Analyzing Knowledge Construction

Ulrike Cress
IWM Tuebingen, Germany
Knowledge Construction Lab
OUTLINE

• My theoretical background

• Our co-evolution model about individual learning and collaborative knowledge construction (peer production)

• Typical lab studies

• Studies with Wikipedia log data

• What network theory can provide for the analysis of knowledge construction
MY BACKGROUND

I am a cognitive psychologist

- knowledge as network
- spreading activation
- basis of meaning making
MY BACKGROUND

I am a cognitive psychologist

- knowledge as network
- spreading activation
- basis of meaning making
MY BACKGROUND

I am a cognitive psychologist

- knowledge as network
- spreading activation
- basis of meaning making
- cognitive conflict: learning through irritation
- learning: new concepts, different link strengths
Knowledge as network
- mental concepts
- meaning of a perception is constructed
- meaning of an entity is defined by its relation to other concepts
COMMUNICATION
COMMUNICATION

lalalal

nununu

gagaga

kukuku

sasasas

lululu
COMMUNICATION
COMMUNICATION

Communication as Network

- people
- utterances
- topics

- meaning of an utterance is constructed
- meaning is defined by its relation to other utterances
- „external knowledge“; artefacts
THEORY DEVELOPMENT

CO-EVOLUTION MODEL

Theoretical papers
- Knowledge Management Research & Practice (2010)
- AI & Society (2011)
- Educational Psychologist (2015)

Empirical papers
- Computers and Education (2012; 2013)
- Interacting with Computers (2011)
- Information, Communication and Society (2010)
- Journal of Medial Internet Research (2014)
- Medical Education (2013)
- Computers in Human Behavior (2012)
- PLOS ONE (2014)
- Journal of Computer-Supported Collaborative Learning (2014)
AUTOPOIETIC SYSTEMS (Maturana; Luhmann)

Systems
- do not exist per se
- exist through their operations
- operationally closed
- but open for their environments

- like a species in the evolutionary process: adapting to it’s niche
KNOWLEDGE CONSTRUCTION - PEER PRODUCTION

social system

cognitive system
COGNITIVE SYSTEM

- autopoietic, exists through own operations (cognitions)
- operationally closed
- can be stimulated through its environment
- irritation
- deals with irritation

→ LEARNING
  more facts
  deeper understanding, deeper knowledge
  change in attitude
SOCIAL SYSTEM

- autopoietic system
- operates through communication
- operationally closed
- develops rules, norms, culture
  - Wikipedia
  - Metapedia
- can be stimulated through its environment
- develops through operating on irritations

→ KNOWLEDGE CONSTRUCTION
STRUCTURAL COUPLING

- both systems operate
- each system is environment for the other
- each one can stimulate/irritate the other
- incongruity leads to co-evolution
- the systems develop (drift)
  - learning
  - knowledge construction
EMPIRICALLY ANALYZING THESE DYNAMIC PROCESSES

• controversial domains
• analyzing individual learning (internal)
  – individual expertise / interests / behaviour / attitude
• analyzing knowledge construction (utterances; artefact)
  – communication thread (quality, quantity)
  – uptakes
  – topical/conceptual development
• analyzing coupling / co-evolution
TYPICAL (PSYCHOLOGICAL) STUDIES

- Studies with Wikis/Wikipedia
- Studies with social tagging systems, Internet forums, design patterns, knowledge platforms
STUDIES WITH WIKIs

controversal topic

- pro argument
- con argument
STUDIES WITH WIKIs

controversial topic

- pro argument
- con argument
STUDIES WITH WIKIs

controversal topic

pro argument

con argument
STUDIES WITH WIKIs

OUTCOME MEASURES

KNOWLEDGE CONSTRUCTION
produced wiki-text
• arguments
• integration of arguments
• one-sidedness

