



Technology Mining for Emerging Trends and Developments:

dynamic term clustering and semantic analysis

Pavel Bakhtin, Ozcan Saritas

pbakhtin@hse.ru, osaritas@hse.ru

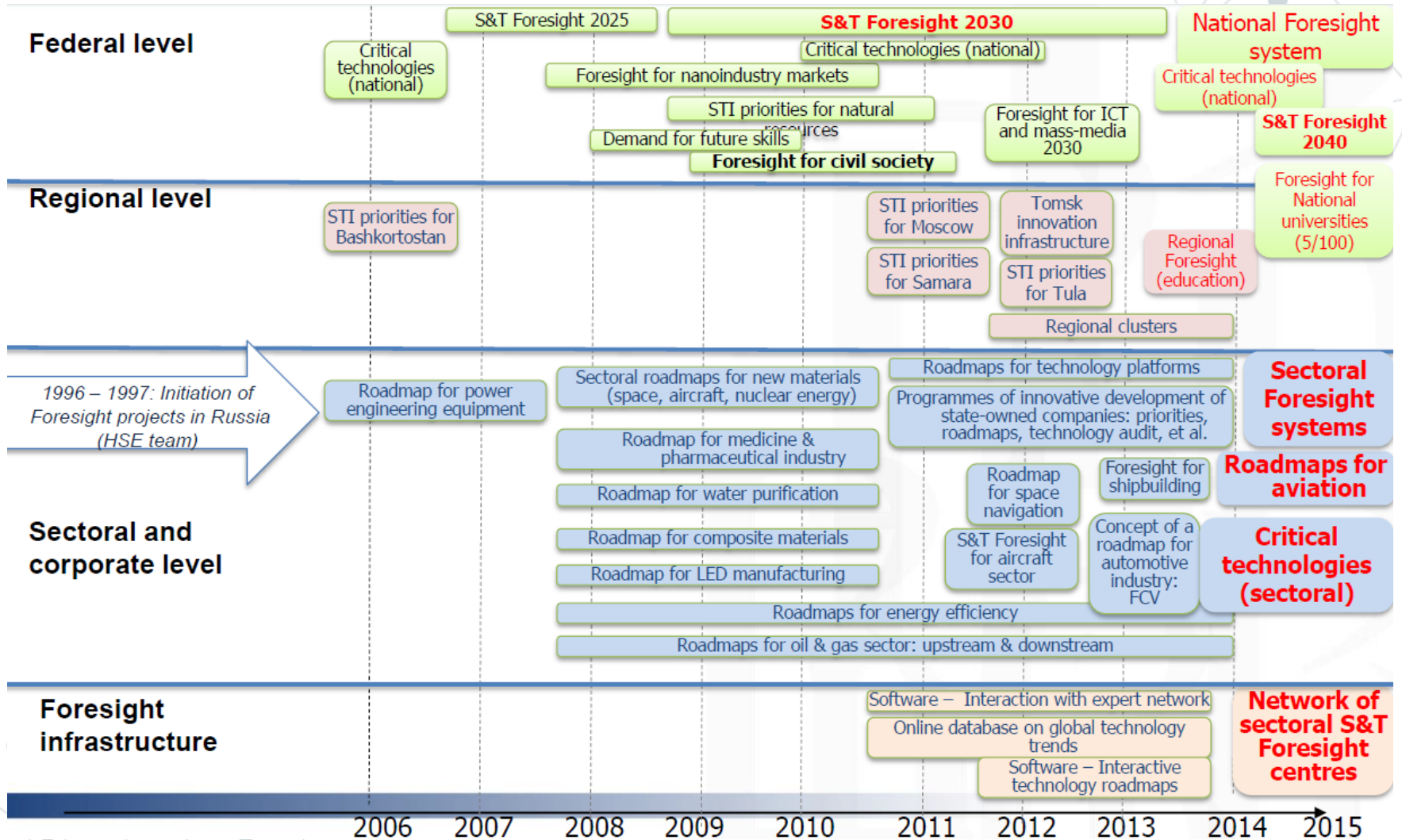
National Research University Higher School of Economics

Outline

- Use of technology intelligence and mining for Foresight
- Commonly used methods and tools
- Research question and proposed methodology
- Case analysis on Photonics
- Findings and conclusions of the research

