Scientific Evolution of Artificial Intelligence in Educational Field Using Text Mining

Text mining technics applied over bibliographic material

Federico Ferrero

EPPS 6302 Methods of Data Collection and Production / Prof. Karl Ho

Research Questions

What are the main topics of Artificial Intelligence investigated in the field of educational research during the last three decades?

Which is its thematic evolution during this period?

Methodology

- **Exploratory exercise with SciMAT:**
  - Science Mapping Analysis software tool (Cobo, López-Herrera, Herrera-Viedma, and Herrera, 2011).
  - Bibliometric science mapping tool based on co-word analysis and h-index. It works in a longitudinal framework in order to detect the different themes treated by the research field across the given time periods.

- **STEPS:**

  1. Collection of the raw data: search in Web of Science Core Collection (ISI WoS) according to the criteria:

  (TS=((artificial intelligence OR success student algorithm OR intelligent tutoring systems OR big data) AND education)) AND DOCUMENT TYPES: (Article) Refined by: WEB OF SCIENCE CATEGORIES about EDUCATION. Timespan: 1980-present. Indexes: SSCI, A&HCI, CPCI-SSH, BKCI-SSH.

  Final corpus: **N= 770** documents.

  Partition periods every **5 years**.

Methodology

- **STEPS:**

2. Normalization of keywords

3. Co-occurrence frequencies of keywords and similarities between items (Equivalence Index)

4. Clustering subgroups of keywords that are solidly linked.

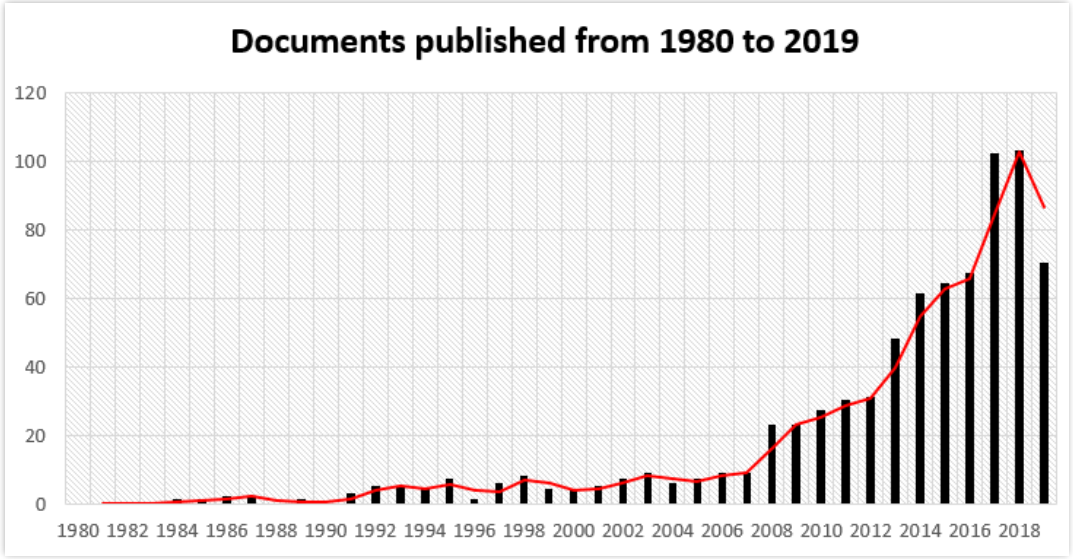
5. Analysis and interpretation:

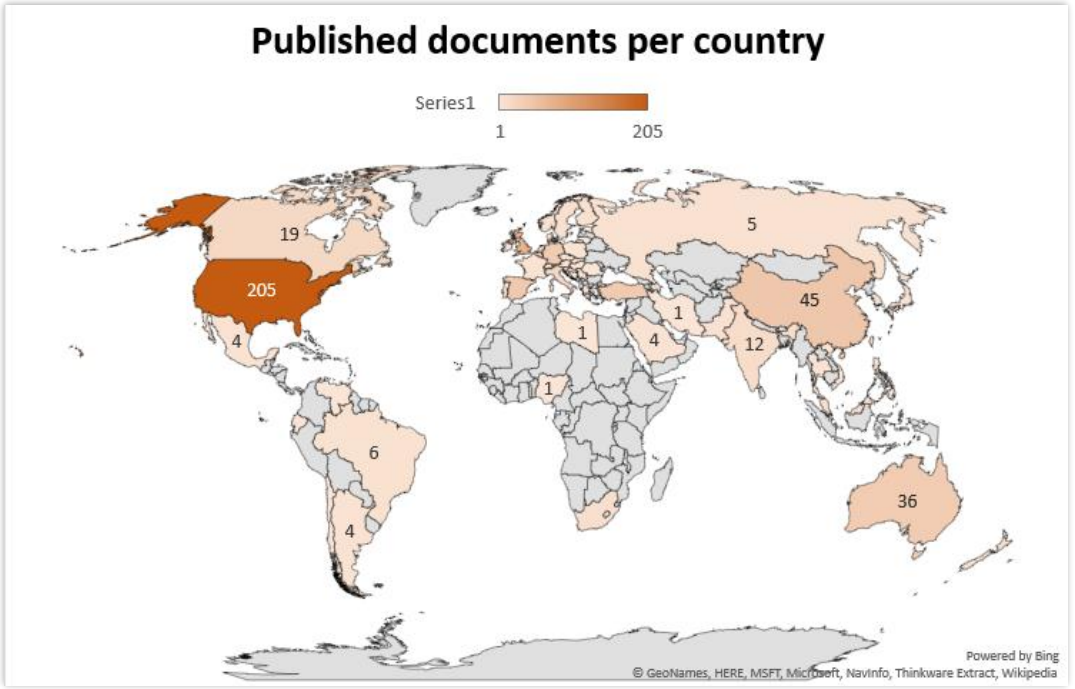
1. *Strategic diagram*
2. *Stability between periods diagram*
3. *Thematic evolution diagram*
4. *Thematic diagram*

Methodology

- **Analysis Configuration**
  - Unit of analysis: Keywords
  - Kind of network: Co-occurence
  - Normalization measure: Equivalence index (0-1)
  - Cluster algorithm: Centers simples
    - Max cluster size: 6
    - Min cluster size: 3
  - Evolution measure: Jaccard index
  - Overlapping measure: Inclusion index

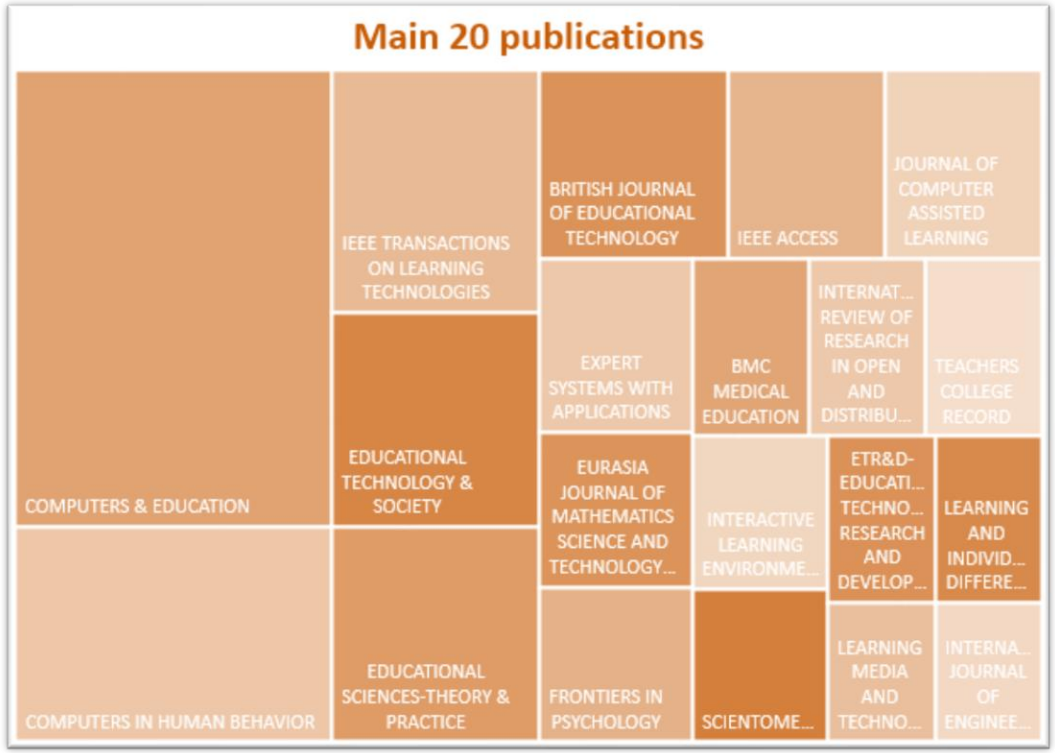
- **Period's results**
  1. 1980-1985
  2. 1986-1990
  3. 1991-1995
  4. 1996-2000
  5. 2001-2005
  6. 2006-2010
  7. 2011-2015
  8. 2016-2019