LEARNING
• knowledge
• attitudes
FINDINGS

KNOWLEDGE CONSTRUCTION

Wiki

LEARNING

level of incongruity

low medium high

level of incongruity
FIRST APPROACH WITH FIELD (BIG) DATA

Goal: Analyze the **conceptual** development knowledge

Study in Wikipedia: CAUSES OF SCHIZOPHRENIA

Data: direct neighbours of the Wikipedia-page “causes of schizophrenia”
- development of the *linkage* of pages (knowledge construction)
- development of *individuals* (learning)
- years 2003-2008
DEVELOPMENT OF CORPUS

square size = nr. of in-links
CORPUS 2003

black squares: grouped pages (k-cycles length 3)
grey squares: boundary spanner
small triangle: not grouped

Psychoanalyt. Cluster

Biolog-Cluster
AUTHORS’ EDITS

Individual edits on Wikipedia-pages

Expert classification of pages
- Psychoanalytic: Freud; ÜBER-ICH
- Biological: HIPPOCAMPUS
- Social: DOUBLE COMMUNICATION
- integrated socio-biological (diathese-stress model) (VULNERABILITY)

Classification of authors according their contributions
- Psychoanalytic
- Biological
- Social
- Socio-biological
DEVELOPMENT OF AUTHORS: LEARNING
RELEVANCE OF BOUNDARY-SPANNER

Social cluster:

Biolog. Cluster
RELEVANCE OF BOUNDARY-SPANNER
RELEVANCE OF BOUNDARY-SPANNER

Hypothesis: knowledge progress happens in the intersection of different domains/communities

INNOVATION HAPPENS AT INTERSECTIONS.
- VALDIS KREBS, LEADING EXPERT IN NETWORK ANALYSIS
RELEVANCE OF BOUNDARY-SPANNER

The role of boundary-spanning articles for knowledge creation.

Who creates boundary-spanning (interdisciplinary) articles?

Study with two different but overlapping topics: pedagogy, psychology
DOMAIN CLASSIFICATION

Category: Psychology

From Wikipedia, the free encyclopedia

Category: Pedagogy

From Wikipedia, the free encyclopedia

Psychology is a collective of industrial disciplines concerned with the understanding of behavior and prediction of behavioral pathologies.

Pedagogy is the art or science of teaching.

Subcategories

This category has the following 13 subcategories, out of 13 total.

- Applied learning (2 C, 33 P)
- Critical pedagogy (51 P)
- Curricula (5 C, 67 P)
- Didactics (1 C, 12 P)
- Educational psychology (18 C, 269 P)
- Educationists (5 C, 42 P)
- Personality (6 C, 59 P)
- Philosophy of psychology (9 C, 20 P)
- Pedagogic integrated development environments (24 P)
- Internships (3 C, 29 P)
- Language-teaching methodology (3 C, 33 P)
- Learning programs (21 C, 29 P)
- Montessori education (2 C, 11 P)
- Online education (1 C, 17 P)
- Women and
ARTICLE CORPUS

Network of 11,000 articles
undirected hyperlinks as edges
RELEVANCE OF ARTICLE

Pedagogy

eigenvector

Psychology

eigenvector
RELEVANCE OF ARTICLE

Combined network

betweenness centrality
CATEGORIZATION OF AUTHORS

identification of all edits from an author (> 150 characters)

Categorization as
• specialist (either Psych. or Ped.)
• generalist (Psych. and Ped.)