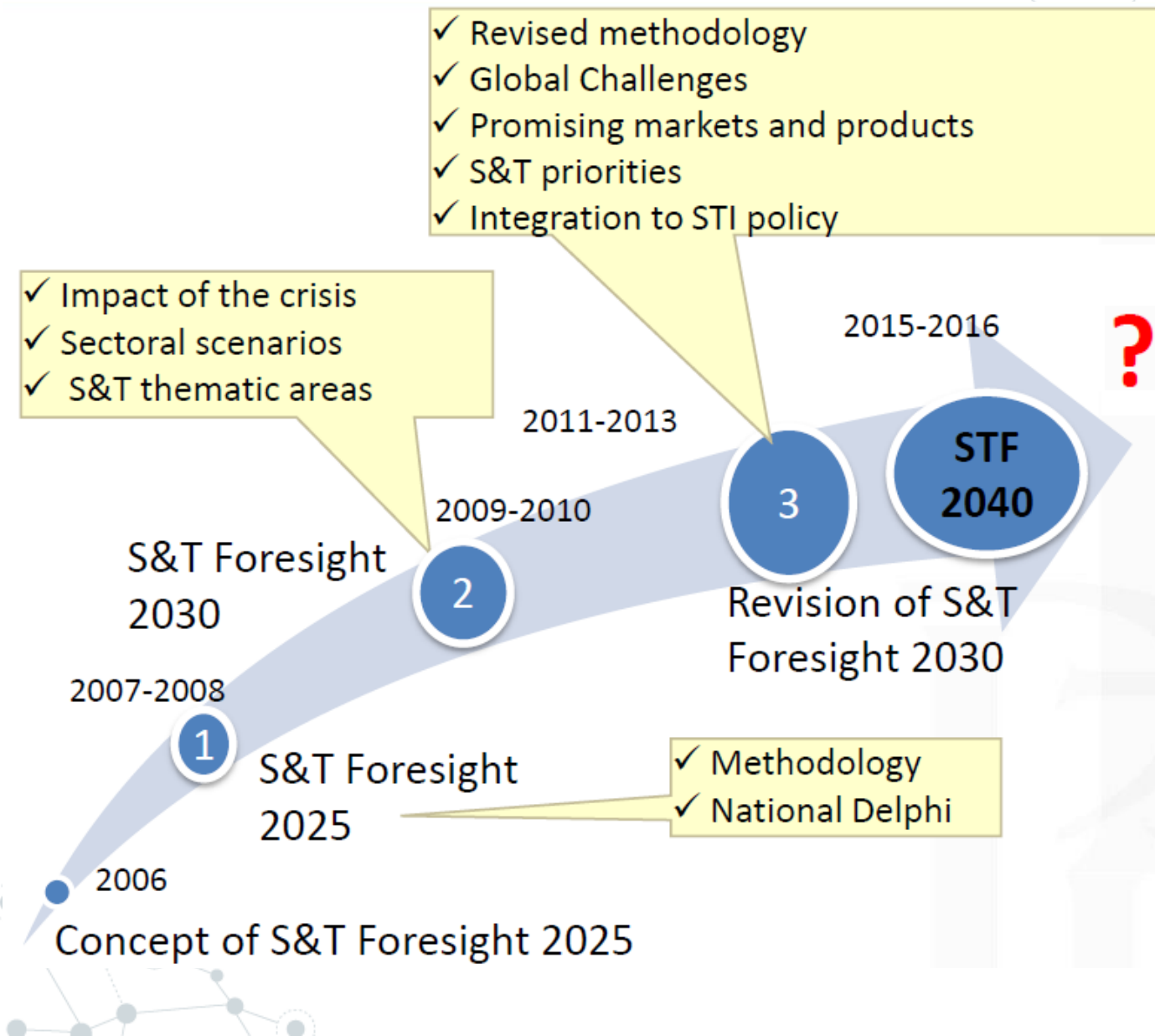


Overview of Foresight activities in Russia



(Chulok, 2015)

The evolution drivers and methodology



Russia 2040: Science & Technology Foresight



- Focusing on the key issues of socio-economic development
- Forecasting of innovative technologies, products and markets (demand vs. supply STI strategies)
- Expanding the time horizon towards 2040 and beyond
- Development of long-term scenarios
- Interdisciplinary and participative approaches
- Focusing on the strategic breakthroughs, “wild cards” and weak signals

Rationales for using Tech intelligence & mining

- Dynamic nature of scientific domains: new concepts, inventions, knowledge, technologies, markets, and innovation
- Increasing amount of available information in the areas of STI
- Higher speed of socio-economic and technological change
- Greater interdisciplinary research, technology diffusion and spillover-effect
- Growing need for
 - reducing the effort required to obtain useful information
 - detecting key trends
 - constructing 'new' hypotheses for future developments

Methods and tools



Bibliometrics

- Patent and publication quantitative meta information integration
- Key words and phrases clusters, co-occurrence and correlation matrices
- Factor maps

NLP and machine learning

- Key words and phrases frequencies
- Topic modeling
- Property-function analysis, TRIZ
- Subject Action Object (SAO) analysis

Research question

Integration and elaboration on existing methods and tools in order to:

- Identify trends and other types of developments
- Cluster most relevant objects in the collection
- Analyze networks and developments over the time
- Visualize and interpret the results to enhance further decision-making

Methodology and approach



Collection of S&T
data

Processing,
filtering and
cleaning of data

DATA LAYER

Thematic term
clustering (co-
occurrence and
ontology)

Dynamic term
clustering of data
(correlation
based on
timeline)

CLUSTER LAYER

Cross-cluster intersection analysis

VISUALIZATION LAYER

Semantic
analysis (SAO)

