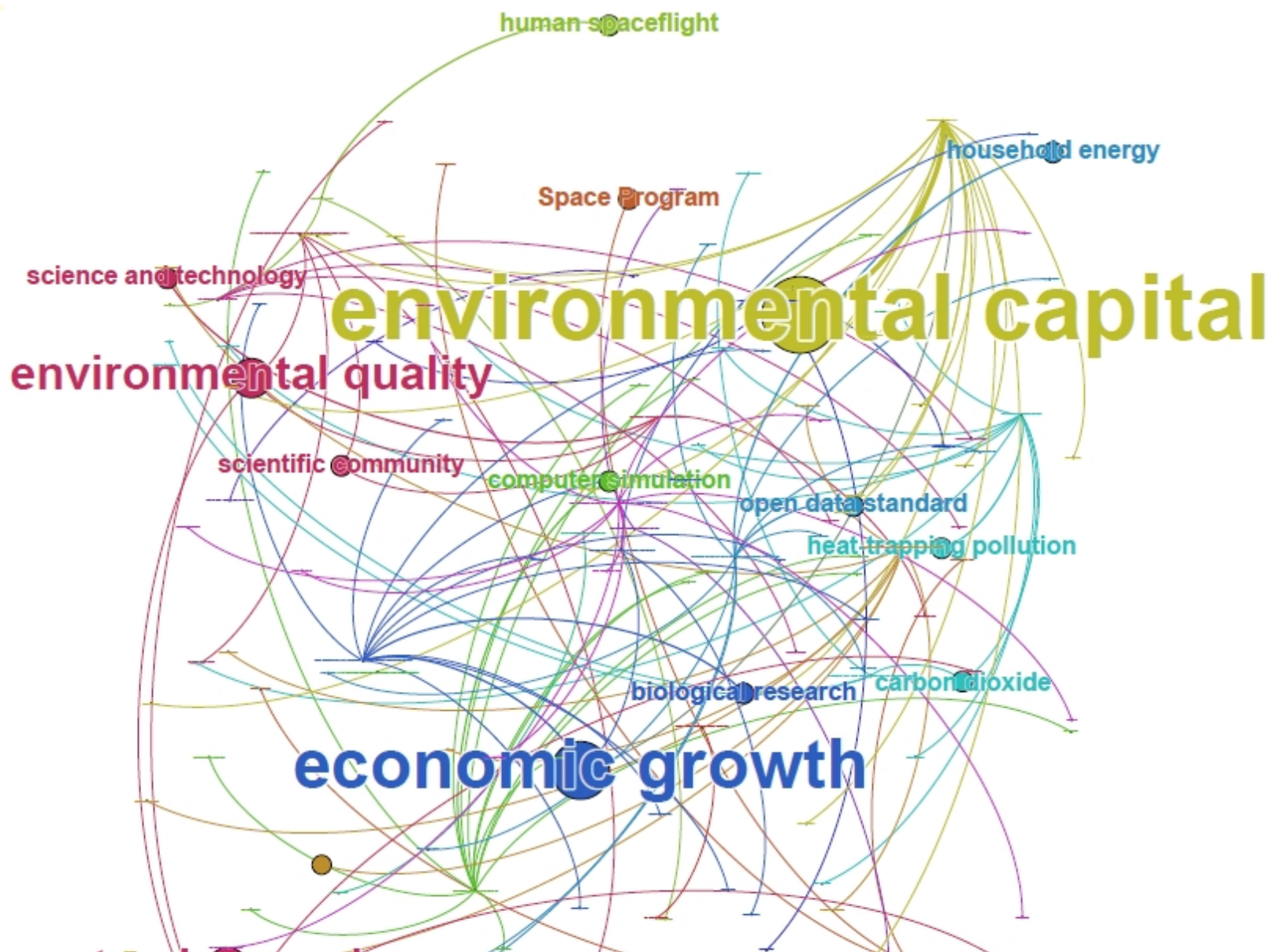
- Released directory analysis
- Hot topics analysis
- Feature topics analysis
- Major Initiatives or Program analysis



