Semantic networks in AutoMap

The research was prepared within the framework of the Academic Fund Program at the National Research University Higher School of Economics (HSE) in 2017 — 2018 (grant No. 17-05-0024) and by the Russian Academic Excellence Project "5-100".

http://www.zdes.spbu.ru/assets/files/20161218_SummerSchool.pdf

http://www.casos.cs.cmu.edu/projects/automap/

Semantic network

"A semantic network or net is a graphic notation for representing knowledge in patterns of interconnected nodes and arcs." (Sowa, 1992)

- Co-occurrence (collocation)
- Co-occurrence and relevance score (VOSviewer)
- Cosine similarity (Loet)
- Relational clustering
- Syntactic analysis
- Morphology
- Topic networks

Technical foundations of semantic networks

Adjacency matrix

same number of columns and rows

$$\boldsymbol{A} = \begin{array}{c} \\ \text{Jay} \\ \text{Adina} \\ \text{Iina} \\ \text{Noelle} \\ \text{Peter} \\ \vdots \end{array} \begin{array}{c} \begin{array}{cccccc} \text{Jay} & \text{Adina} & \text{Iina} & \text{Noelle} & \text{Peter} & \cdots \end{array} \\ \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & \cdots \\ 0 & 0 & 1 & 1 & 1 & \cdots \\ 0 & 1 & 0 & 0 & 1 & \cdots \\ 0 & 0 & 0 & 0 & 0 & \cdots \\ 0 & 0 & 1 & 0 & 0 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix} \end{array}$$

Bimodal (2-mode) networks

$$AO = \begin{array}{c} \\ \text{Jay} \\ \text{Adina} \\ \text{Iina} \\ \text{Noelle} \\ \text{Peter} \end{array} \begin{array}{c} \begin{array}{cccc} \text{CMU} & \text{VU} & \text{UvA} & \text{WUR} \end{array} \\ \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \end{bmatrix} \end{array}$$