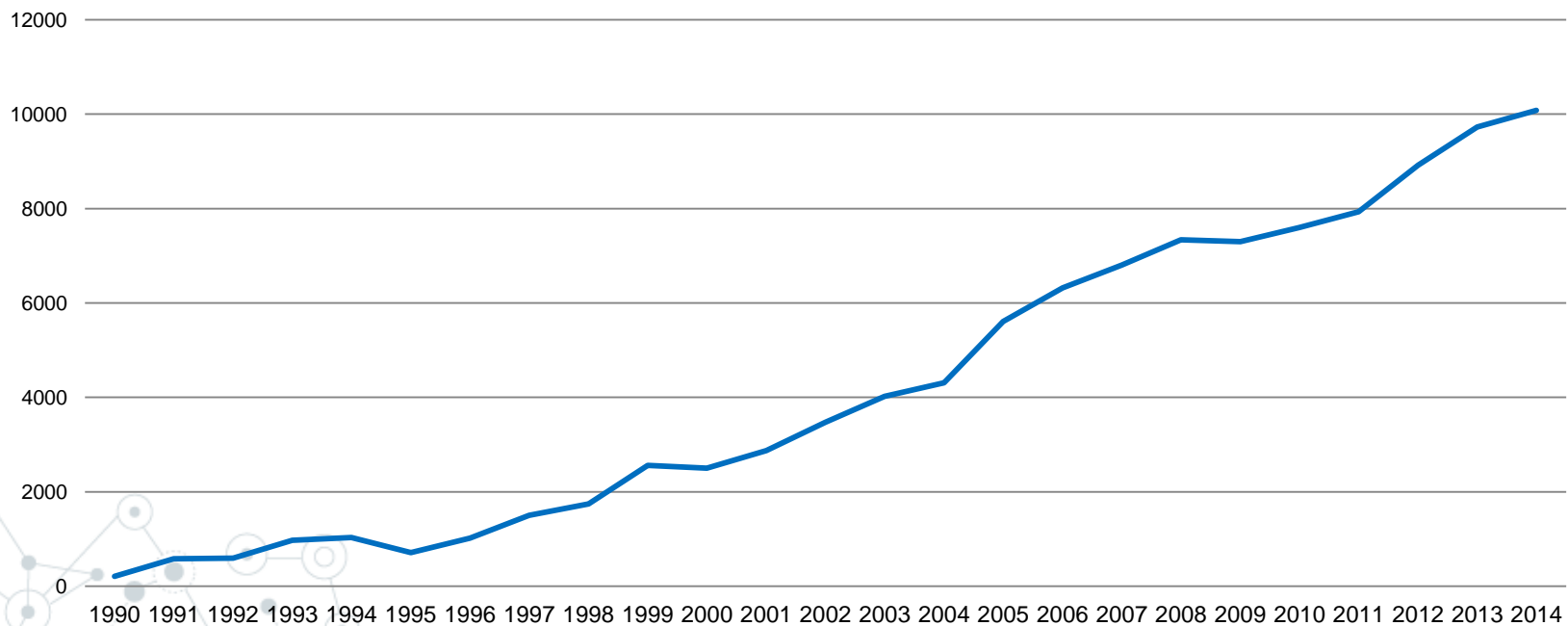
Expert evaluation

SEMANTIC LAYER

Data layer



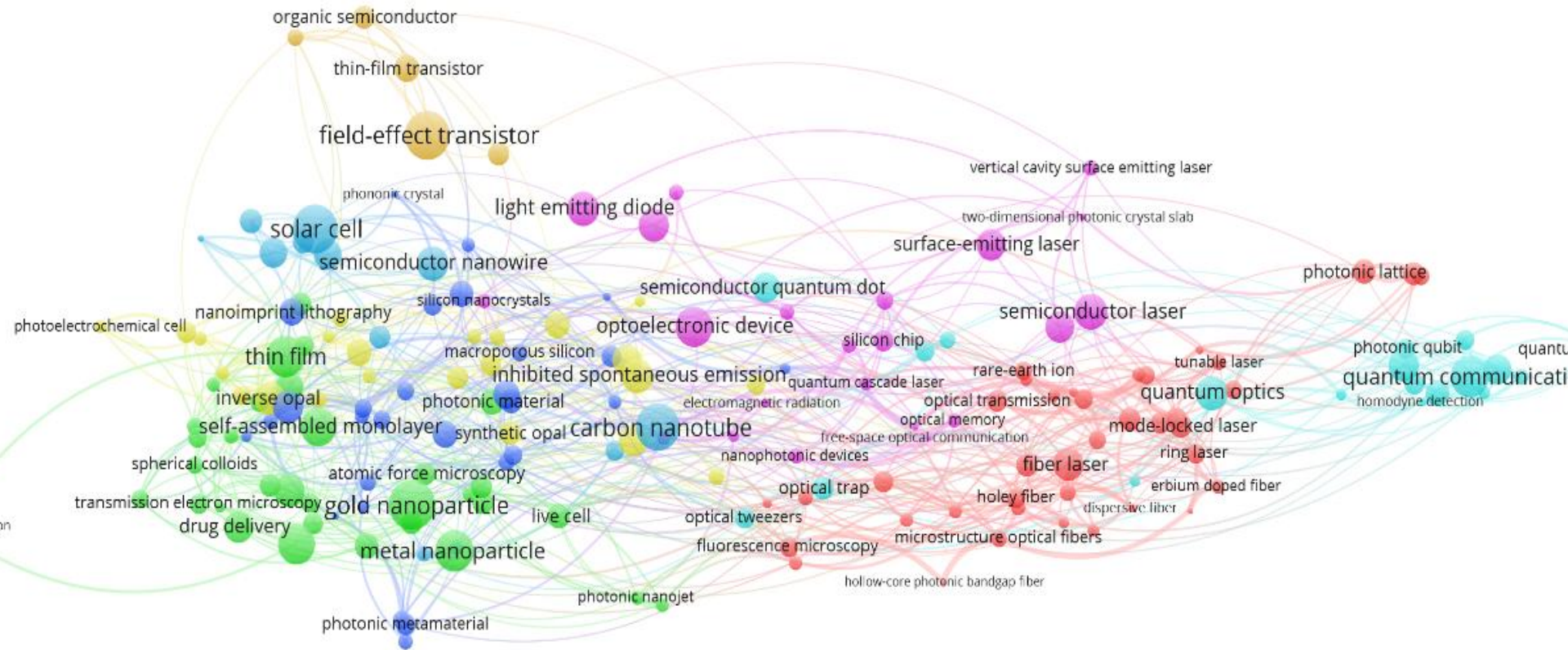
- Query formulation for publication/patent and other S&T databases
 - Extraction of 10% most referenced documents for each year
 - Data cleaning and filtering (using VantagePoint)
- WoS publications on Photonics**