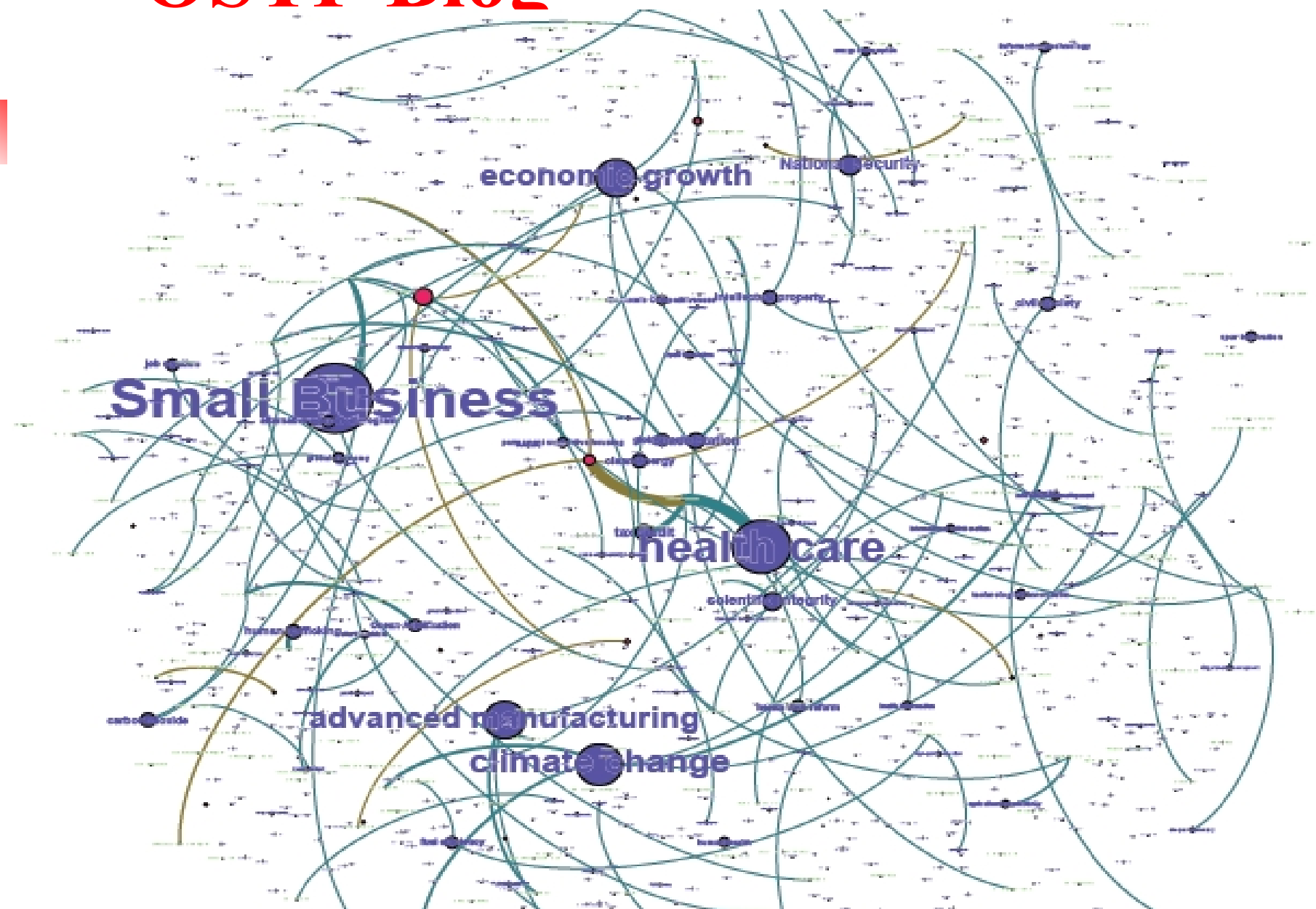
4.2 The Policy Focuses Analysis

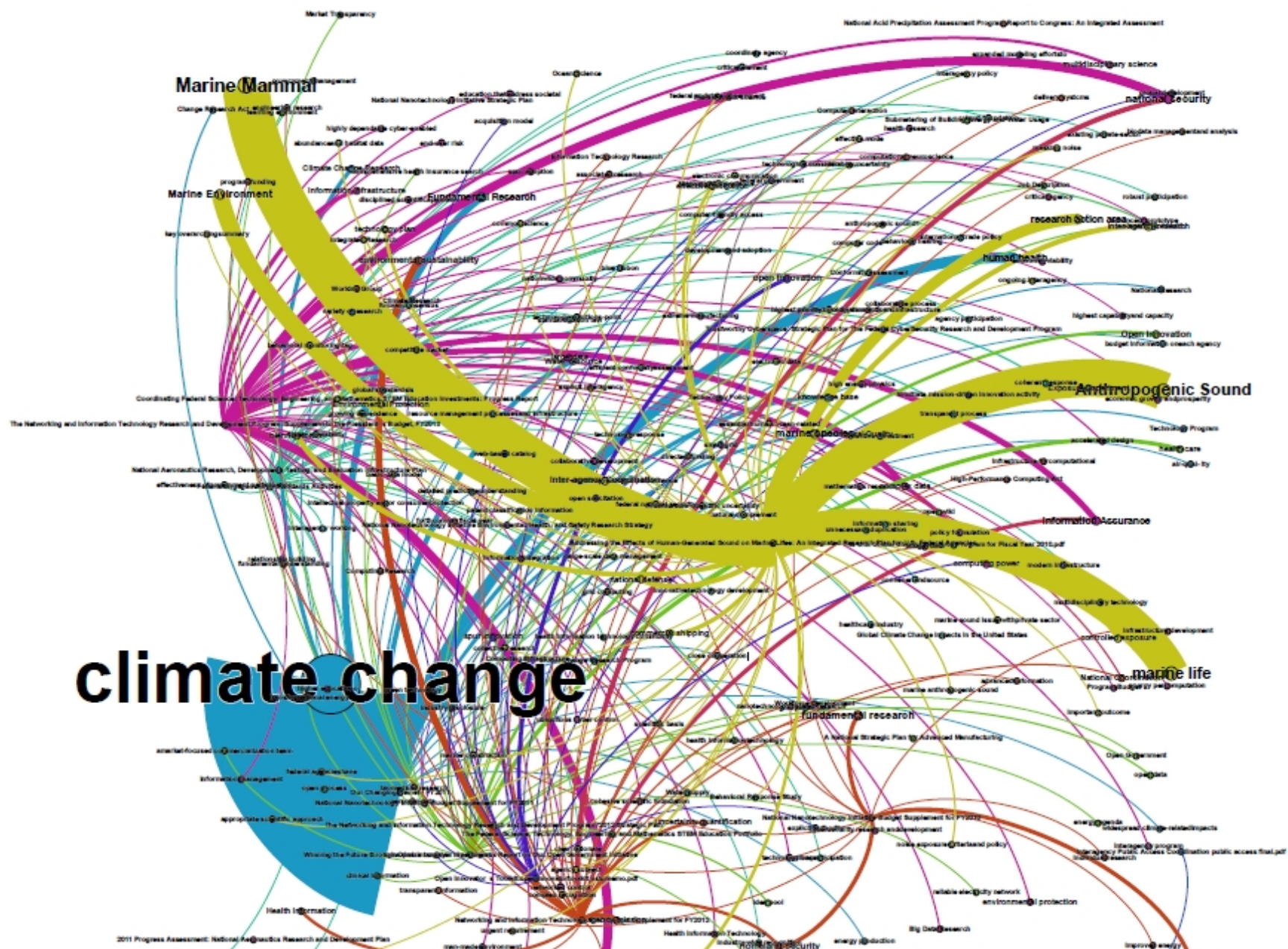
- Released directory analysis
 - To analyze the policy focuses of each directory
 - Four directories are chose
 - OSTP Pressroom
 - OSTP Blog
 - NSTC Documents & Reports
 - PCAST Documents & Reports;