2-mode to 1-mode transformation

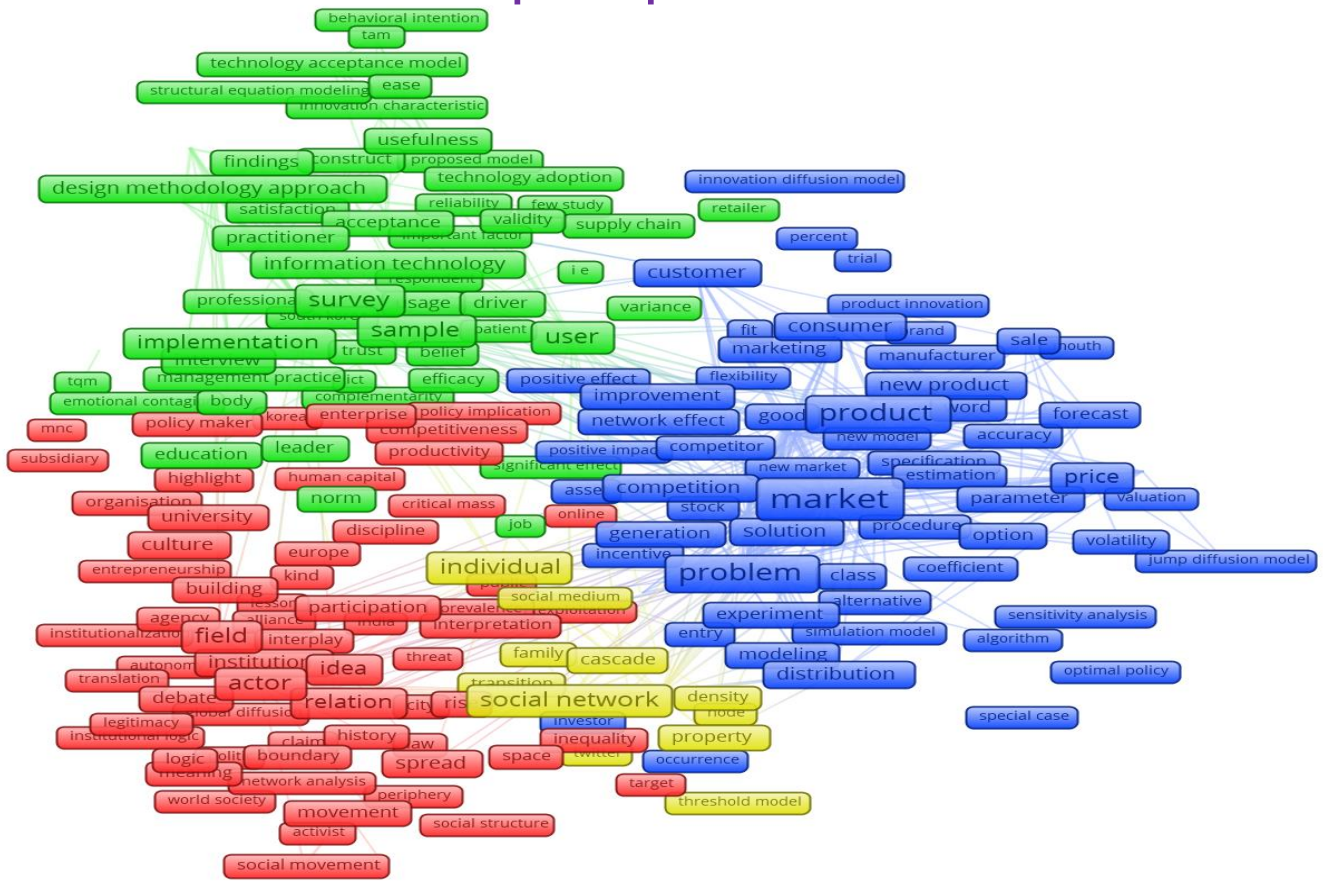
matrix multiplication

$$\begin{array}{c} \\ \text{Jay} \\ \text{Adina} \\ \text{Iina} \\ \text{Noelle} \\ \text{Peter} \end{array} \begin{array}{c} \begin{array}{cccc} \text{CMU} & \text{VU} & \text{UvA} & \text{WUR} \end{array} \\ \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \end{bmatrix} \end{array} \times \begin{array}{c} \\ \text{CMU} \\ \text{VU} \\ \text{UvA} \\ \text{WUR} \end{array} \begin{array}{c} \begin{array}{ccccc} \text{Jay} & \text{Adina} & \text{Iina} & \text{Noelle} & \text{Peter} \end{array} \\ \begin{bmatrix} 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 \end{bmatrix} \end{array}$$