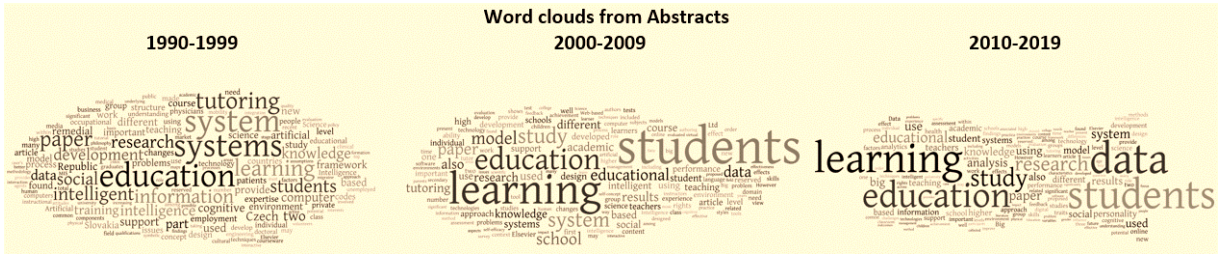




- **First record:** 1984
- **Explosion** from 2008

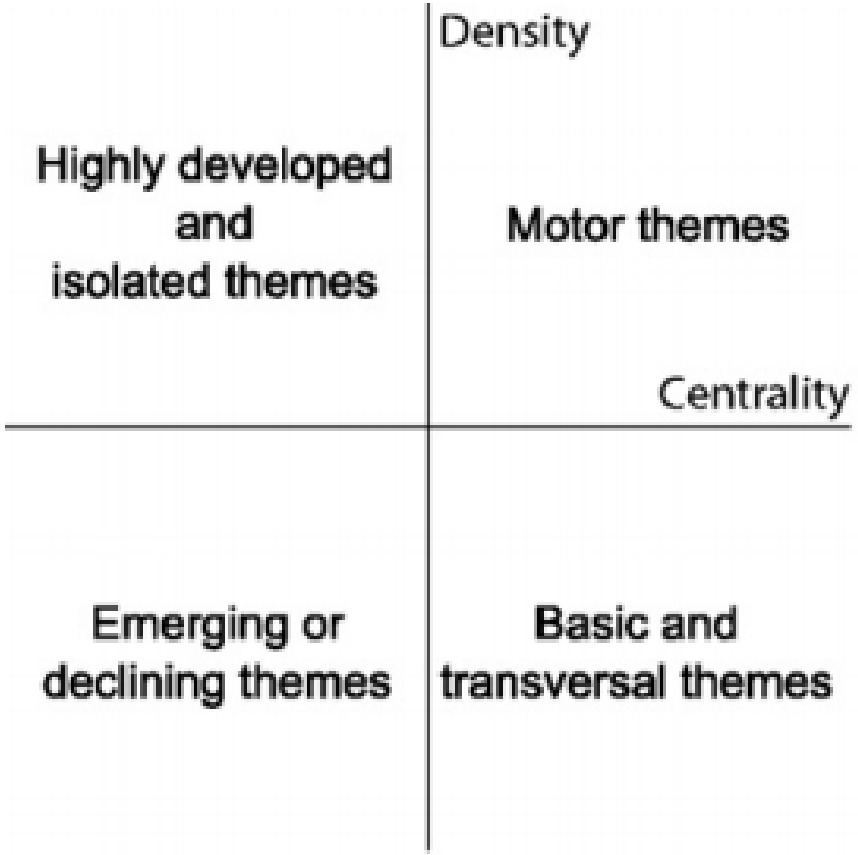
- **Main countries**: USA, UK, China and Spain

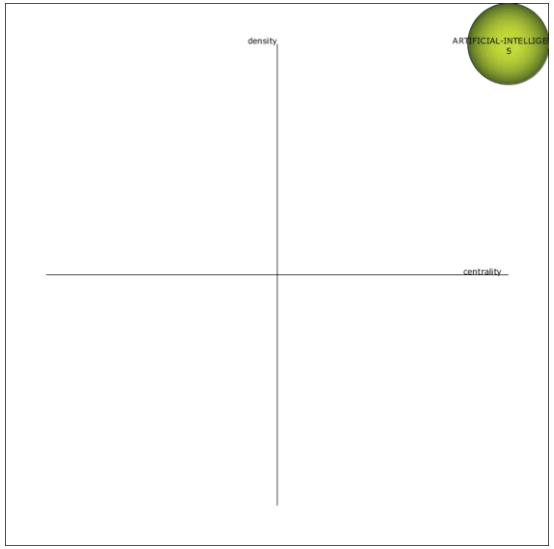




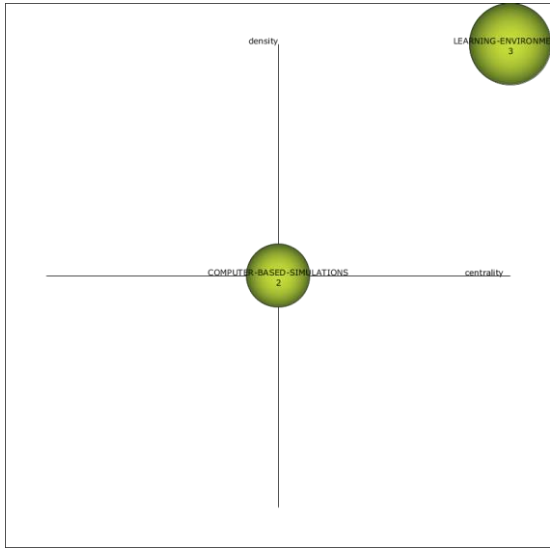
Intelligent systems > learning & students > learning & data

Strategic diagram quadrants

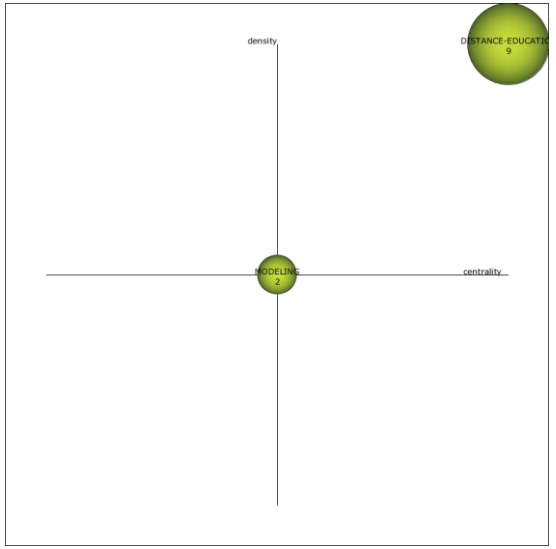




Strategic diagram 1991-1995

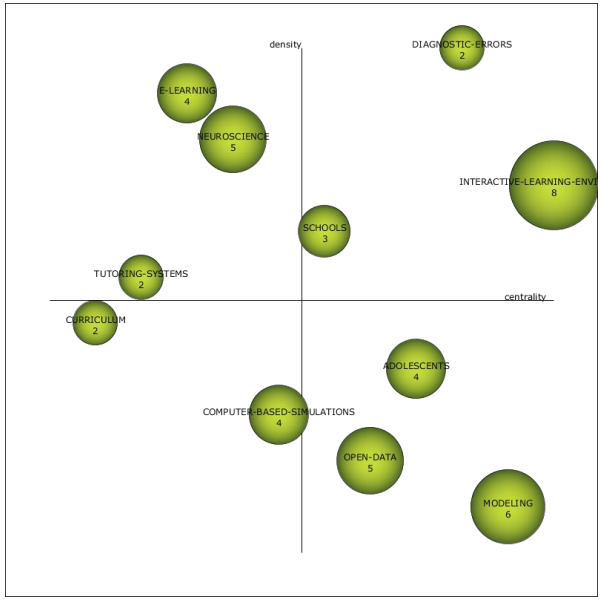


Strategic diagram 1996-2000

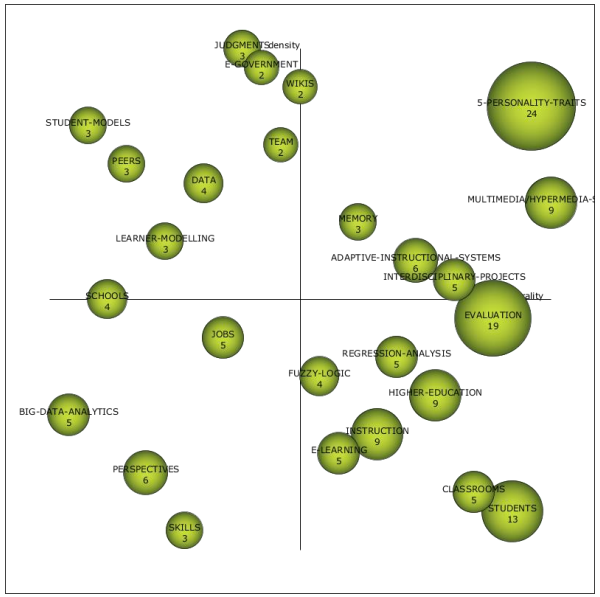


Strategic diagram 2001-2005

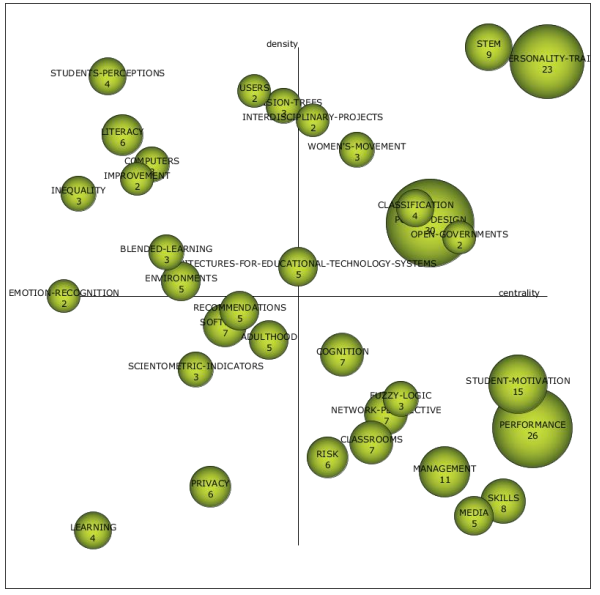
Evolution Strategic Diagrams (1991-2019)