Cluster layer: thematic (1)

- Clustering of cleaned data based on co-occurrence with VOSViewer
- Validation of clusters based on research area ontology
- Development of thesaurus using validated clusters
- Use of thesaurus for further VantagePoint analysis

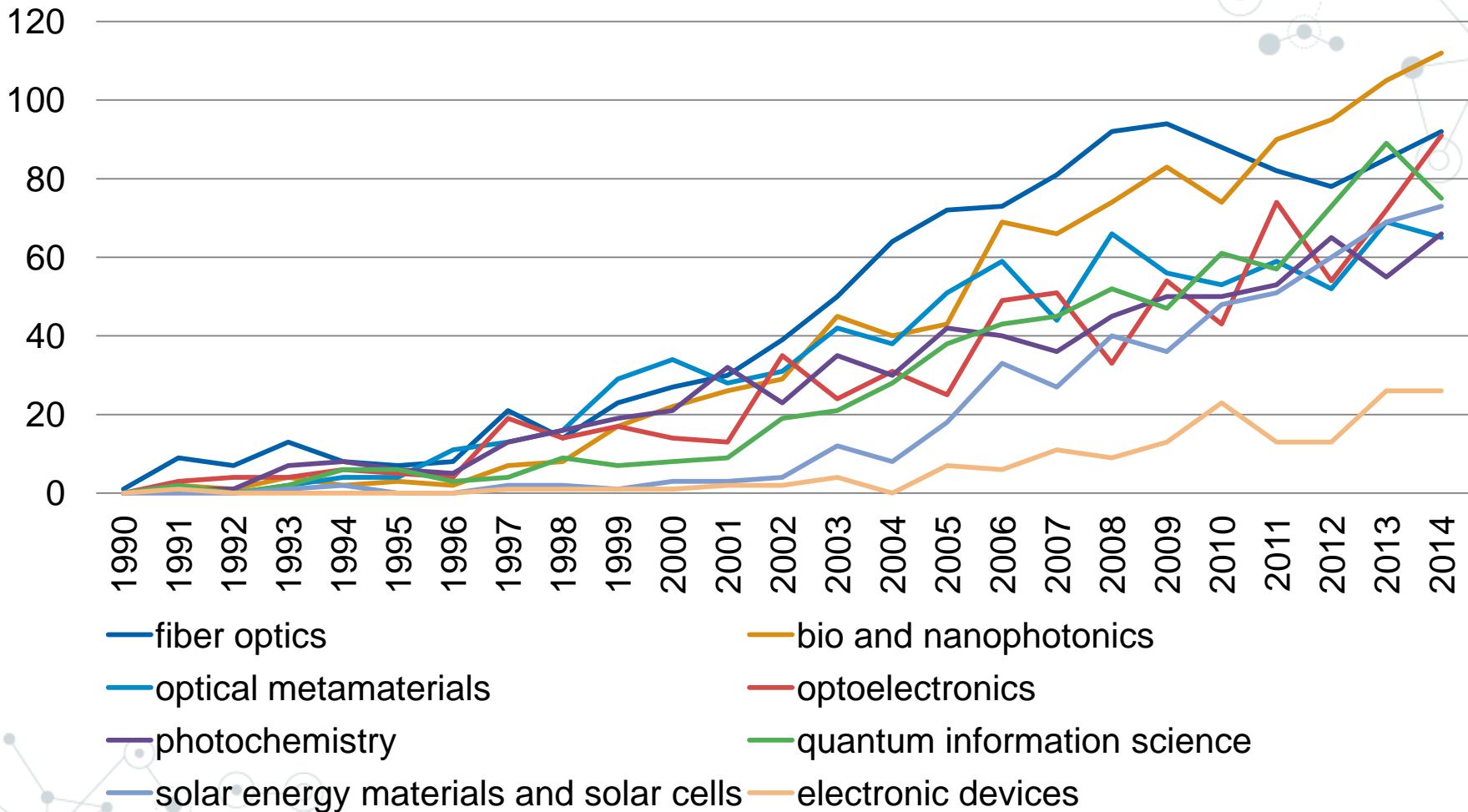
Cluster layer: thematic (2)



Cluster layer: thematic (3)



