CHIST-ERA Projects Seminar 2021
Visual Analytics for Decision Making under Uncertainty (VADMU)

Alejandro Benito-Santos (PROVIDEDH, VisUSAL)

April 14, 2021

❖ One central challenge of big data analysis is to understand their reliability, scope, and accuracy, and to communicate these aspects to users in an intuitive manner.

❖ New visual analysis tools and approaches are needed to support trusted and efficient decision making under uncertainty in a variety of application contexts.

❖ These are expected to require for example uncertainty analysis of ensemble data, sensitivity analysis of input-output models, and supported decision making that will allow expert users to understand the reliability and conflicts inherent in the analysis.

Visual Data Exploration
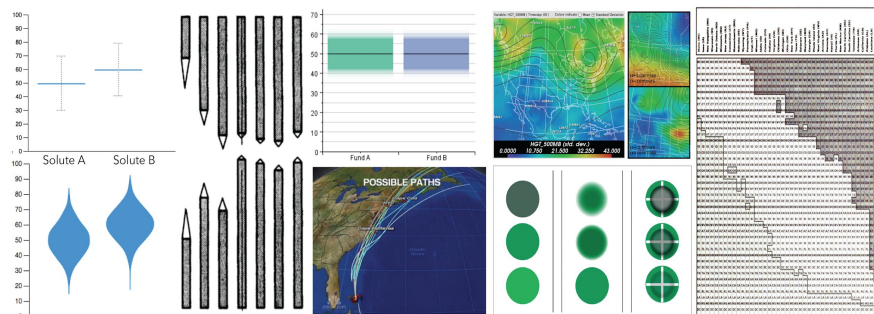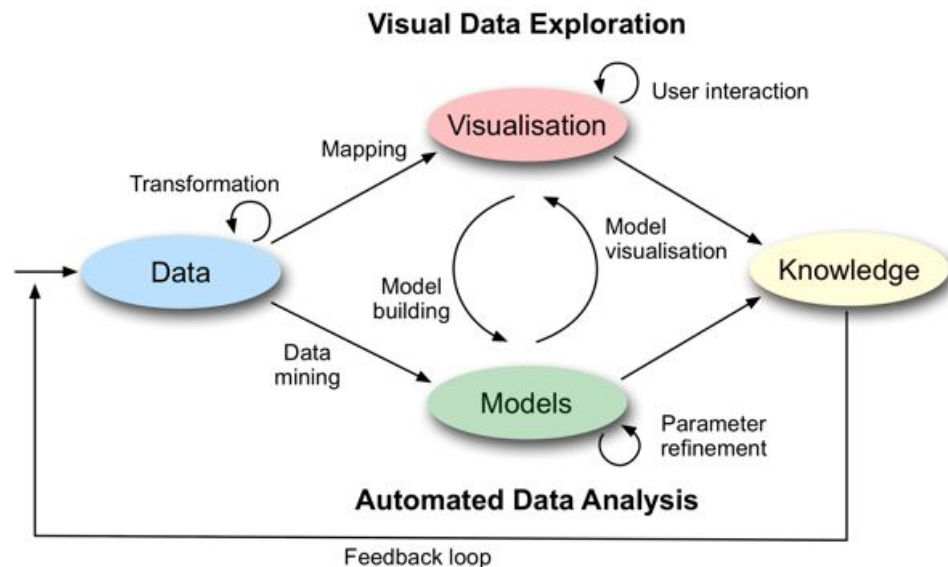
Visual Analytics for Decision Making under Uncertainty

❖ **PROVIDEDH: PROgressive VIsual DEcision-making for Digital Humanities**
  - ✓ Assess and track the degree of uncertainty of Digital Humanities research objects and how they are affected when different computational models are applied to them?
  - ✓ Is it possible to convey this evolution of uncertainty with interactive multimodal interfaces that progressively adapt to the moment of decision-making?