Jay x Adina
= 1x1 + 1x1 + 0x0 + 0x1

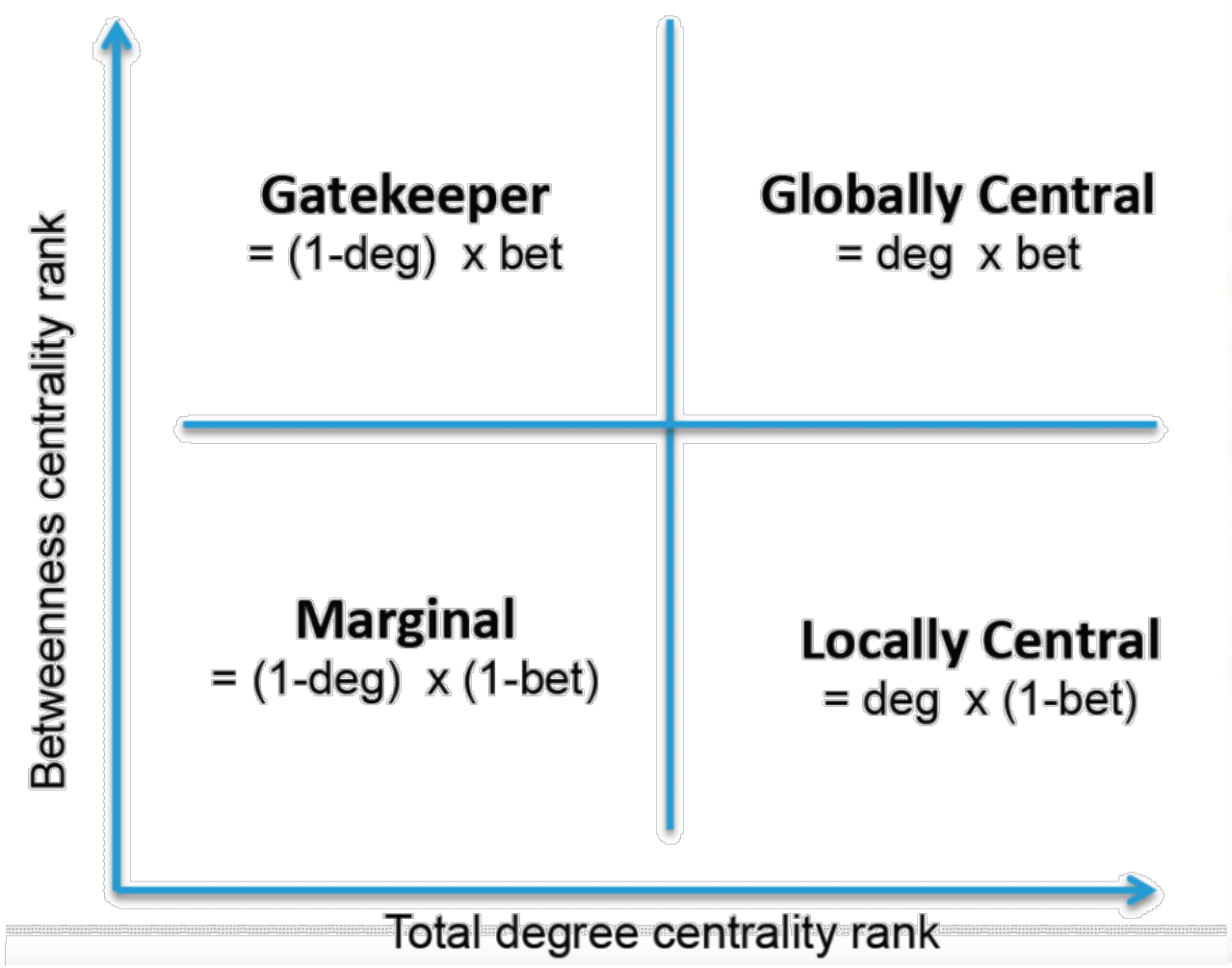
$$= \begin{array}{c} \\ \text{Jay} \\ \text{Adina} \\ \text{Iina} \\ \text{Noelle} \\ \text{Peter} \end{array} \begin{array}{c} \begin{array}{ccccc} \text{Jay} & \text{Adina} & \text{Iina} & \text{Noelle} & \text{Peter} \end{array} \\ \begin{bmatrix} 2 & 2 & 1 & 0 & 1 \\ 2 & 3 & 1 & 1 & 1 \\ 1 & 1 & 2 & 0 & 1 \\ 0 & 1 & 0 & 1 & 0 \\ 1 & 1 & 1 & 0 & 1 \end{bmatrix} \end{array}$$

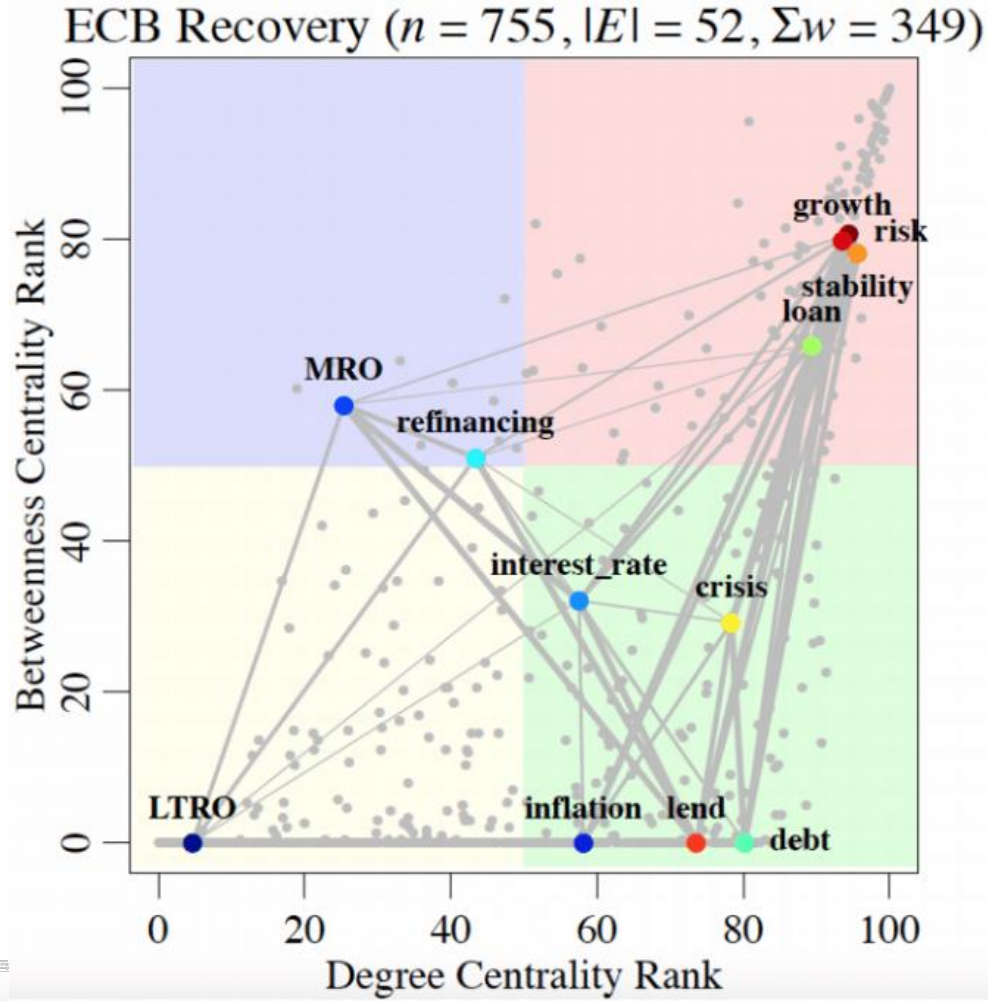
Semantic network of papers on social diffusion