OSTP Pressroom



OSTP Blog





PCAST (President's Council of Advisors on Science and Technology) Documents & Reports

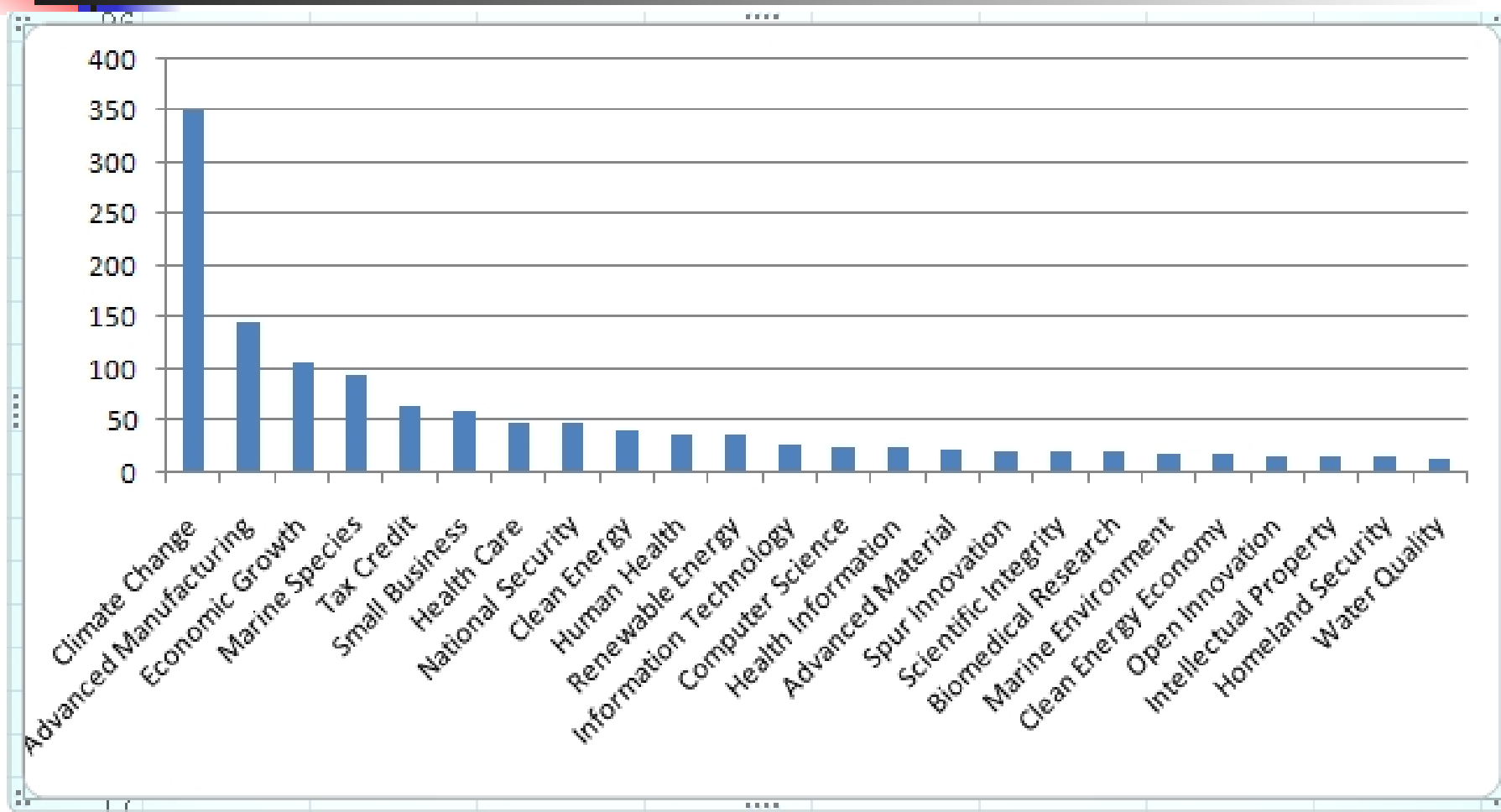




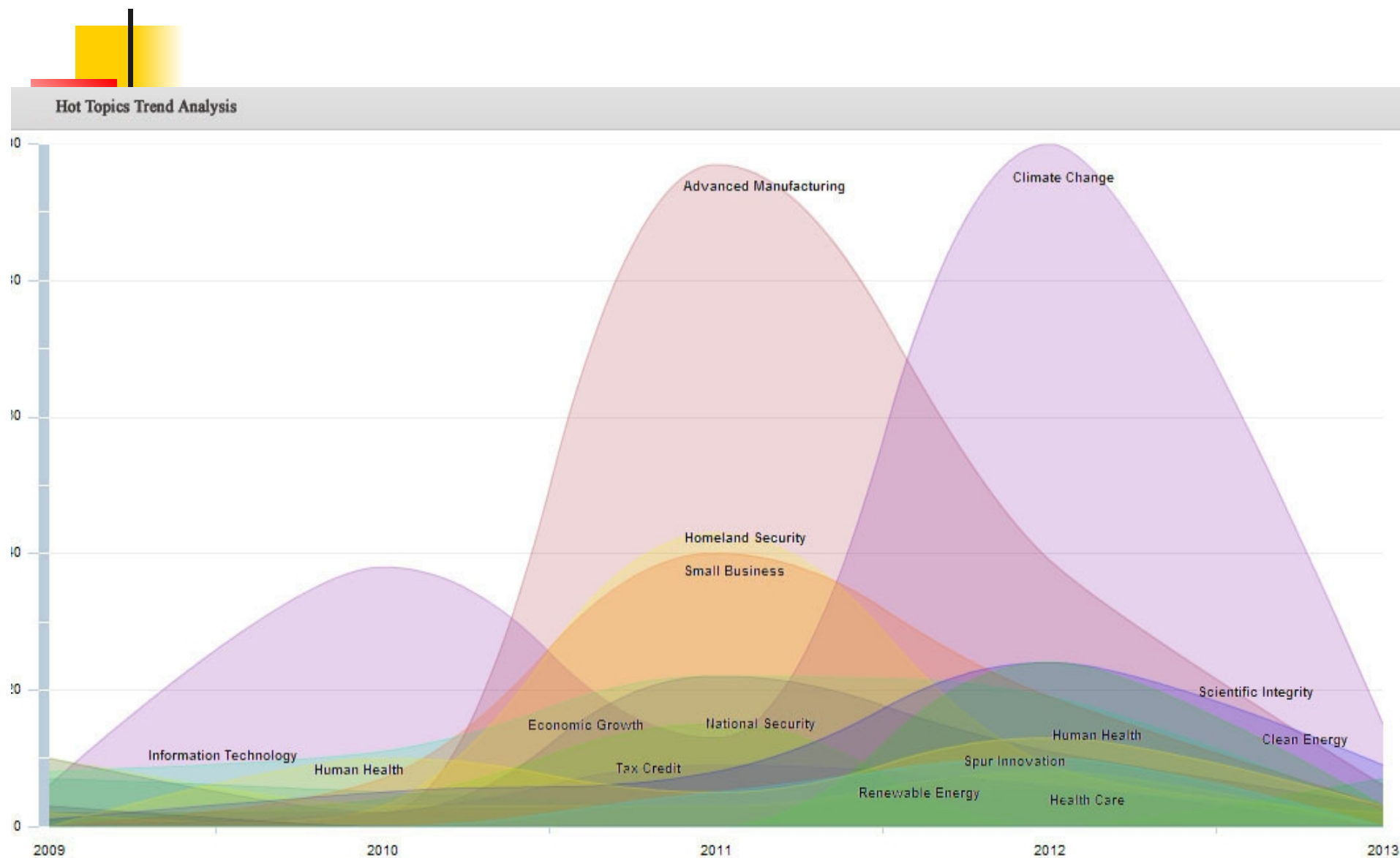
4.2 The Policy Focuses Analysis

- Hot topics analysis
 - Identify hot topics focused by OSTP Web site
 - Find changing status of hot topics each year

Hot topics



changing status of hot topics



4.2 The Policy Focuses Analysis

- Feature topics analysis
 - Feature topics is a column in OSTP Web site
 - It indicates the policy focuses of OSTP



LATEST BLOG POSTS

August 8, 2013 at 9:10 AM EDT

Honoring Tech Inclusion Champions of Change at the White House

Last week, the White House honored 11 heroes as Champions of Change for Tech Inclusion—leaders who have done extraordinary work to connect kids from underrepresented and underserved communities to tech skills and opportunities. President Obama has emphasized time and again that equipping students with science, technology, engineering, and math (STEM) skills is key to our Nation's competitiveness in the global marketplace.

August 8, 2013 at 1:41 PM EDT

We the Geeks: "Robots"

Join our next We the Geeks Google+ Hangout on Friday, August 9th to discuss robots.

July 29, 2013 at 10:44 AM EDT

Honoring Open Government and Civic Hacking Champions of Change