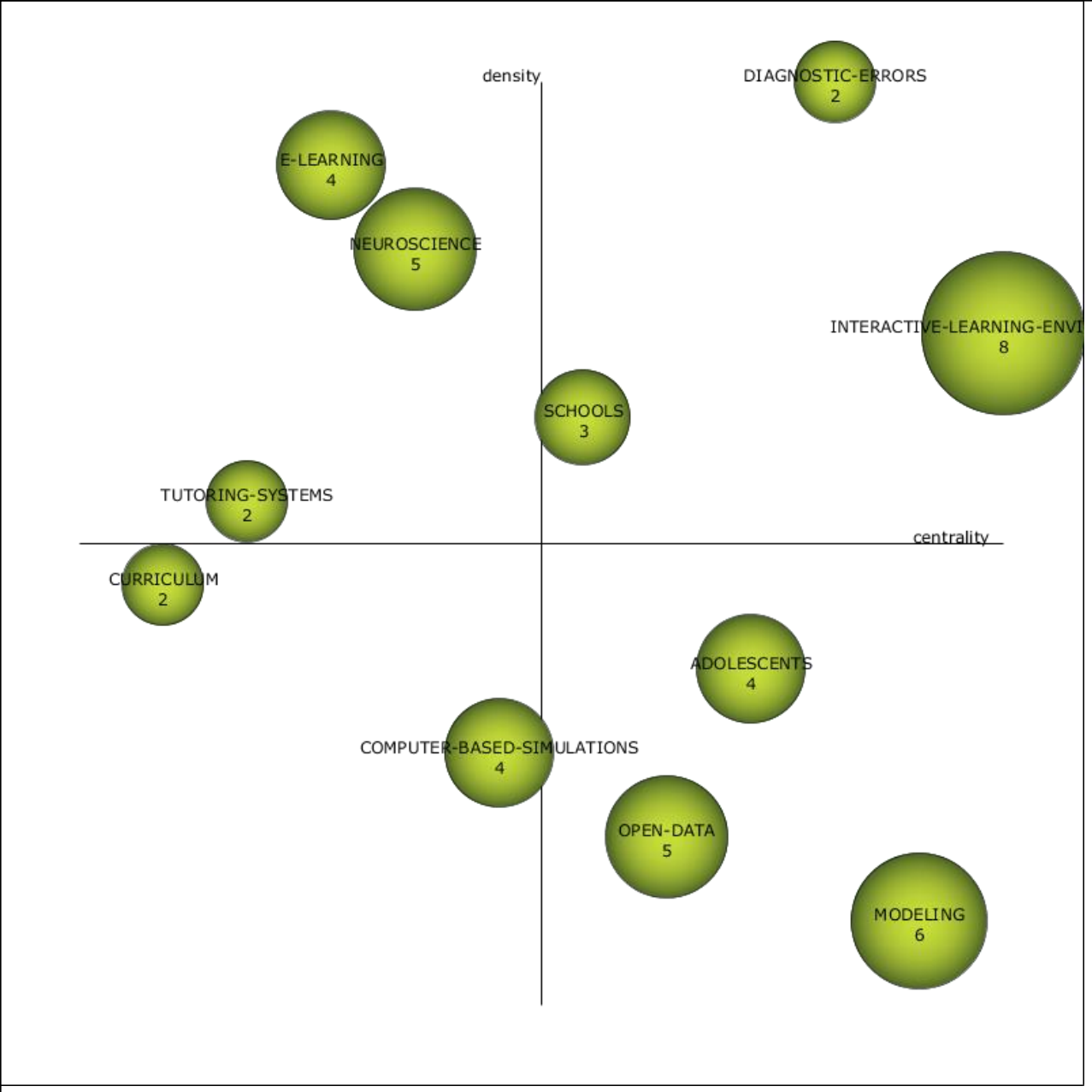
Strategic diagram 2006-2010



Strategic diagram 2011-2015



Strategic diagram 2016-2019

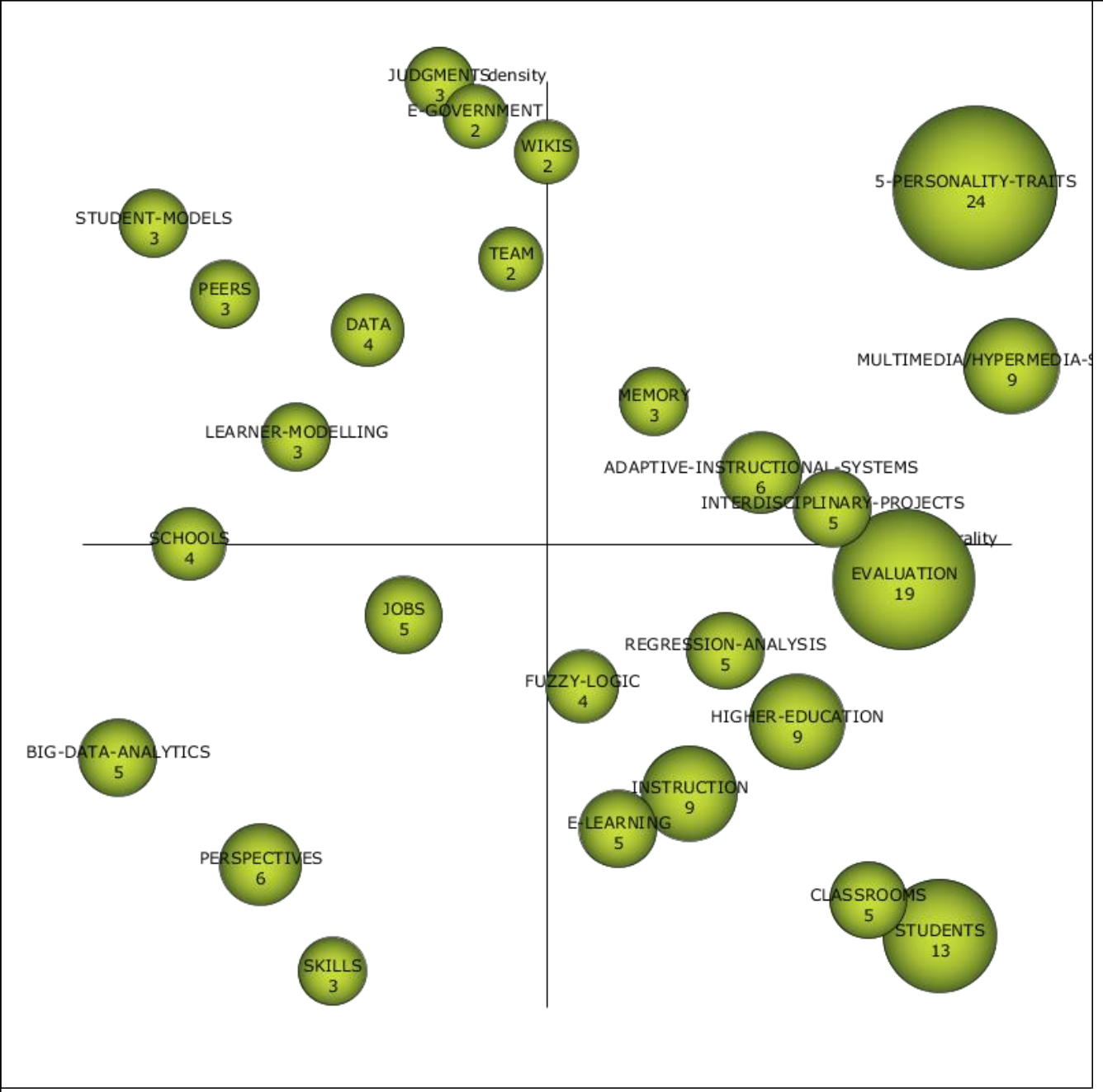


Strategic diagram 2006-2010

- Explosion in thematic diversification.

- **MOTOR**: **Interactive-Learning-Environments**.
- **BASIC**: Modeling, Open Data.
- **ISOLATED**: Neuroscience, E-Learning.
- **EMERGING**: Computer-Based-Simulations.