Social movements *Knowledge diffusion*	*Organizations* *Innovation diffusion*	*Market* *New product diffusion*	*Social interaction* *Information diffusion*
activist, conflict, discourse, governance, nation	organizational adoption, human resource, technology adoption, emotional contagion	consumer, forecast, market	family, friends, information sharing, Twitter

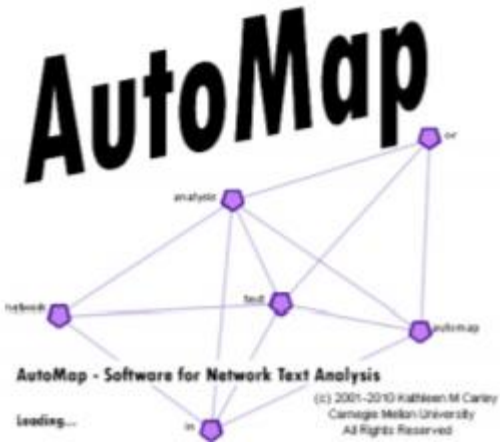
Combined centralities





Nerghes, A., Lee, J. S., Groenewegen, P., & Hellsten, I. (2015). Mapping discursive dynamics of the financial crisis: a structural perspective of concept roles in semantic networks. *Computational Social Networks*, *2* (1), 16.

AutoMap



AutoMap is a text mining tool that enables the extraction of network data from texts.

AutoMap can extract four types of information:

- content (concepts, frequencies and meta-data such as sentence length)
- semantic networks (concepts and relationships)
- meta-networks (ontologically coded concepts and relationships – named entities and links)
- sentiment and node attributes (attributes of named entities)

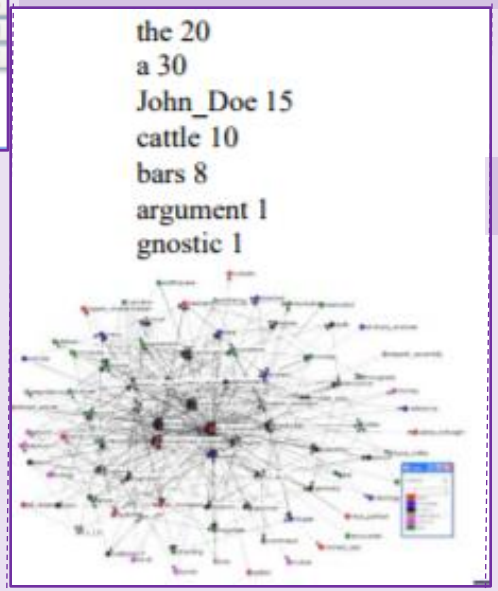
Natural Language Processing and Relational Extraction Routines in AutoMap

- Stemming
- Reduction and Normalization
- Part of Speech Tagging
- Anaphora Resolution
- Feature Identification
- Named Entity Extraction
- Ontological Text Coding
- Node and edge attributes
- Email Data Analysis
- Entropy Assessment
- Content Analysis