• Wikipedia expertise (no. of edits)

Do these features predict if an author’s contributions are central?
What predicts the centrality of an author’s contribution in the Psychology corpus

\[ Y_i = \alpha + \beta X_i + \delta Z_i + \varepsilon_i \]

- \( Y_i = \) log eigenvector of author i
- \( X_i = 1 \) if author i is a generalist
  i.e. 0 if author i is a specialist in Psychology
- \( Z_i = \) log number of articles of author i
- \( \varepsilon_i = \) Residuum

<table>
<thead>
<tr>
<th>Domain</th>
<th>Centrality Measures</th>
<th>( \beta ) generalist/specialist</th>
<th>( \delta ) general expertise author</th>
</tr>
</thead>
<tbody>
<tr>
<td>psychology</td>
<td>Eigenvector</td>
<td>-0.36***</td>
<td>0.30***</td>
</tr>
</tbody>
</table>
REGRESSION MODEL

What predicts the centrality of an author’s contribution? in the Pedagogy Corpus?

\[ Y_i = \alpha + \beta X_i + \delta Z_i + \varepsilon_i \]

- \( Y_i \) = log eigenvector of author \( i \)
- \( X_i \) = 1 if author \( i \) ein generalist
  = 0 if author \( i \) ein specialist in Pedagogy
- \( Z_i \) = log number of articles of author \( i \)
- \( \varepsilon_i \) = Residuum

<table>
<thead>
<tr>
<th>Domain</th>
<th>Centrality Measures</th>
<th>( \beta ) generalist/specialist</th>
<th>( \delta ) general expertise author</th>
</tr>
</thead>
<tbody>
<tr>
<td>psychology</td>
<td>Eigenvector</td>
<td>-0.36***</td>
<td>0.30***</td>
</tr>
<tr>
<td>pedagogy</td>
<td>Eigenvector</td>
<td>-0.15</td>
<td>0.38***</td>
</tr>
</tbody>
</table>
REGRESSION MODEL

What predicts the centrality of an author’s contribution for the combined network?

\[ Y_i = \alpha + \beta X_i + \delta Z_i + \varepsilon_i \]

- \( Y_i \) = log Betweenness of author \( i \)
- \( X_i = 1 \), if author \( i \) ein generalist
  = 0 if author \( i \) ein specialist
- \( Z_i = \log \) number of articles of author \( i \)
- \( \varepsilon_i = \) Residuum

<table>
<thead>
<tr>
<th>Domain</th>
<th>Centrality Measures</th>
<th>( \beta ) generalist/specialist</th>
<th>( \delta ) general expertise author</th>
</tr>
</thead>
<tbody>
<tr>
<td>psychology</td>
<td>Eigenvector</td>
<td>-0.36***</td>
<td>0.30***</td>
</tr>
<tr>
<td>pedagogy</td>
<td>Eigenvector</td>
<td>-0.15</td>
<td>0.38***</td>
</tr>
<tr>
<td>combined network</td>
<td>Betweenness</td>
<td>0.54***</td>
<td>0.60***</td>
</tr>
</tbody>
</table>
TEMPORAL DEVELOPMENT OF NEW KNOWLEDGE

<table>
<thead>
<tr>
<th>period</th>
<th>Psychology</th>
<th>Pedagogy</th>
<th>Intersection</th>
<th>whole network</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>articles</td>
<td>links</td>
<td>articles</td>
<td>links</td>
</tr>
<tr>
<td>2006</td>
<td>2176</td>
<td>12834</td>
<td>1357</td>
<td>5007</td>
</tr>
<tr>
<td>2007</td>
<td>2911</td>
<td>20311</td>
<td>1980</td>
<td>7861</td>
</tr>
<tr>
<td>2008</td>
<td>3472</td>
<td>26615</td>
<td>2556</td>
<td>10925</td>
</tr>
<tr>
<td>2009</td>
<td>3908</td>
<td>31955</td>
<td>3108</td>
<td>13564</td>
</tr>
<tr>
<td>2010</td>
<td>4262</td>
<td>36028</td>
<td>3595</td>
<td>16074</td>
</tr>
<tr>
<td>2011</td>
<td>4660</td>
<td>40983</td>
<td>4166</td>
<td>18793</td>
</tr>
<tr>
<td>2012</td>
<td>5085</td>
<td>44939</td>
<td>4696</td>
<td>22518</td>
</tr>
</tbody>
</table>
### MULTILEVEL LOGISTIC MODEL (periods nested in articles)