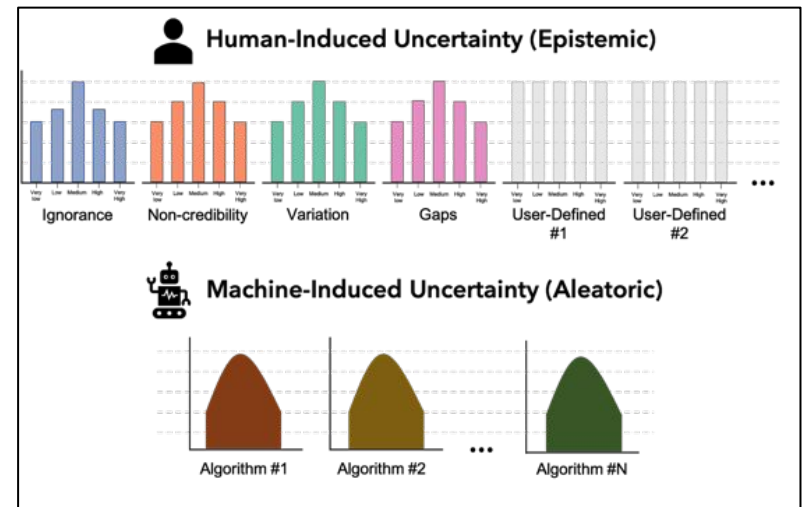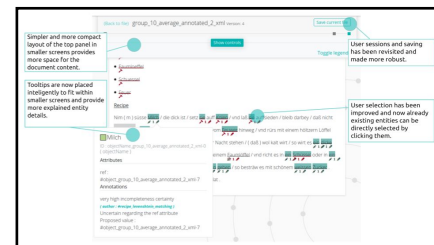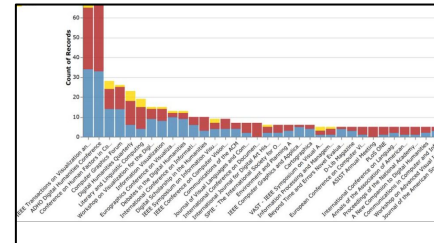❖ **IVAN: Interactive and Visual Analysis of Networks**
  - ✓ Create a visual analysis system for the exploration of dynamic or time-dependent networks, in particular using modern clustering methods.

- ❖ **Algorithms need human guidance; humans need algorithmic support** for complex problems.
  - ✔ It is of **vital importance** to find ways in which **users and algorithms can collaborate.**
  - ✔ Explorations of **ensembles and parameter spaces** of algorithms were proposed.
- ❖ **Participatory design workshops** are key to identify questions and capture the users' mental models
- ❖ Different proposals to convey uncertainty in different application settings and measure how this affects collective and/or individual sensemaking.

❖ A **literature review** on **visualization** for the **digital humanities**
- ✔ To further **characterize** and **define** the **design space** of **digital humanities**.
- ✔ To **identify commonalities** and **differences** between information and uncertainty visualization applied to experimental sciences and humanities research.

❖ An **uncertainty taxonomy** developed through different **evaluation techniques** carried out alongside experts.
- ✔ Interviews with experts
- ✔ Participatory design workshops
- ✔ Engagement with the VIS/DH communities
  - ▪ **Special Issues**
    - • Uncertainty in Digital Humanities (MDPI Informatics)
    - • Uncertainty Visualization and Decision Making (Frontiers in Computer Science/Psychology)
  - ▪ **Active participation in VIS/DH conferences (DARIAH, IEEE VIS, DH Conference,...)**
    - • Organized workshops
    - • Publications

❖ A **framework and specification** to annotate uncertainty in XML-TEI documents.

❖ A **collaborative platform** that enables **collective reasoning and sensemaking** on a corpus with a focus on **uncertainty management** and **representation**.

✔ **Close reading module**: for annotating the corpus and fixating perspectives/interpretations on the research objects.