Period 2006-2010			
Theme Name	Number of Documents	h-Index	Number of citations
INTERACTIVE-LEARNING-ENVIRONMENTS	8	7	203
MODELING	6	4	121
OPEN-DATA	5	5	164
NEUROSCIENCE	5	4	154
E-LEARNING	4	4	228
COMPUTER-BASED-SIMULATIONS	4	3	21
ADOLESCENTS	4	4	67
SCHOOLS	3	3	36
TUTORING-SYSTEMS	2	2	89
DIAGNOSTIC-ERRORS	2	2	23
CURRICULUM	2	1	44

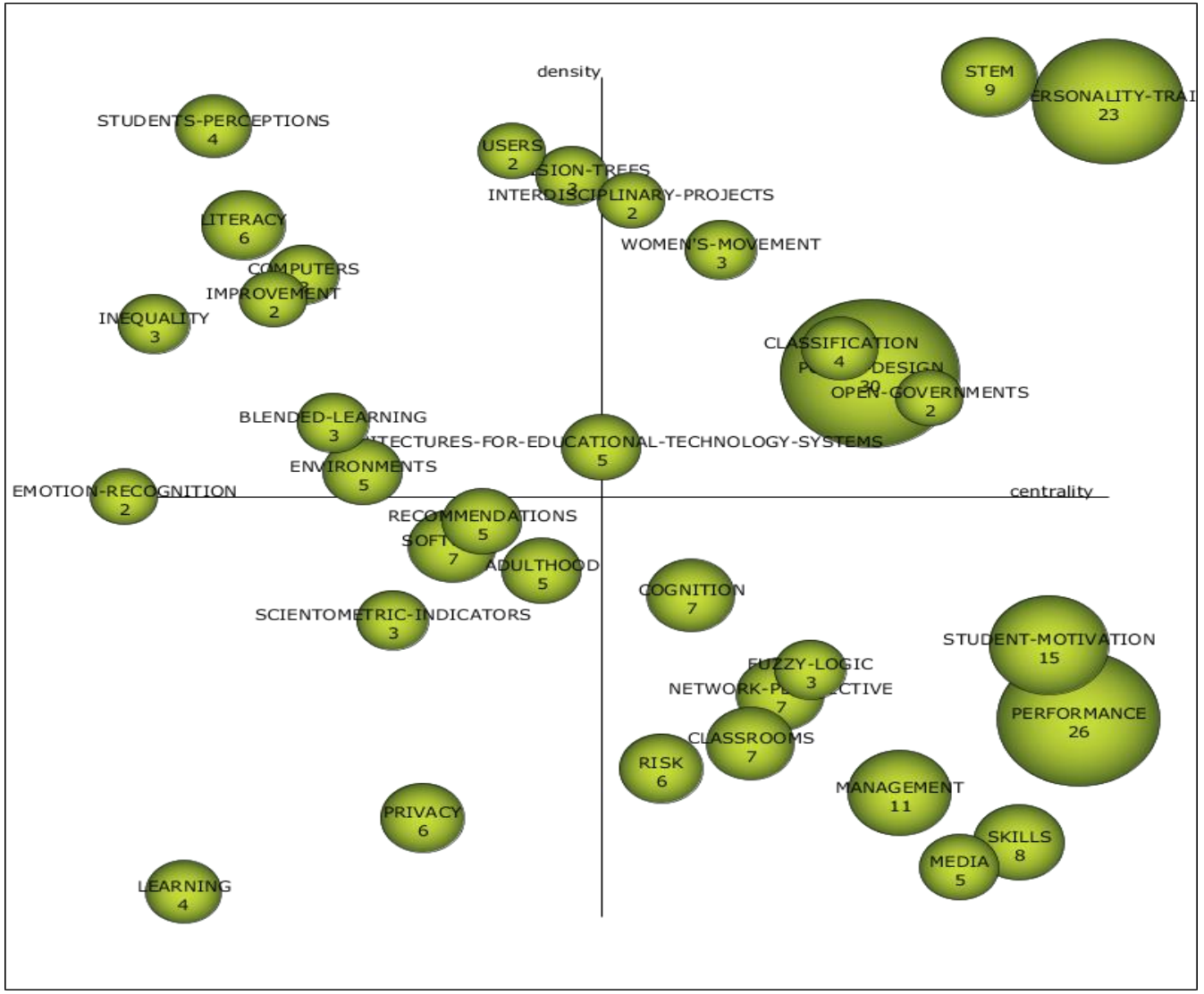


Strategic diagram
2011-2015

- Sustained thematic proliferation: 26 themes (15 more).

- **MOTOR**: **5-Personality-Traits** **(h-index and citations).**
- **BASIC**: Evaluation.
- **ISOLATED**: Students-Models, Learner-Modelling.
- **EMERGING**: Perspectives, Big-Data-Analytics.

Period 2011-2015			
Theme Name	Number of Documents	h-Index	Number of citations
5-PERSONALITY-TRAITS	24	12	422
EVALUATION	19	10	291
STUDENTS	13	7	326
MULTIMEDIA/HYPERMEDIA-SYSTEMS	9	8	218
HIGHER-EDUCATION	9	6	210
INSTRUCTION	9	7	211
ADAPTIVE-INSTRUCTIONAL-SYSTEMS	6	5	154
PERSPECTIVES	6	4	162
INTERDISCIPLINARY-PROJECTS	5	5	137
REGRESSION-ANALYSIS	5	5	102
JOBS	5	4	166
CLASSROOMS	5	3	30
E-LEARNING	5	4	88
BIG-DATA-ANALYTICS	5	5	269
FUZZY-LOGIC	4	4	51
SCHOOLS	4	4	64
DATA	4	3	40
JUDGMENTS	3	2	31
STUDENT-MODELS	3	2	113
MEMORY	3	2	67
PEERS	3	3	184
LEARNER-MODELLING	3	3	86
SKILLS	3	3	17
E-GOVERNMENT	2	2	97
WIKIS	2	1	41
TEAM	2	1	6

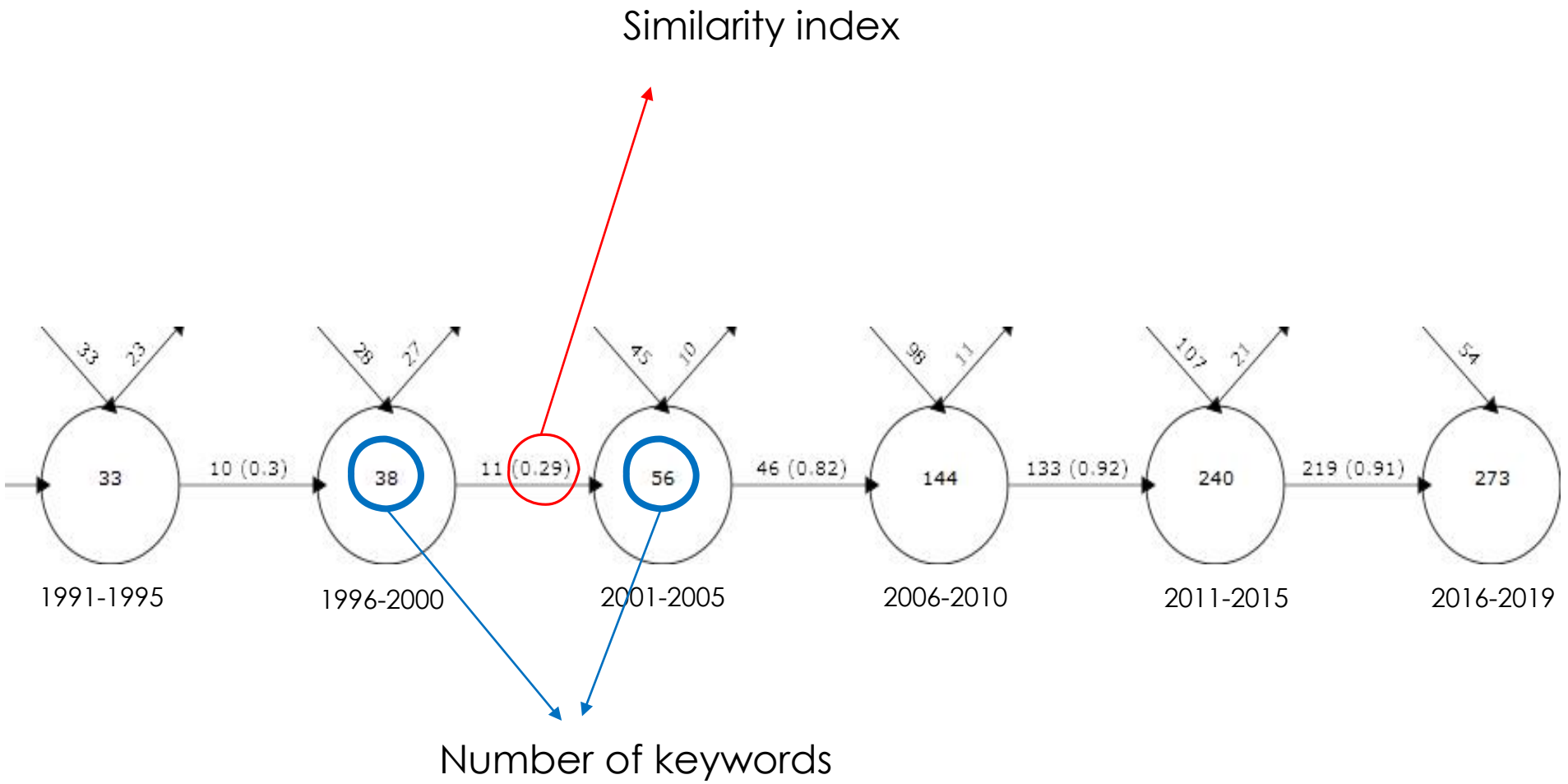


Strategic diagram
2016-2019

- Sustained thematic proliferation: 34 themes.
- **MOTOR**: **Policy Design** **(citations).**
- **BASIC**: Performance.
- **ISOLATED**: multiple themes.
- **EMERGING**: Privacy.