Last Tuesday, the White House honored 14 Open Government and Civic Hacking Champions of Change—extraordinary Americans working to improve their communities through technology, innovation, and civic participation. As entrepreneurs, innovators, organizers, and community leaders, these "Champions of Change" have made a tremendous positive impact.

July 24, 2013 at 4:47 PM EDT

Top White House Officials Convened All Star

FEATURED TOPICS

CLIMATE ACTION PLAN:

On June 25, 2013, President Obama laid out his comprehensive plan for steady action to reduce greenhouse gas pollution in America, prepare our country for the impacts of climate change, and lead global efforts to fight it. The Plan is a recognition that climate change is unequivocal, its primary cause is greenhouse gas pollution, and it is threatening the health of our communities, families, and economy. Learn more:

- Text of the President's Speech
- Weekly Address
- Infographic
- Fact Sheet
- The Climate Action Plan
- OSTP Blog Post
- What Business Leaders are Saying

OPEN DATA:

PRESSROOM

July 16, 2013 - Media A Meeting

June 24, 2013 - Media A Highlights Citizen Science Change Event

June 24, 2013 - Fact Sheet Materials Genome Initiative

June 18, 2013 - Press Fact Highlight Open Science

May 9, 2013 - Press Release Administration Release Rules to Enhance Government Fuel Economic Growth