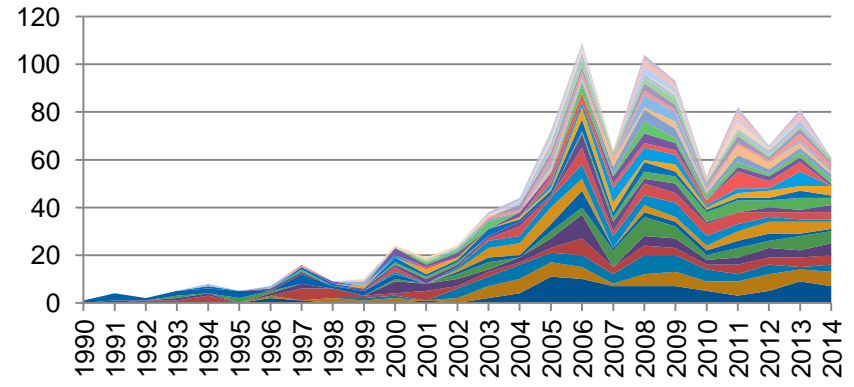
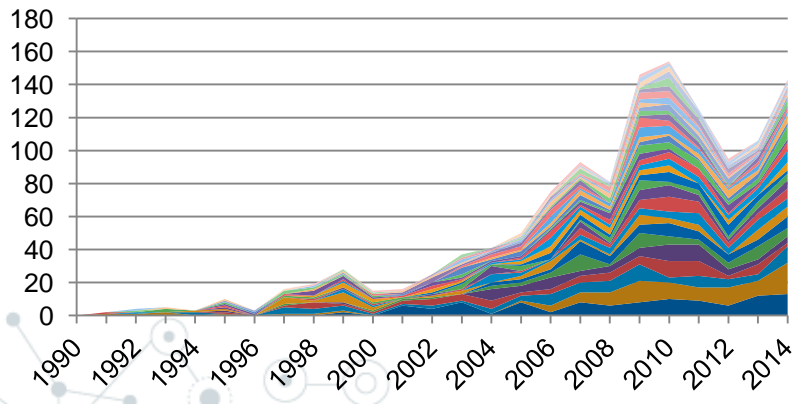
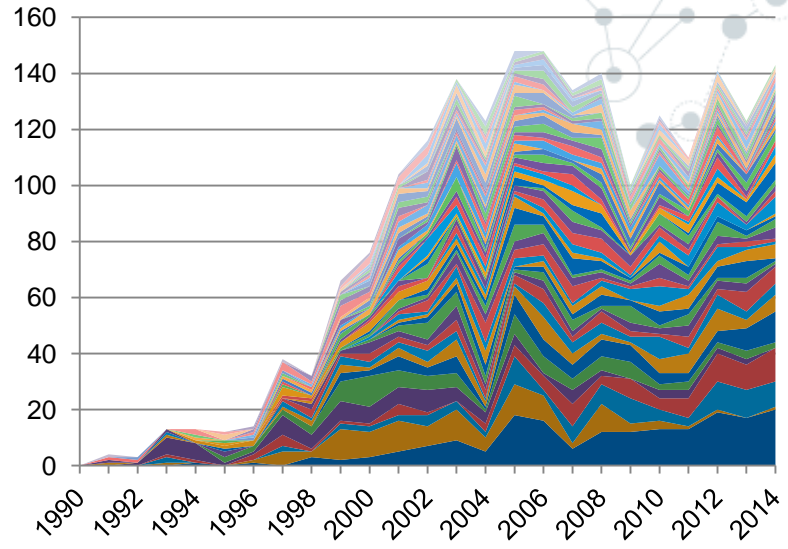
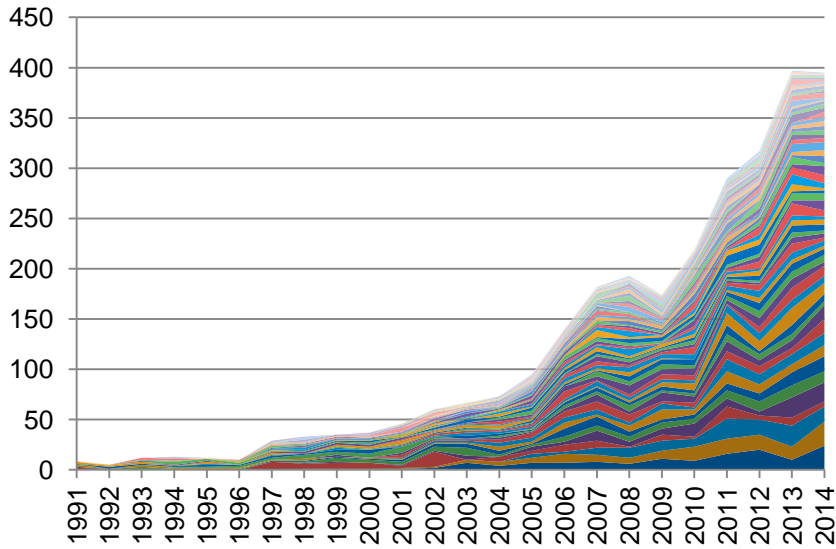
WoS publications for clusters



Cluster layer: dynamic (1)

- Cleaned data term correlation analysis based on S&T documents dynamics over the time period
- Clustering of terms based on correlation scores with VOSViewer
- Development of thesaurus using validated clusters
- Use of thesaurus for further VantagePoint analysis