**new edits**

<table>
<thead>
<tr>
<th></th>
<th>level</th>
<th>est. value</th>
<th>z</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>combined network</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Intercept)</td>
<td>level</td>
<td>0.63</td>
<td>7.11</td>
<td>1.2e-12***</td>
</tr>
<tr>
<td>creation year</td>
<td>Article</td>
<td>-0.39</td>
<td>-34.40</td>
<td>&lt;2e-16***</td>
</tr>
<tr>
<td>article age</td>
<td>Periode</td>
<td>-0.31</td>
<td>-32.86</td>
<td>&lt;2e-16***</td>
</tr>
<tr>
<td>t-1 log betweenness</td>
<td>Periode</td>
<td>0.09</td>
<td>12.10</td>
<td>&lt;2e-16***</td>
</tr>
<tr>
<td>t-1 log no. of edits</td>
<td>Periode</td>
<td>0.65</td>
<td>33.02</td>
<td>&lt;2e-16***</td>
</tr>
<tr>
<td>excellent articles</td>
<td>Article</td>
<td>0.18</td>
<td>0.94</td>
<td>0.35</td>
</tr>
<tr>
<td>log controversiality</td>
<td>Article</td>
<td>0.20</td>
<td>16.34</td>
<td>&lt;2e-16***</td>
</tr>
</tbody>
</table>
MULTILEVEL LOGISTIC MODEL
(periods nested in articles)
newly created articles as neighbors

<table>
<thead>
<tr>
<th>Model</th>
<th>Level</th>
<th>Est. Value</th>
<th>Z</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>combined network</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Intercept)</td>
<td></td>
<td>2.56</td>
<td>23.19</td>
<td>1.2e-12***</td>
</tr>
<tr>
<td>creation year</td>
<td>Article</td>
<td>-0.33</td>
<td>-25.27</td>
<td>&lt;2e-16***</td>
</tr>
<tr>
<td>article age</td>
<td>Periode</td>
<td>-0.22</td>
<td>-21.57</td>
<td>&lt;2e-16***</td>
</tr>
<tr>
<td>t-1 log betweenness</td>
<td>Periode</td>
<td>0.31</td>
<td>12.10</td>
<td>&lt;2e-16***</td>
</tr>
<tr>
<td>t-1 log no. of edits</td>
<td>Periode</td>
<td>0.26</td>
<td>11.40</td>
<td>&lt;2e-16***</td>
</tr>
<tr>
<td>excellent articles</td>
<td>Article</td>
<td>0.04</td>
<td>0.20</td>
<td>0.83</td>
</tr>
<tr>
<td>log controversiality</td>
<td>Article</td>
<td>0.05</td>
<td>4.26</td>
<td>&lt;2e-05***</td>
</tr>
</tbody>
</table>
MULTILEVEL LOGISTIC MODEL (periods nested in articles)
edit count of the neighboring articles

<table>
<thead>
<tr>
<th></th>
<th>level</th>
<th>est. value</th>
<th>z</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>combined network</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Intercept)</td>
<td></td>
<td>131.95</td>
<td>32.29</td>
<td>&lt;2e-12***</td>
</tr>
<tr>
<td>creation year</td>
<td>Article</td>
<td>-7.73</td>
<td>-15.97</td>
<td>&lt;2e-16***</td>
</tr>
<tr>
<td>article age</td>
<td>Periode</td>
<td>-8.13</td>
<td>-23.32</td>
<td>&lt;2e-16***</td>
</tr>
<tr>
<td>Δ neighbors since t-1</td>
<td>Periode</td>
<td>20.42</td>
<td>180.30</td>
<td>&lt;2e-16***</td>
</tr>
<tr>
<td>t-1 log betweenness</td>
<td>Periode</td>
<td>5.86</td>
<td>17.82</td>
<td>&lt;2e-16***</td>
</tr>
<tr>
<td>t-1 log no. of edits</td>
<td>Period</td>
<td>9.98</td>
<td>11.82</td>
<td>&lt;2e-16***</td>
</tr>
<tr>
<td>excellent articles</td>
<td>Article</td>
<td>53.85</td>
<td>6.30</td>
<td>0.83</td>
</tr>
<tr>
<td>log controversiality</td>
<td>Article</td>
<td>6.28</td>
<td>12.68</td>
<td>&lt;2e-05***</td>
</tr>
</tbody>
</table>
MULTILEVEL LOGISTIC MODEL (periods nested in articles)