April 29, 2013 - Fact Sheet Partners with Industry to Credentialed for High-D

April 22, 2013 - NEW Director to Host White House Science

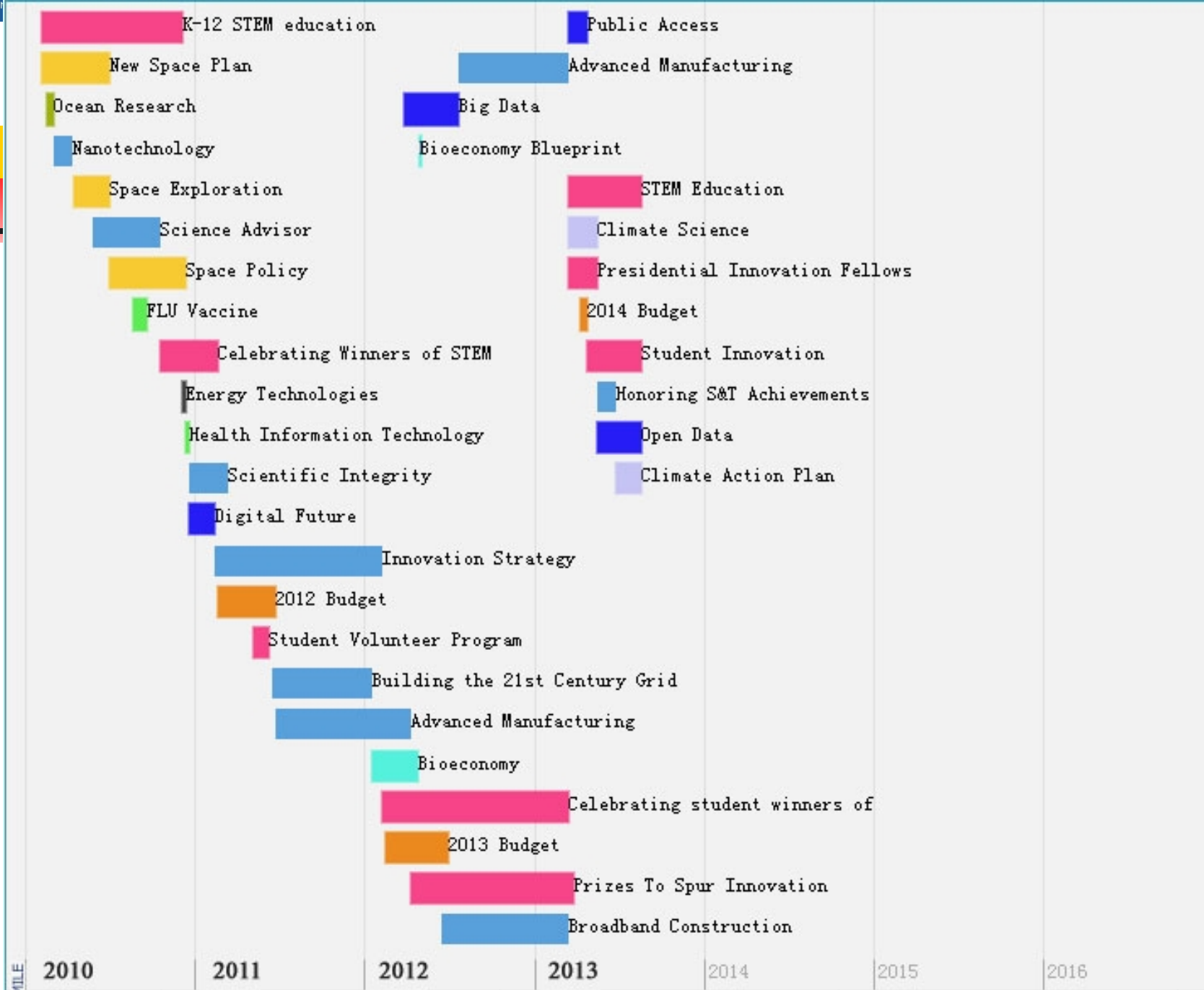
April 16, 2013 - Press Fact Administration Release Economy and Resiliency

April 10, 2013 - Press Fact Federal R&D Budget



Feature Topics Analysis

立





4.2 The Policy Focuses Analysis

- Major Initiatives or Programs analysis
 - An initiative or program means action taken for some policy
 - Perform analysis to
 - figure out the most important initiatives OSTP developed
 - What are they concerned?
 - Which one is active player?
 - Are those initiatives related?