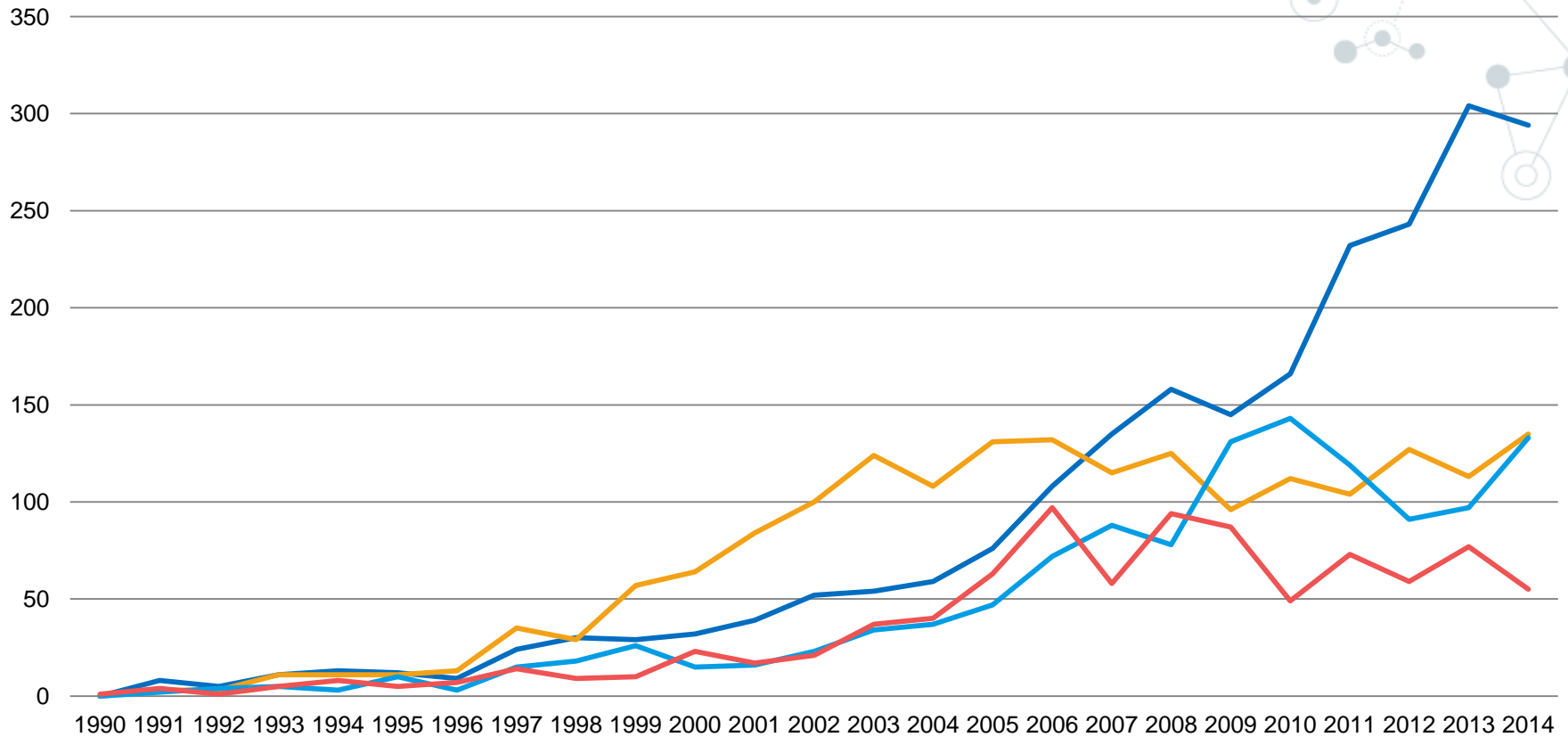
Cluster layer: dynamic (2)



Cluster layer: dynamic (3)