Period 2016-2019			
Theme Name	Number of Documents	h-Index	Number of citations
POLICY-DESIGN	30	7	190
PERFORMANCE	26	5	86
5-PERSONALITY-TRAITS	23	6	89
STUDENT-MOTIVATION	15	4	58
MANAGEMENT	11	3	27
STEM	9	4	50
SKILLS	8	4	57
SOFTWARE	7	3	88
NETWORK-PERSPECTIVE	7	3	29
COGNITION	7	1	6
CLASSROOMS	7	3	20
LITERACY	6	1	13
RISK	6	1	6
PRIVACY	6	3	44
ADULTHOOD	5	2	12
ARCHITECTURES-FOR-EDUCATIONAL-TECHNOLOGY-SYSTEMS	5	2	18
RECOMMENDATIONS	5	1	14
MEDIA	5	2	23
ENVIRONMENTS	5	3	23
CLASSIFICATION	4	3	40
STUDENTS-PERCEPTIONS	4	2	32
LEARNING	4	2	17
WOMEN'S-MOVEMENT	3	1	5
COMPUTERS	3	2	14
DECISION-TREES	3	2	17
BLENDED-LEARNING	3	2	10
INEQUALITY	3	1	4
FUZZY-LOGIC	3	1	2
SCIENTOMETRIC-INDICATORS	3	3	20
INTERDISCIPLINARY-PROJECTS	2	2	4
IMPROVEMENT	2	2	11
USERS	2	1	1
OPEN-GOVERNMENTS	2	2	7
EMOTION-RECOGNITION	2	1	1

- **# of keywords increases**: **33** (1991-1995) to **273** (2016-2019).
- **Jump in 2006-2010**: 56 to 144.
- **Similarity Index increases** from 0.3 to 0.91: terminology is shared and maintained while the research field is consolidated.



Overlapping map
(stability across consecutive periods)

Not only **thematic proliferation** but also incipient **maintenance of interest** in specific themes (especially 2011-2015 and 2016-2019).

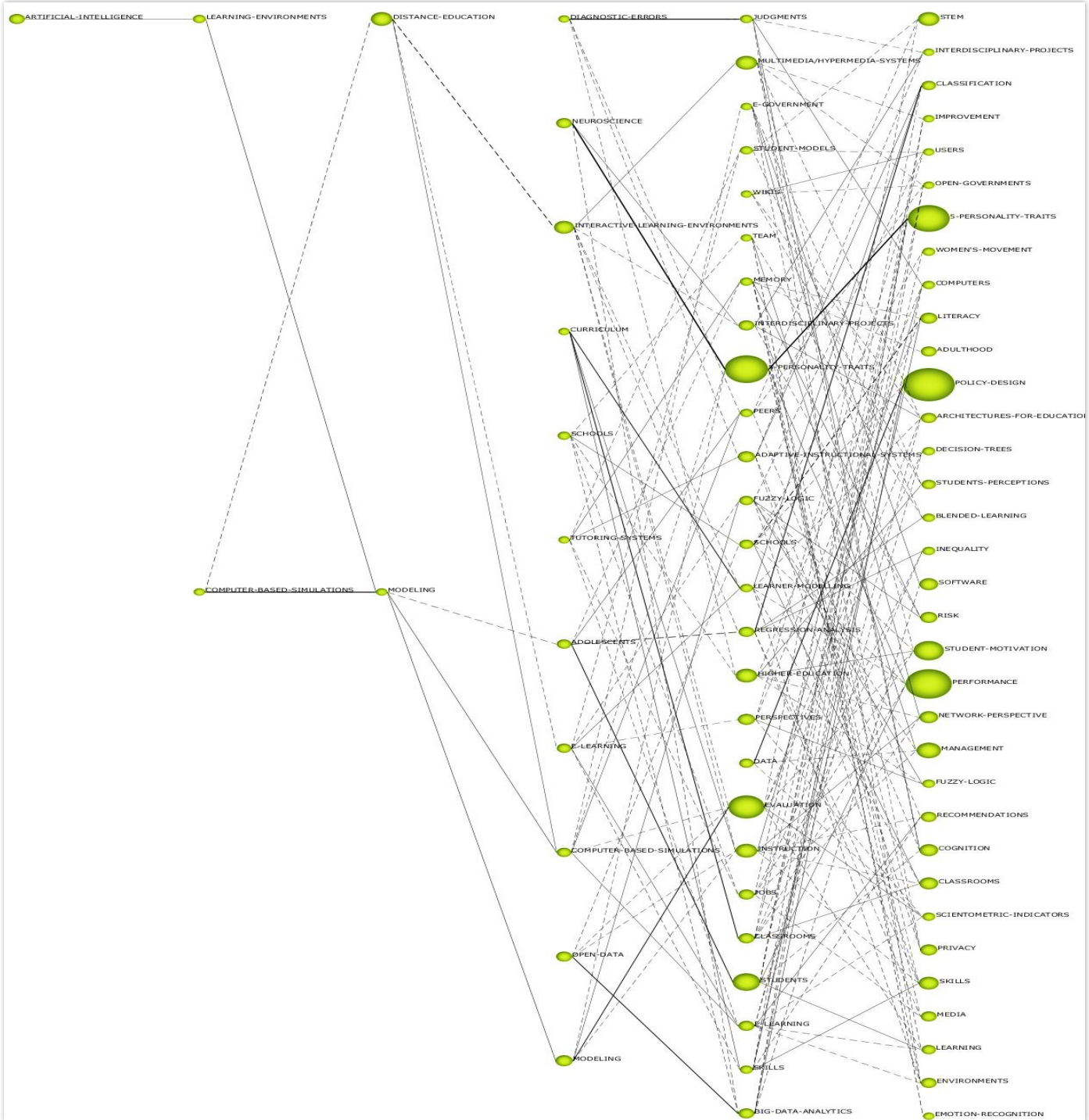
Repeated Themes				
Theme Name	Number of Documents	h-Index	Number of citations	Repeated period
COMPUTER-BASED-SIMULATIONS	2	2	27	1996-2000 & 2006-2010
MODELING	2	2	122	2001-2005 & 2006-2010
E-LEARNING	4	4	228	2006-2010 & 2011-2015
SCHOOLS	3	3	36	2006-2010 & 2011-2015
5-PERSONALITY-TRAITS	24	12	422	2011-2015 & 2016-2019
INTERDISCIPLINARY-PROJECTS	5	5	137	2011-2015 & 2016-2019
FUZZY-LOGIC	4	4	51	2011-2015 & 2016-2019
CLASSROOMS	5	3	30	2011-2015 & 2016-2019
SKILLS	3	3	17	2011-2015 & 2016-2019