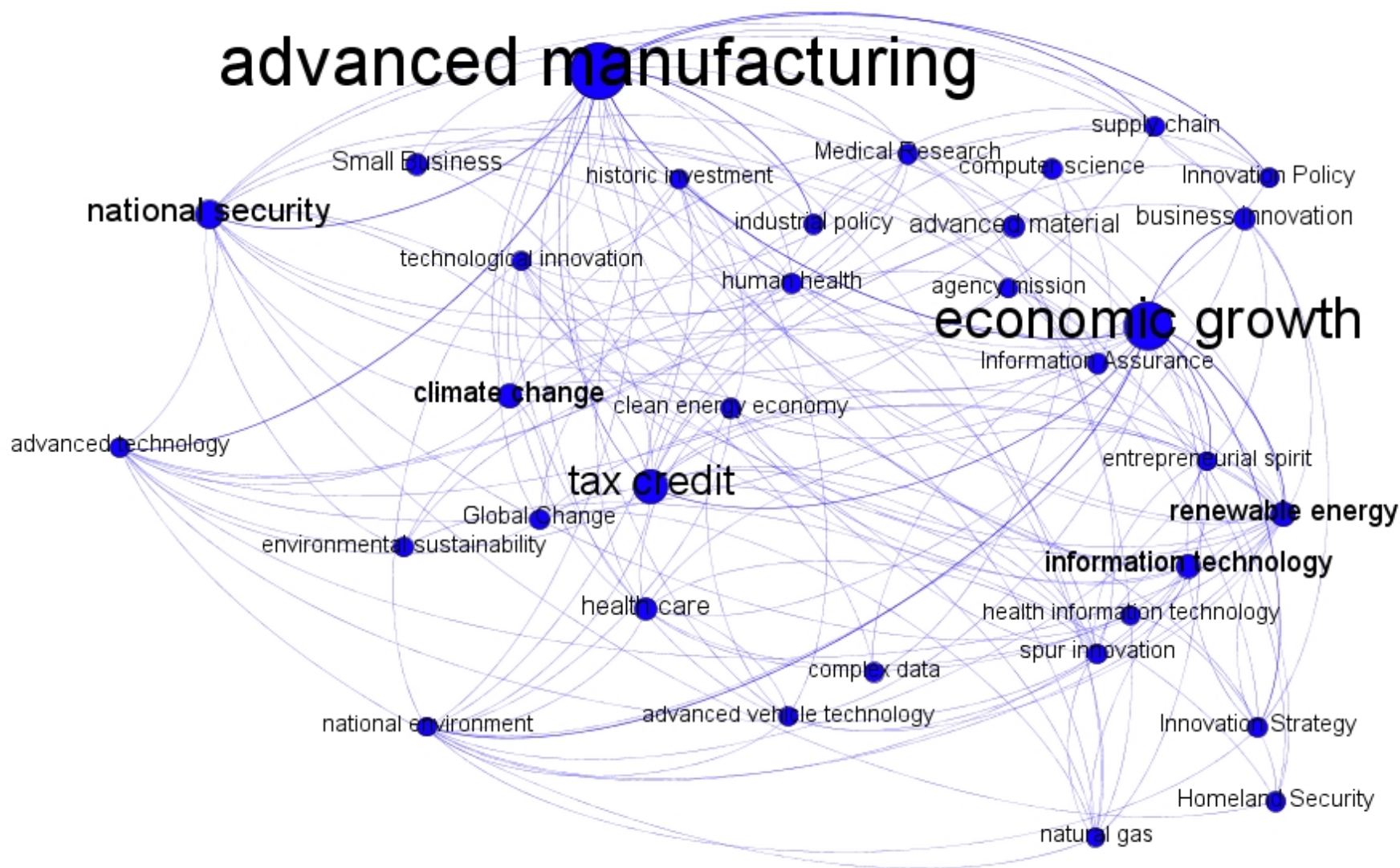
4.2 The Policy Focuses Analysis

■ Major initiatives or program lists

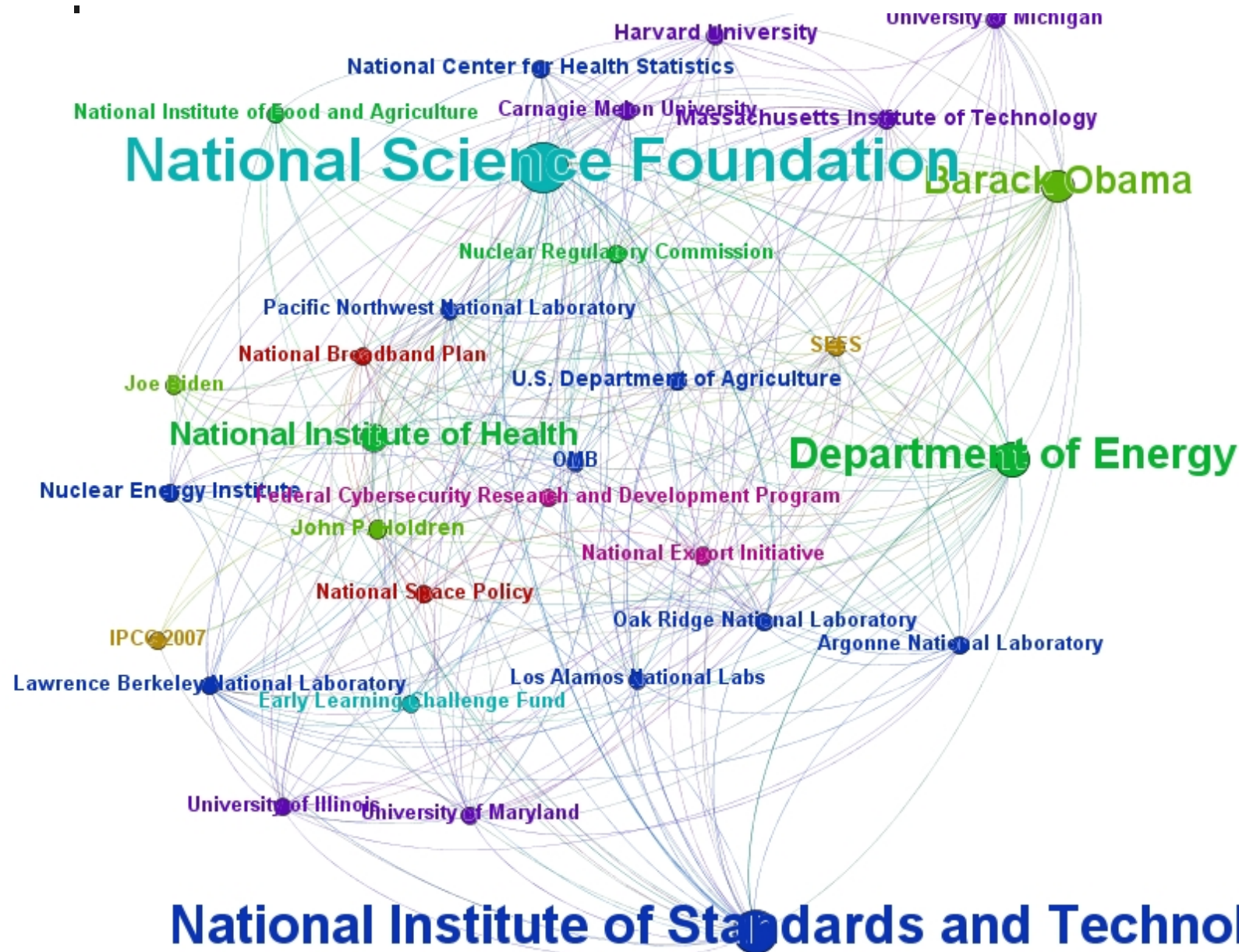
Initiative or Program Name	Frequency	Reports Counts	Appear Time
Smart Grid	185	20	2009-9-20
National Nanotechnology Initiative	120	28	2009-5-13
Information Technology Research and Development Program	81	13	2009-5-13
Methane Opportunities for Vehicular Energy	65	45	2009-9-20
Materials Genome Initiative	51	14	2011-6-24
Federal Energy Management Program	31	8	2012-5-15
Startup America initiative	20	11	2011-3-2
US Global Change Research Program	18	13	2010-3-24
National Acid Precipitation Assessment Program	18	5	2012-1-20
National Robotics Initiative	16	11	2011-6-24
Open Government Initiative	12	10	2009-9-20
Wireless Innovation and Infrastructure Initiative	10	10	2011-2-15
Rural Energy for America Program	10	9	2009-9-20
Energy Data Initiative	10	4	2012-5-24
Advanced Manufacturing Initiative	8	2	2011-6-23