Dynamic term clusters based on WoS publications



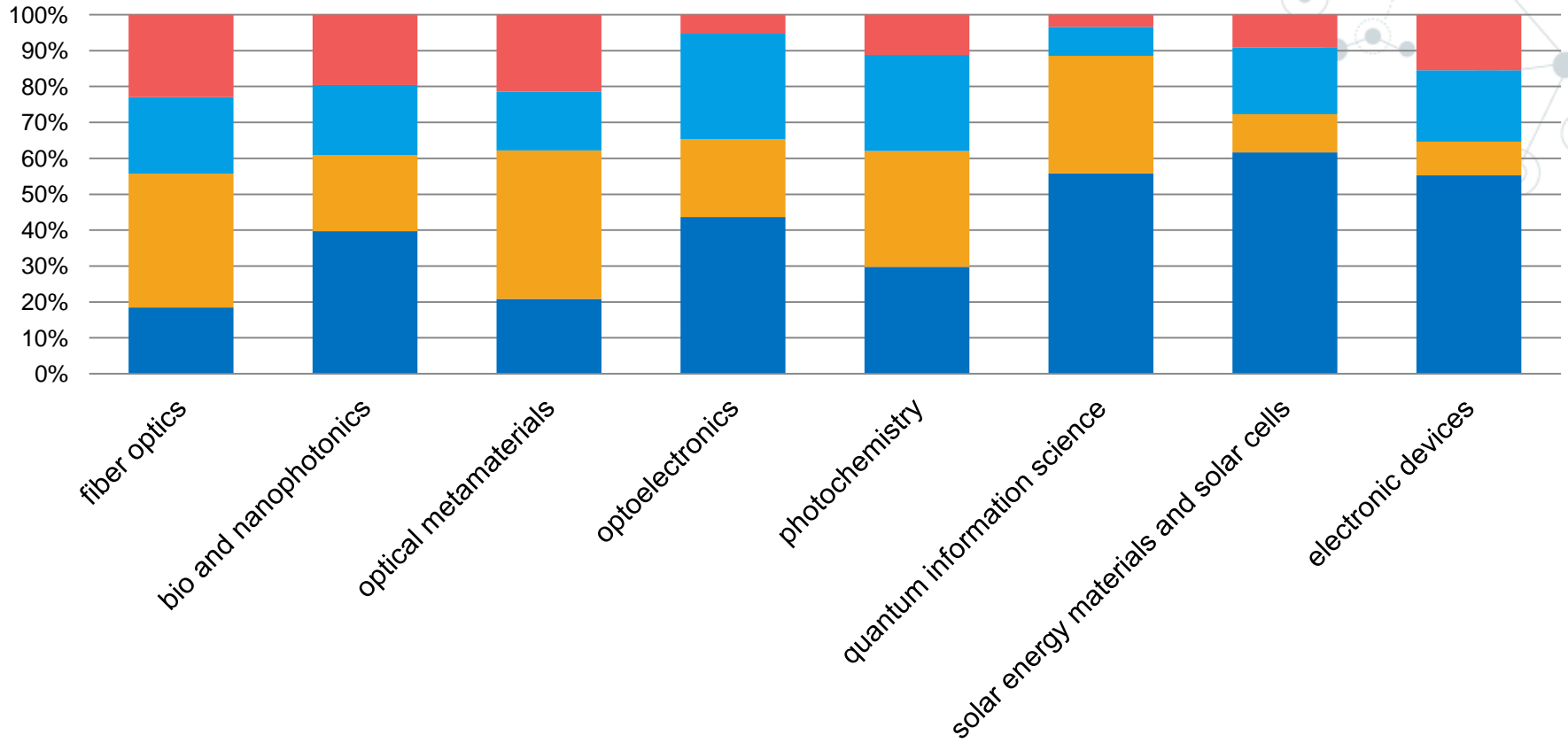
— emerging trends
— early developments

— mature stable developments
— declining developments

Visualization layer (1)