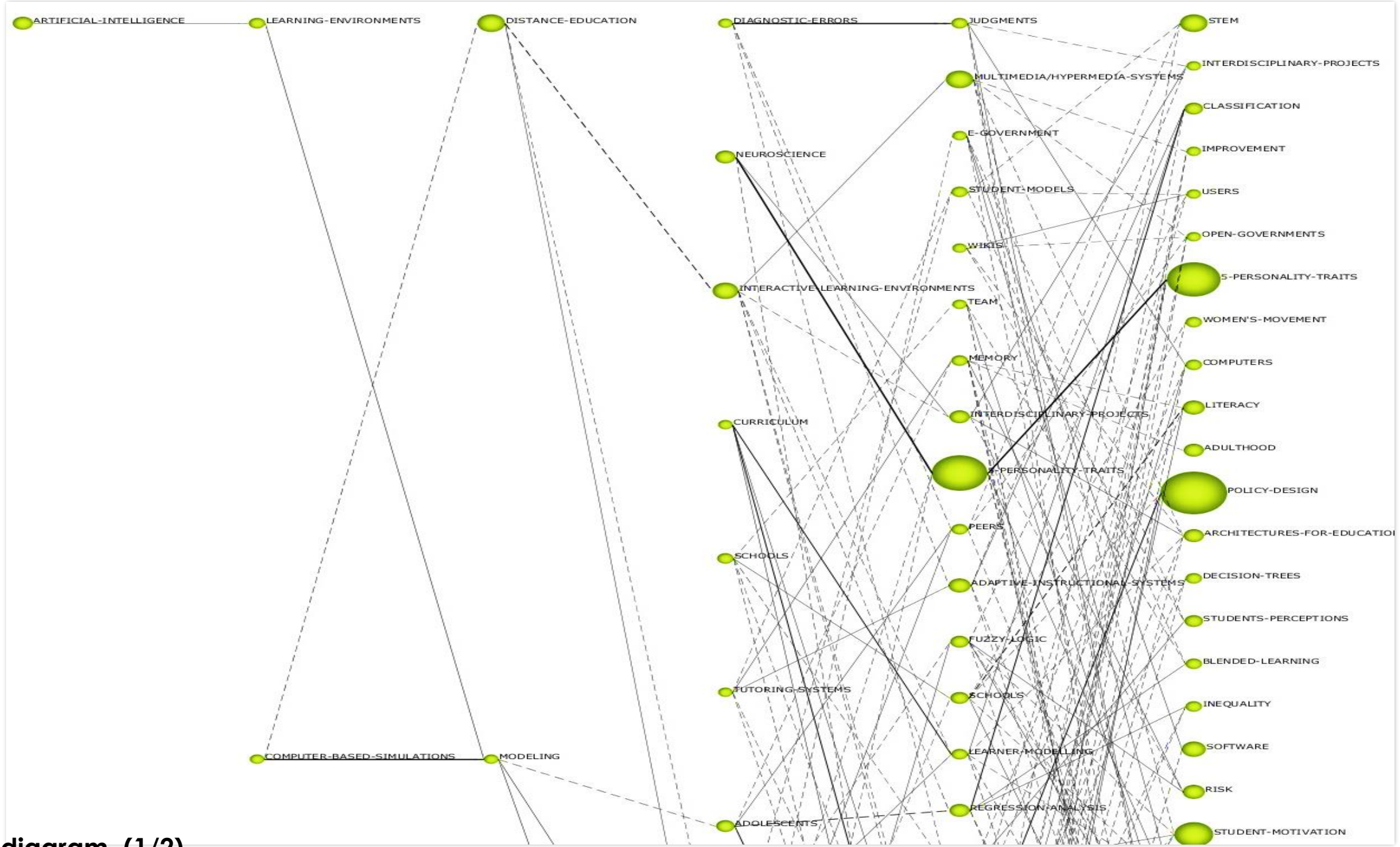
Evolution diagram *(complete)*

- **High density of links** especially among the last three periods.

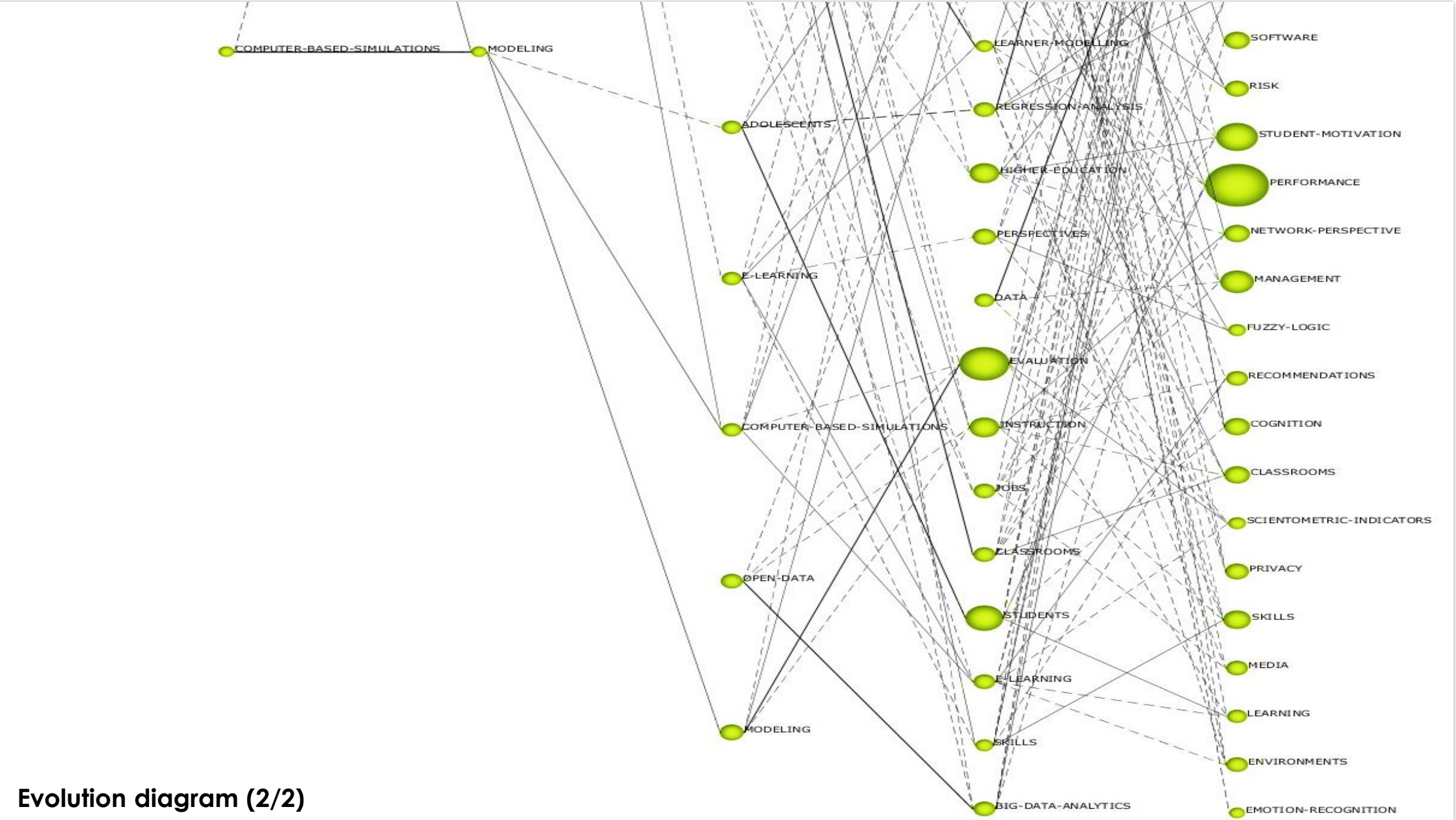
- **Dotted lines predominate**: which prevents talking about consolidated lines of interest for research in the field.
- **"PERFORMANCE"** is nurtured by several previous topics.

- Some continuous thematic areas expressed by **solid lines**:
  - "NEUROSCIENCE" > "5-PERSONALITY-TRAITS" > "5-PERSONALITY-TRAITS"
  - "DATA" > "POLICY-DESIGN"
  - "MODELING" > "EVALUATION"

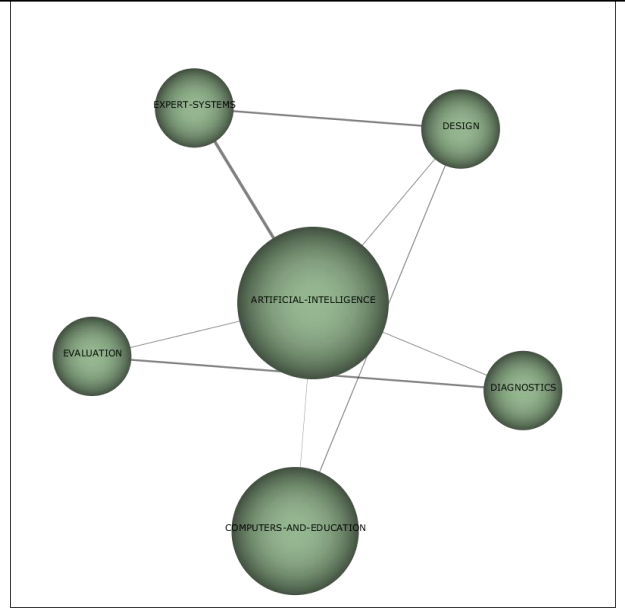




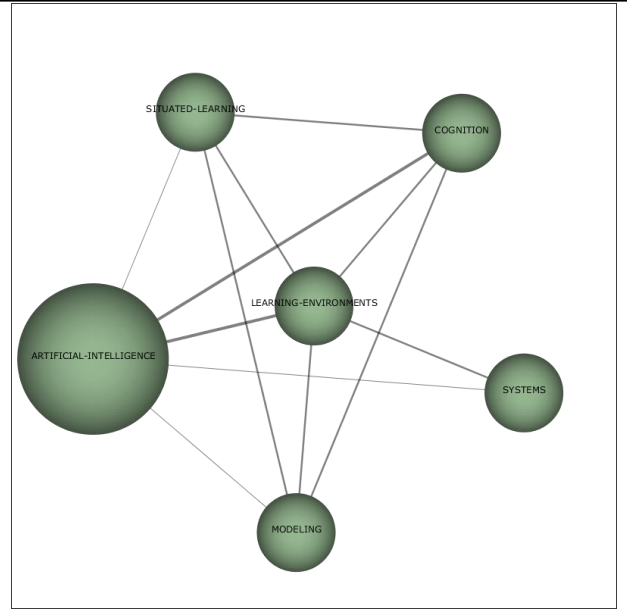
Evolution diagram (1/2)