Unstructured Data: Texts

AutoMap: Data Extraction

CSV & DyNetML Output



Post Processing

- Event inference
- Belief inference
- Node attribute data fusion
- Lat-Lon data fusion

ORA: Analysis

Specialized Text analytics:

- Theme identification – LSA
- Hot Topics
- Semantic Network
- Parts of Speech
- Communicators
- Communicative Power

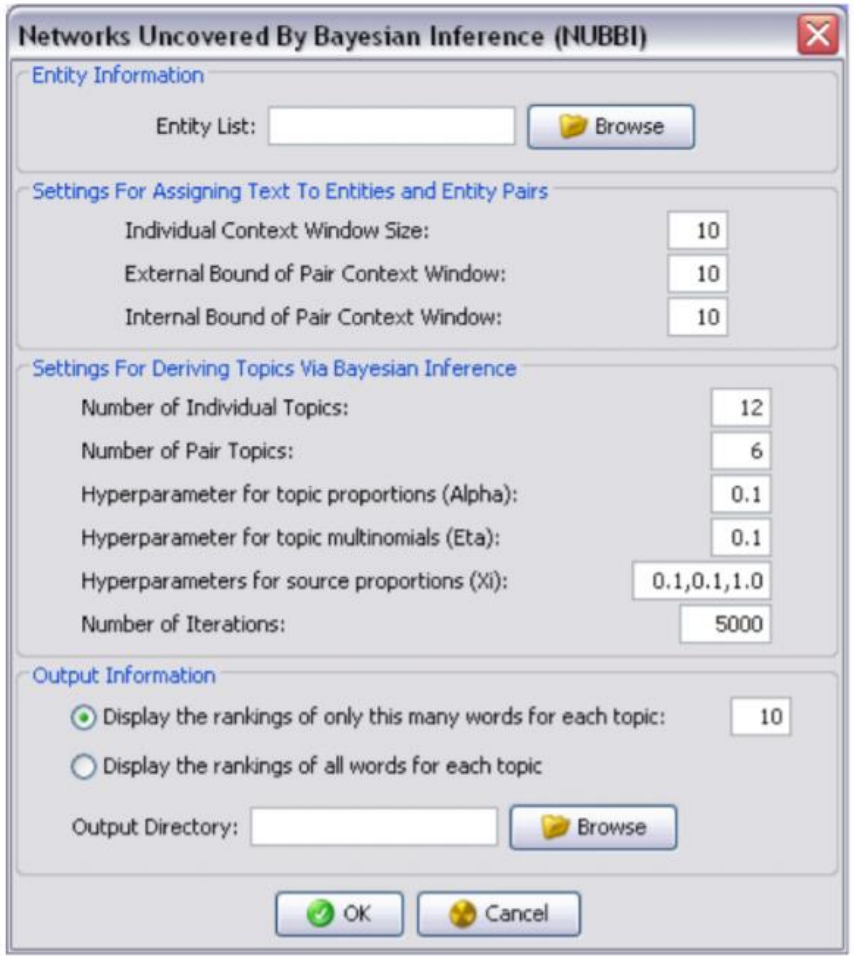
Text cleaning

1. **Text cleaning**:
   - *Remove extra spaces
   - *Convert British to American spelling
   - *Fix common typos
   - *Expand common contractions

2. **Text preparation:**
   - *Remove single letters
   - *N-gram conversion
   - *Remove pronouns
   - *Remove noise verbs
   - *Remove prepositions
   - *Remove All Noise Words
   - *Remove Day and Month words
   - *Remove Numbers as Words
   - *Remove Possessive Form
   - *Remove Complete Numbers

3. **Text refinement:**
   - *Remove Symbols (entirely)
   - *Replace HTML Symbols
   - *Apply Stemming (Kstemmer – Stem capitalization)

Topic modeling in AutoMap with NUBBI



nubbiNetworks.xml

- A collection of DyNetML networks and node classes.

Node classes

• There will be a node class for the specified class of each observed entity (as dictated in the entity list thesaurus)

• A node class for all of the different topics, called "Topics", and of type "Resource". Each node has a set of ranked words listed as 79 attributes. From this set of attributes, you can develop an appropriate contextual label for each topic.

• A node class for all of the different concepts, called "Concepts" and of type "Knowledge".