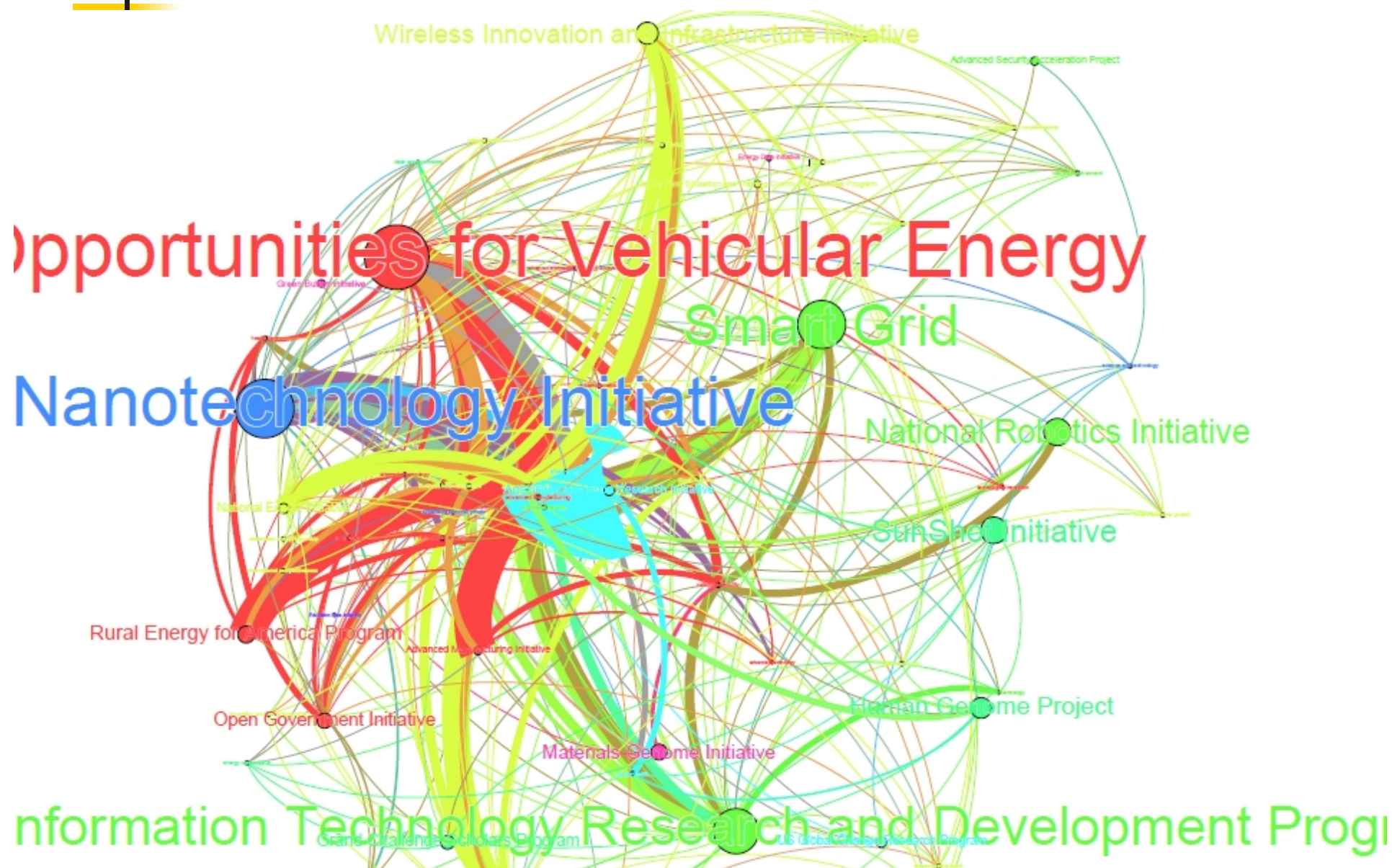
Major concerns in those Initiative



active players in those initiatives



Initiative relationship

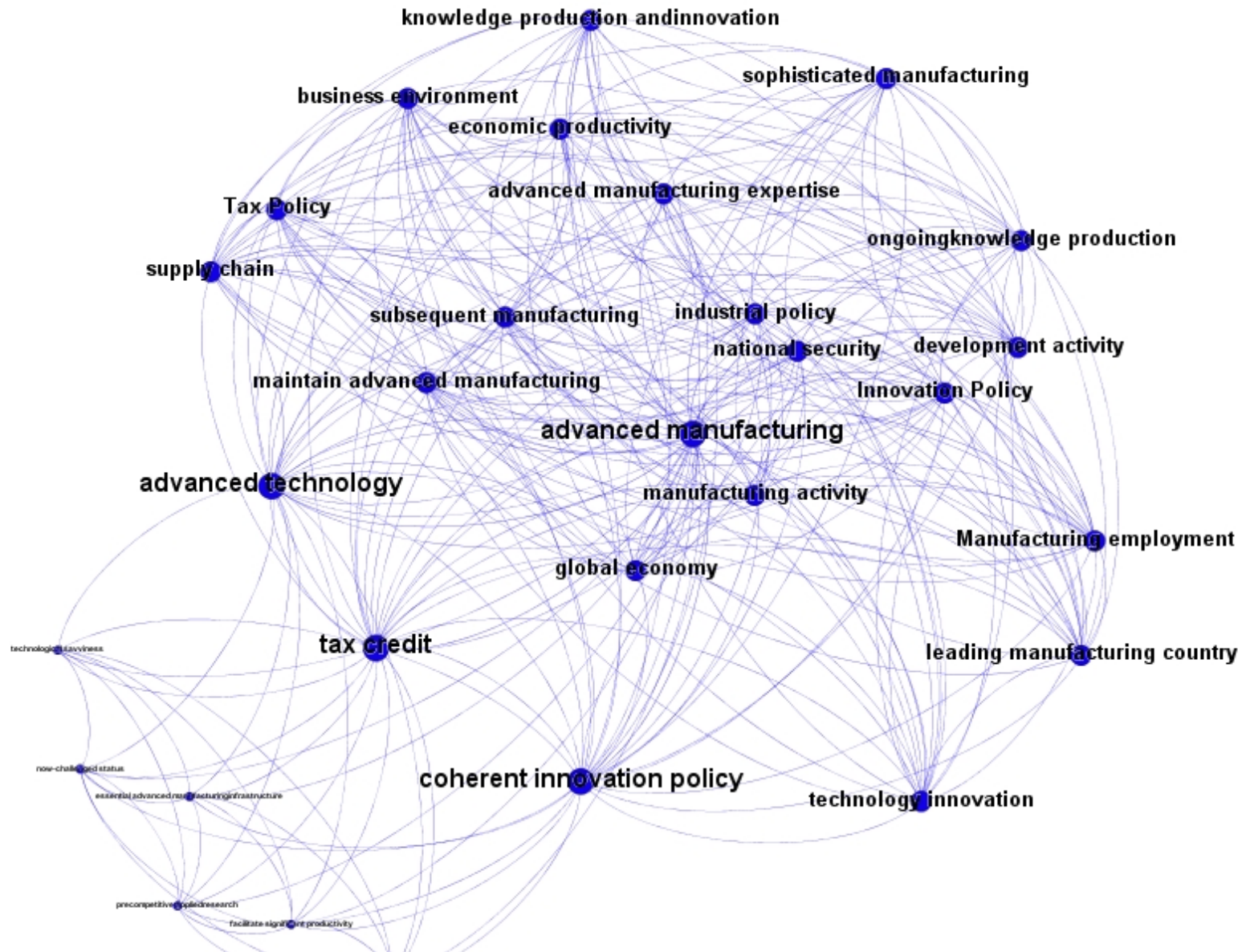




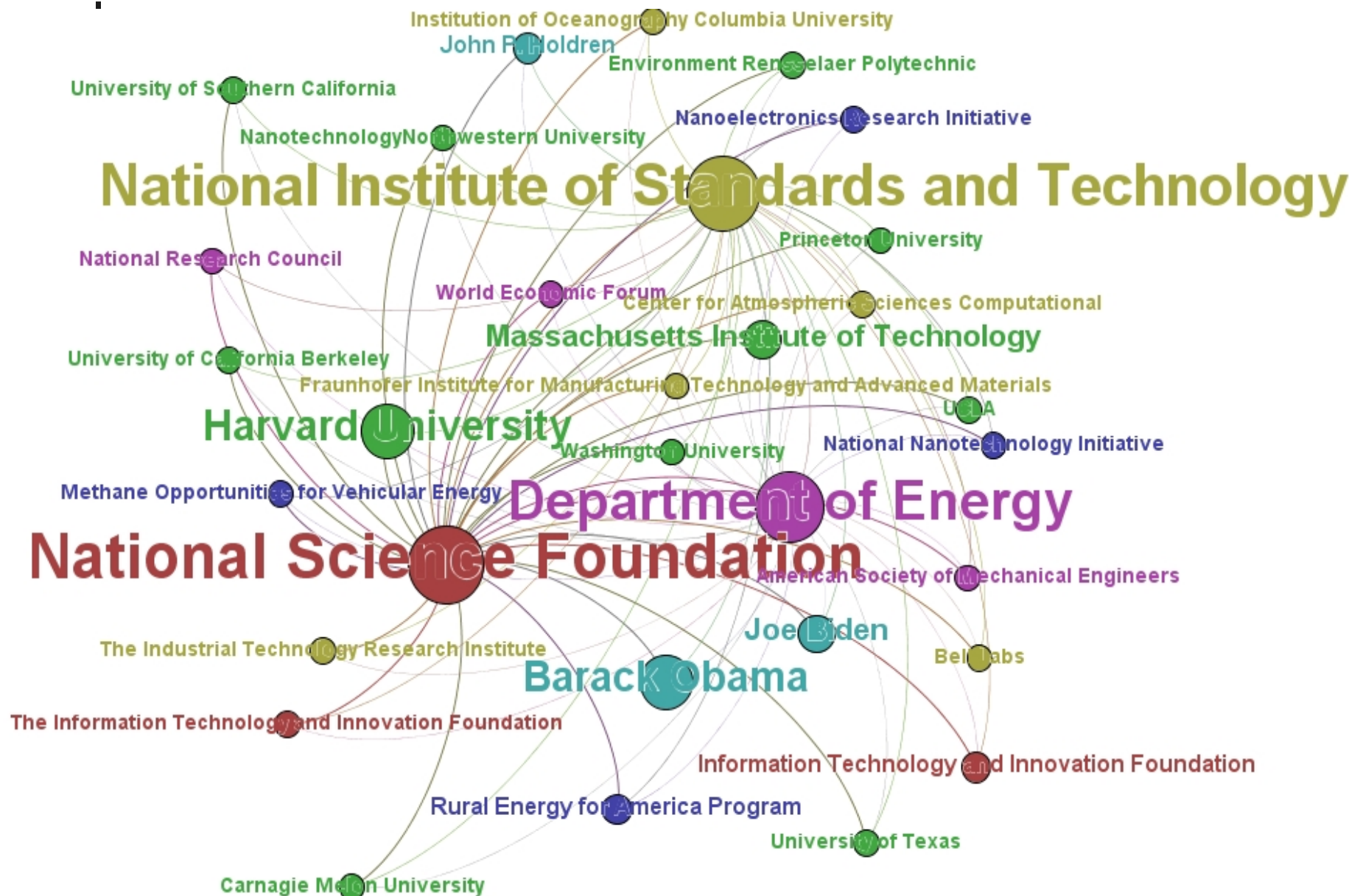
4.3 Inter-relationship Analysis

- **Inter-relationship Analysis**
 - disclose inter-relationship within OSTP web site
 - Take “Advanced Manufacturing Initiative” as an example

Inter-Topics in Advanced Manufacturing Initiative



Inter-Objects in Advanced Manufacturing Initiative





Conclusion and Discussion

- It seem a sound method to profile the policy taken by Obama administration
- But weakness
 - Simple, based mainly on frequency and co-occurrence
 - rely heavenly on the accuracy of the term and objects extraction
 - Not checked with policy expert



- Thank You for Your Attention!
- 谢谢!