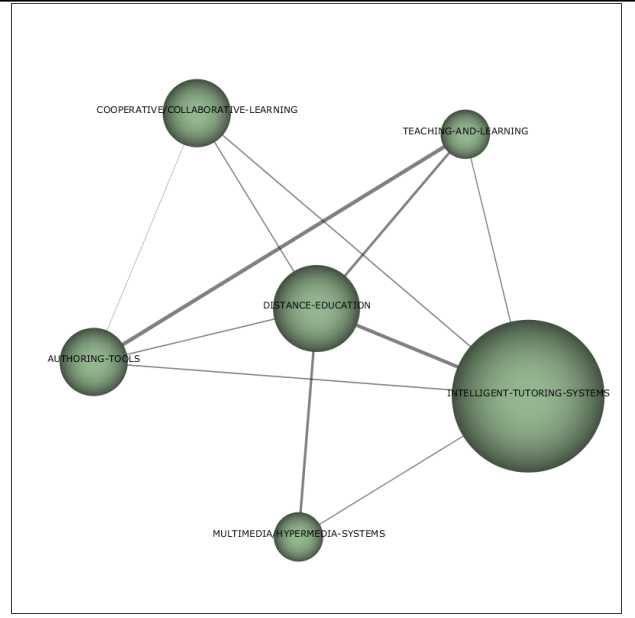
Evolution diagram (2/2)



ARTIFICIAL INTELLIGENCE
1991-1995

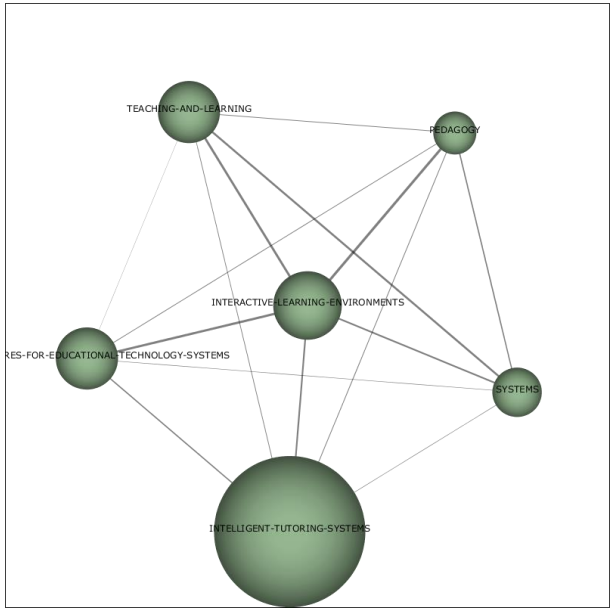


LEARNING ENVIRONMENTS
1996-2000

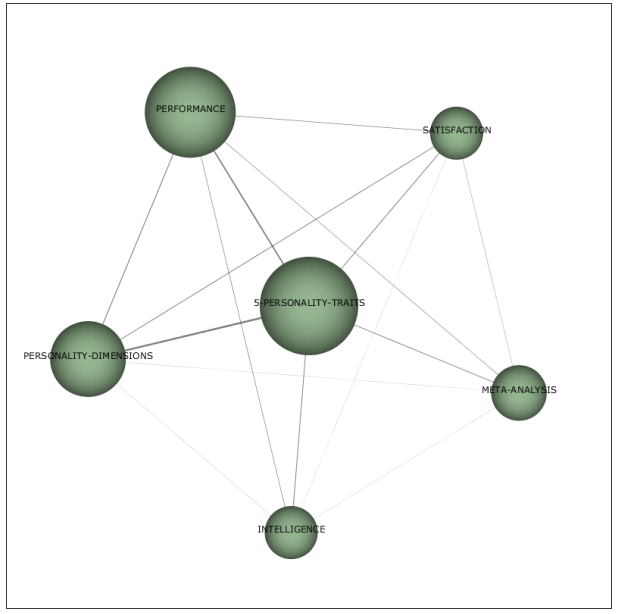


DISTANCE EDUCATION
2001-2005

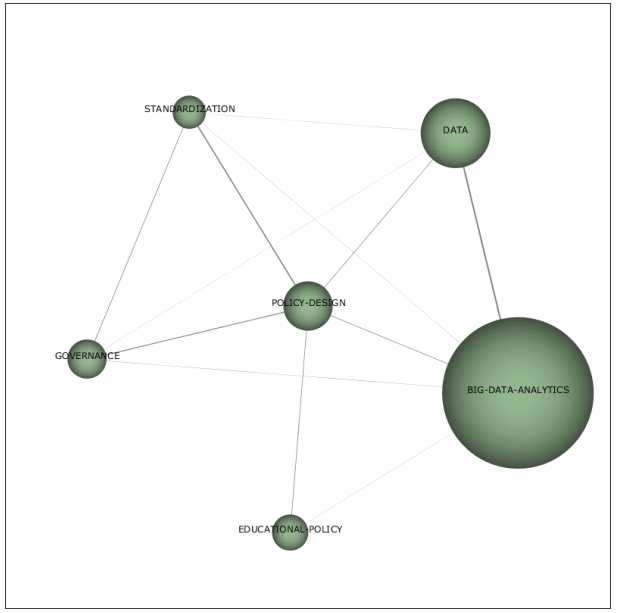
Keywords configuration per each main theme per period (1991-2019)



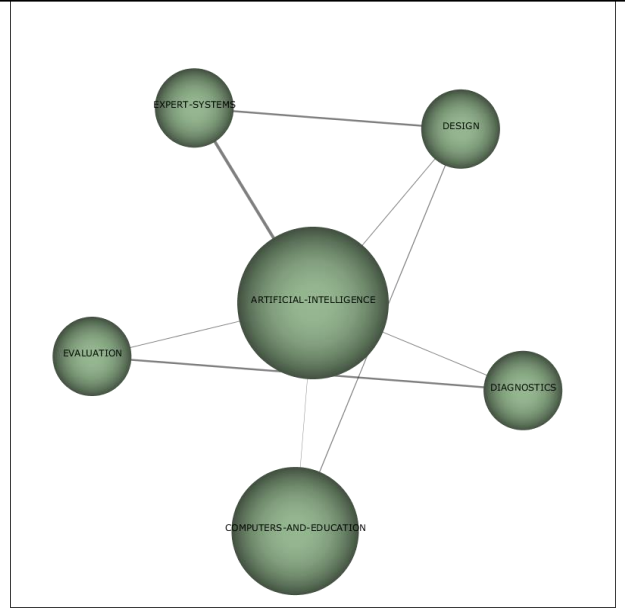
INTERACTIVE LEARNING ENVIRONMENTS
2006-2010