Intersection of thematic cluster and dynamic cluster data



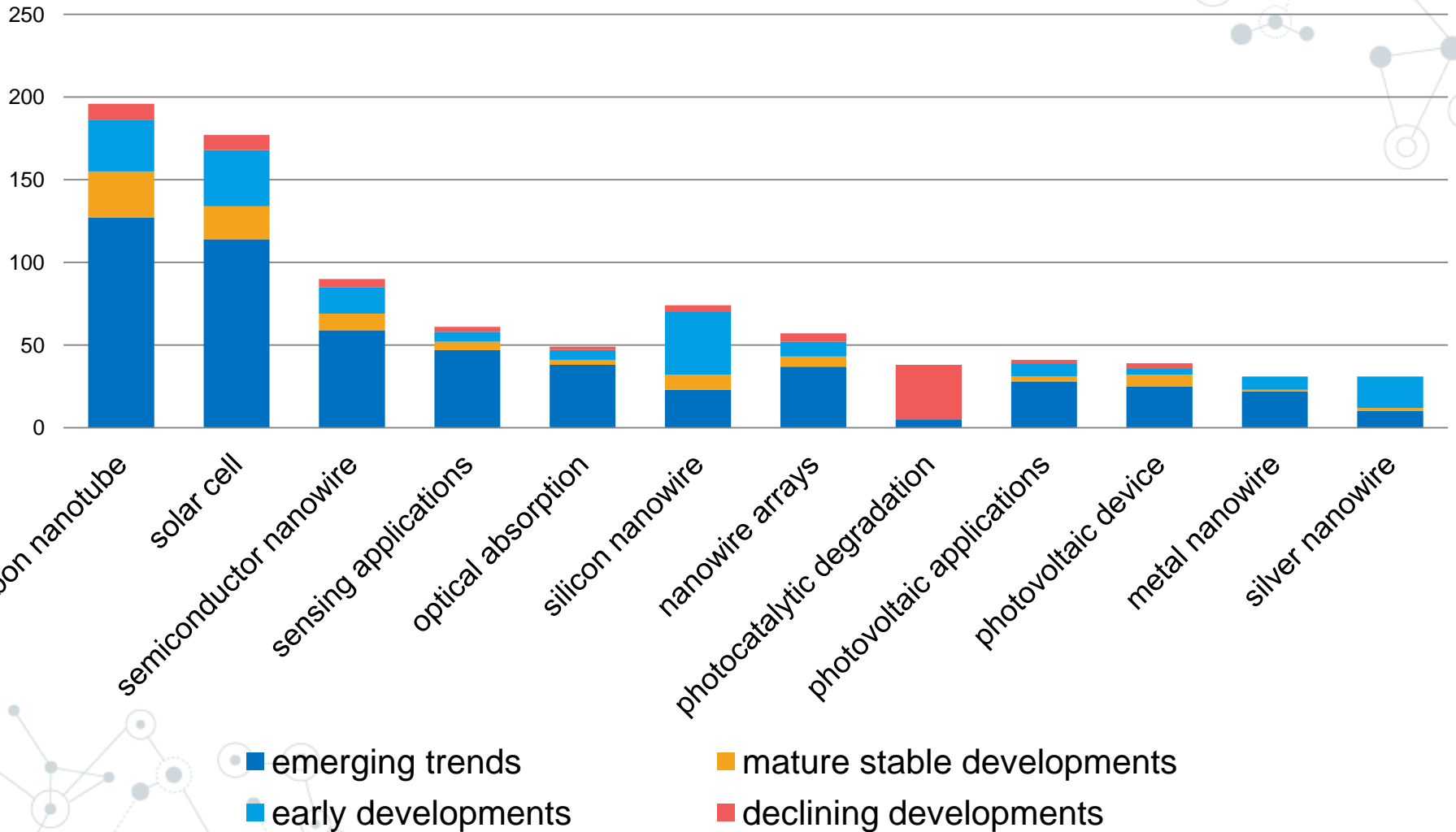
■ emerging trends
■ early developments

■ mature stable developments
■ declining developments

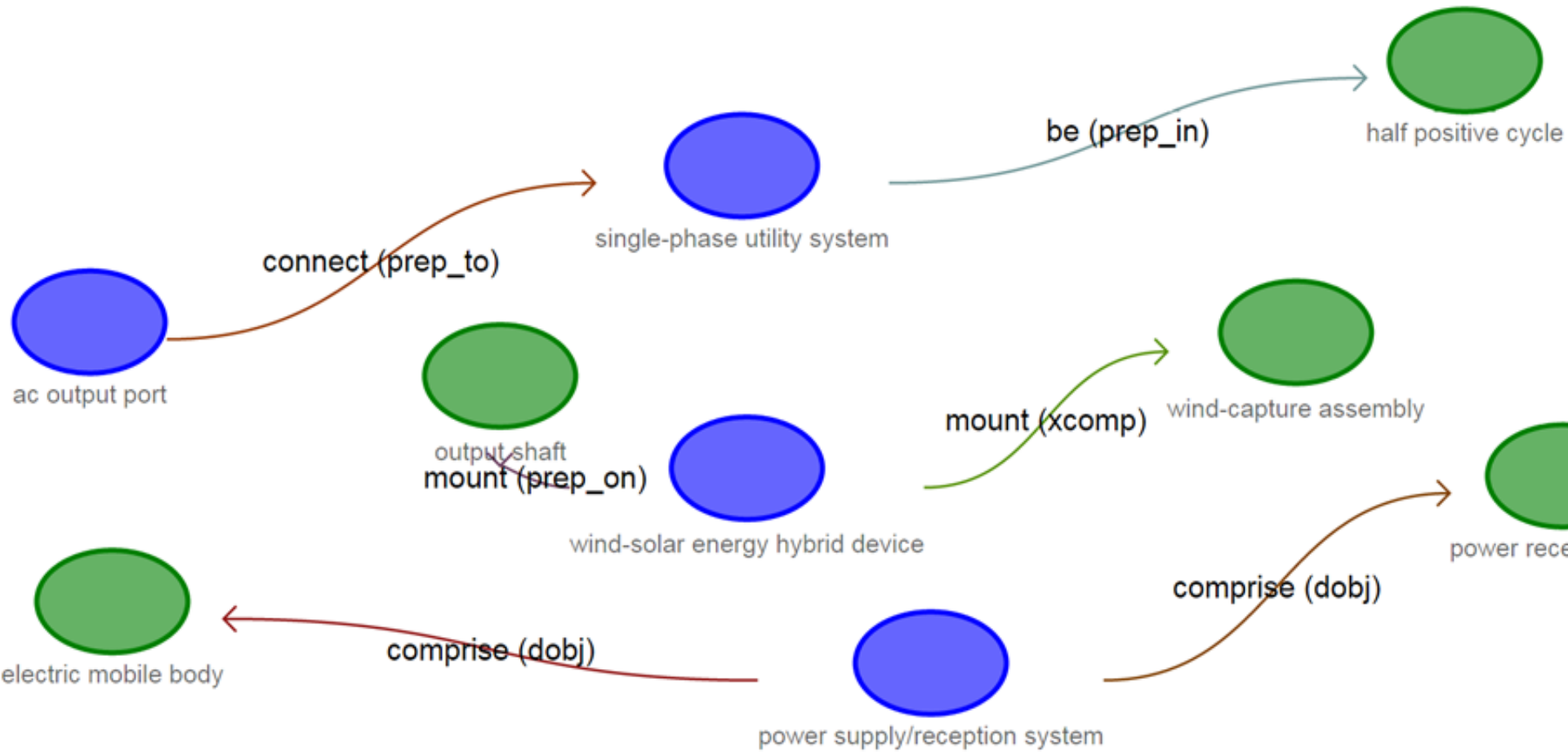
Visualization layer (2)



Solar energy materials and solar cells trends



Semantic layer



Conclusions



- Novel methodology for the identification of emerging trends and developments
- Dynamic term clustering based on the correlation of term distribution across time
- Cross-cluster analysis based on joint use of co-occurrence and correlation clusters
- Semantic analysis based on SAO to interpret the analytical outputs



Thank you for your attention.
Questions are appreciated!

Pavel Bakhtin, Ozcan Saritas

pbakhtin@hse.ru, osaritas@hse.ru

National Research University Higher School of Economics