similar results for

- pedagogy network
- psychology network
- combined network
RESULTS

Preferential attachment
• central articles attract more edits, neighbours, edits in neighbours

• Boundary-spanning articles attract more edits, neighbours, edits in neighbours
   → Role of conflict, discourse
CONTROVERSIES IN WIKIPEDIA

“I love controversy”
STUDIES WITH WIKIs

controversial topic
pro argument
con argument
STUDIES WITH WIKIs

Knowledge Construction

Wiki

Learning

LEARNING

low medium high
USE OF SEMANTICS

Wikipedia: All articles about alternative medicine (400 articles)
- classification of the 500,000 modifications (via machine learning)

classification Modification 1
neutral / alternative / conventional

classification Modification 2
neutral / alternative / conventional
Dynamic development of an *article: aggregation of all edits*
ARTICLE PROFILES

Extremity

Number_of_edits
AUTHOR PROFILES

Dynamic development of an author: aggregation of all edits done by that author

alternative

conventional

number of edits
AUTHOR PROFILES

Extremity

Number_of_edits
INCONGRUITY AUTHOR/ARTICLE
INCONGRUITY: DELTAS AUTHOR-ARTICLE
PREDICTION OF EDITS

significant predictors
– article profile ***
– author profile ***
– interaction ***

conventional ............................................................alternative

<table>
<thead>
<tr>
<th></th>
<th>0 - 0.2</th>
<th>0.2 - 0.4</th>
<th>0.4 - 0.6</th>
<th>0.6 - 0.8</th>
<th>0.8 - 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>author</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0 - 0.2</td>
<td>*</td>
<td>=</td>
<td>*</td>
<td>=</td>
<td>=</td>
</tr>
<tr>
<td>0.2 - 0.4</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>=</td>
<td>=</td>
</tr>
<tr>
<td>0.4 - 0.6</td>
<td>=</td>
<td>*</td>
<td>*</td>
<td>=</td>
<td>=</td>
</tr>
<tr>
<td>0.6 - 0.8</td>
<td>=</td>
<td>=</td>
<td>=</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>0.8 - 1</td>
<td>=</td>
<td>=</td>
<td>=</td>
<td>*</td>
<td>*</td>
</tr>
</tbody>
</table>

\[ \chi^2 = 339.2, \text{ df } = 16, p < 0.0001 \]
INTERPRETATION

• Controversy does not always lead to knowledge production
• authors contribute to article they agree on
• articles become less extreme over time
• authors become less extreme over time

CONCLUSION

What can network theory and methods provide for the analysis of knowledge construction?

Fits to the autopoietic framework
• preferential attachment
• structure shapes dynamics
• no external criteria for relevance, quality etc.
• features of an element are determined through the whole network

Deals with the bimodal person-artefact network
• allows to analyse dynamics of individuals
• allows to analyse dynamics of the artefact
CONCLUSION

What can network theory and methods provide for the analysis of knowledge construction?

Fits to the concept of knowledge and collaboration
- knowledge as network
- communication as network
- learning, knowledge construction as development of conceptual / communicational networks
REFERENCES


COLLABORATORS

• Joachim Kimmerle
• Aileen Oeberst
• Iassen Halatschlyiski
• Johannes Moskaliuk
• Julia Hancke
• Jens Jirschitzka
Thank you for your attention