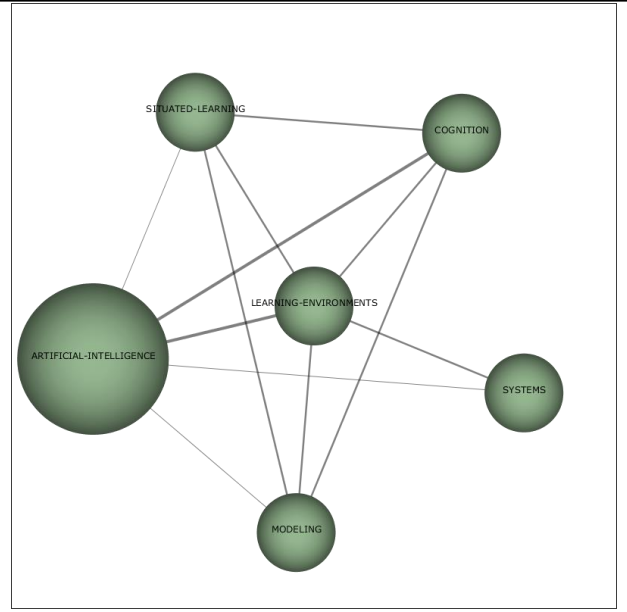
5-PERSONALITY TRAITS
2011-2015



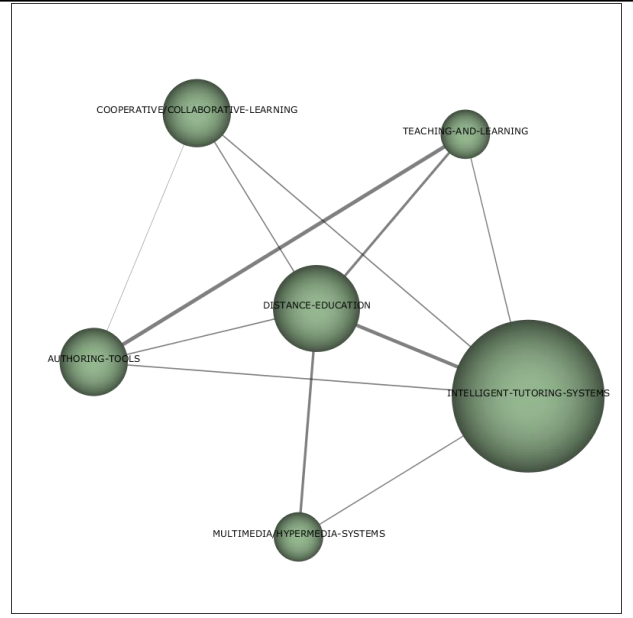
POLICY DESIGN
2016-2019



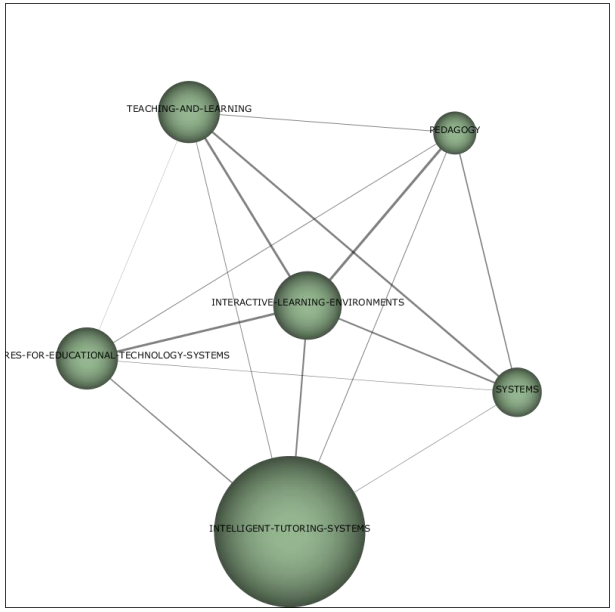
ARTIFICIAL INTELLIGENCE
1991-1995



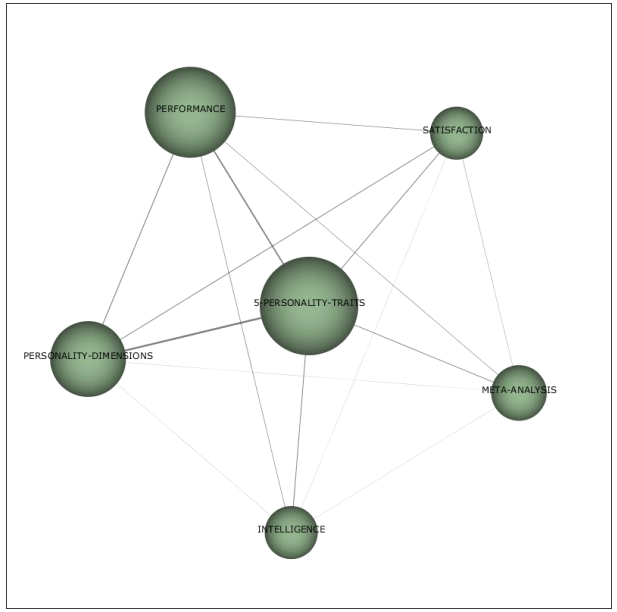
LEARNING ENVIRONMENTS
1996-2000



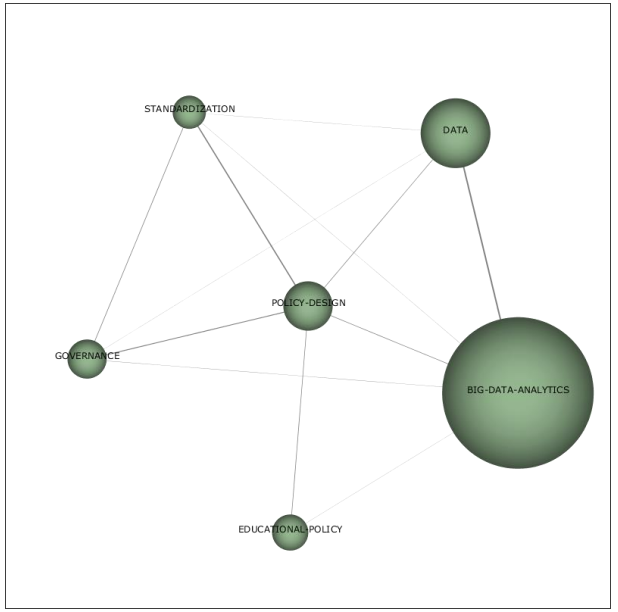
DISTANCE EDUCATION
2001-2005



INTERACTIVE LEARNING ENVIRONMENTS
2006-2010



5-PERSONALITY TRAITS
2011-2015



POLICY DESIGN
2016-2019

1. Focus is on **artificial intelligence systems** in education with the domain of Computer Sciences topics.
2. A more pedagogical concern: **learning** and **distance education**.
3. **Intelligent systems to detect personality traits**.
4. Recently: **policy design** using large amounts of **data**.

Conclusions

1. **AI academic production** in the educational field **prevails in the last decade**, with sustained growth.
2. **Main countries**: USA, UK, China and Spain.
3. **Drastic thematic proliferation** (especially among the last 3 periods) with established core themes.
4. **I**ncipient **consolidation** of research field (especially 2 periods) with solid lines of investigation.
5. Two **thematic circuits** (one more technical and the other pedagogical/psychological) with different emphasis and with different crossings during the periods.
   - Focus is on Computer Sciences topics *(1991-1995)*.
   - A more pedagogical concern (1996-2010).
   - Intelligent systems to detect personality traits (2011-2015).
   - Policy design using large amounts of data (2016-2019).

Bibliography

- A. Breiter, "Datafication in education: a multi-level challenge for IT in educational management," in Stakeholders and Information Technology in Education, T. Brinda, N. Mavengere, I. Haukijarvi, C. Lewin, D. Passey Eds. Berlin: Springer, 2016, pp. 95-103.
- B. K. Daniel, Ed., Big data and Learning Analytics in higher education: current theory and practice. Switzerland: Springer International Publishing, 2017.
- B. Williamson, Big data in education: the digital future of learning, policy and practice. London: SAGE, 2017.
- D. Kehl, P. Guo and S. Kessler, Algorithms in the criminal justice system: assessing the use of risk assessments in sentencing. Berkman Klein Center for Internet & Society, Harvard Law School, 2017.
- E. Siegel, Predictive Analytics. The power to predict who will click, buy, lie or die. New Jersey: John Wiley and Sons, 2016.
- G-H. Kim, S. Trimi, J-H. Chung, "Big-Data Applications in the Government Sector," Communications of the ACM, vol. 57, num. 3, pp. 78-85, 2014.
- J. Danaher, M. J. Hogan, C. Noone, R. Kennedy, A. Behan, A. De Paor, H. Felzmann, M. Haklay, S. Khoo, J. Morison, M. H. Murphy, N. O'Brolchain, B. Schafer and K. Shankar, "Algorithmic governance: Developing a research agenda through the power of collective intelligence," Big Data & Society, pp. 1–21, 2017.
- J. Van Dijck, "Datafication, dataism and dataveillance: big data between scientific paradigm and ideology," Surveillance & Society, vol. 12, num. 2, pp. 197-208, 2014.
- K. Jee, G-H. Kim, "Potentiality of big data in the medical sector: focus on how to reshape the healthcare system," Healthc Inform Res, vol. 19, num. 2, pp. 79-85, 2013.
- M. Batty, "Big data, smart cities and city planning," Dialogues in Human Geography, vol. 3, num. 3, pp. 274–279, 2013.
- M.A Martínez, M.J. Cobo, M. Herrera and E. Herrera-Viedma, "Analyzing the Scientific Evolution of Social Work Using Science Mapping", Research on Social Work Practice, vol. 25, num. 2, pp. 257-277, 2015.
- M.J. Cobo, A.G. López-Herrera, E. Herrera-Viedma and F. Herrera, "An approach for detecting, quantifying, and visualizing the evolution of a research field: A practical application to the Fuzzy Sets Theory field", Journal of Informetrics, vol. 5, num. 1, pp. 146-166, 2011.
- M.J. Cobo, A.G. López-Herrera, E. Herrera-Viedma and F. Herrera, "SciMAT: A new Science Mapping Analysis Software Tool", Journal of the American Society for Information Science and Technology, vol. 63, num. 8, pp. 1609-1630, 2012.
- N. Sclater, A. Peasgood and J. Mullan, Learning Analytics in Higher Education. A review of UK and international practice. United Kingdom: Jisc, 2016.
- N. Selwyn, "Data entry: towards the critical study of digital data and education," Learning, Media and Technology, vol. 40, num. 1, pp. 64–82, 2015.
- O. Simpson, "Predicting student success in open distance learning," Open Learning, vol. 21, num. 2, pp. 125-138, 2006.
- R. S. Baker, "Stupid tutoring systems, intelligent humans," International Journal of Artificial Intelligence in Education, vol. 26, Issue 2, pp. 600-614, 2016.
- V. Aleven, J. Sewall, O. Popescu, F. Xhakaj, D. Chand, R. Baker, Y. Wang, G. Siemens, C. Rosé and D. Gasevic, "The Beginning of a Beautiful Friendship? Intelligent Tutoring Systems and MOOCs" in Artificial Intelligence in Education, AIED 2015, C. Conati, N. Heffernan, A. Mitrovic, M. Verdejo, Eds. Lecture Notes in Computer Science, vol 9112. Springer, 2015, pp. pp 525-528.
- W. Holmes, M. Bialik, and C. Fadel, Artificial intelligence in education: promises and implications for teaching and learning. Boston, MA: The Center for Curriculum Redesign, 2019.

- SciMAT webpage: https://sci2s.ugr.es/scimat/index.html

Thanks very much!