✔ **Distant reading** of the corpus with a focus on collective uncertainty communication and interpretation.

❖ Application of **visualization techniques** aimed at depicting uncertainty in distinct DH tasks (e.g., visualizing uncertainty in tagged texts or the evolution of bounded uncertainty in a collection over time).

There is no way the machine or human can work alone

- ❖ Participatory design workshops are key to identify questions and capture the users' mental models
- ❖ Two full-day Creative Visualization-Opportunities (CVO) Workshops*
- ❖ Requirement analysis: Results used as a foundation for the development of visual analysis systems

Paris Workshop, 2019

Geneva Workshop, 2020

- ● Eight DH Researchers
- ● Historic trading networks

- ● Eight Medical Researchers
- ● EEG data of Epilepsy patients

Our achievements are in three principle areas (with output examples):

1. **Novel algorithms for network clustering**
   - Novel algorithm based on graph Slepians that allows to zoom on nodes of interest
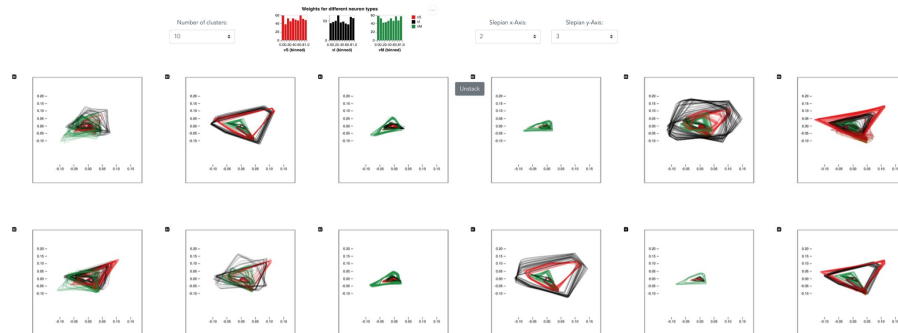
2. **Comparing multiple network clusterings**
   - Integrating Prior Knowledge into clustering process ("PK-Clustering")
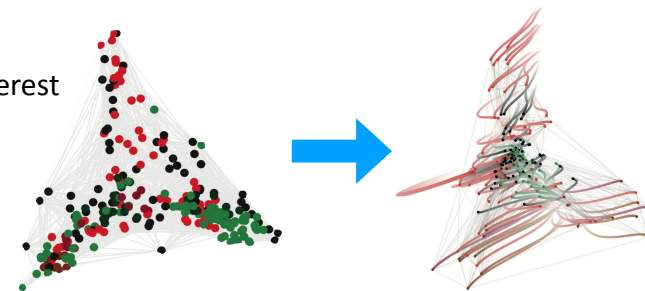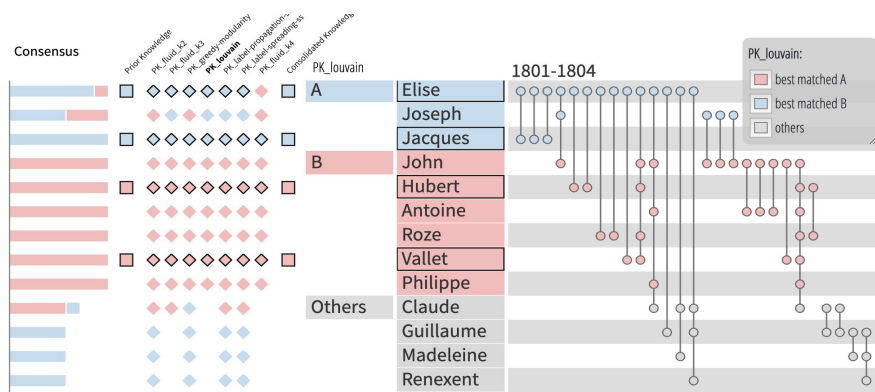
3. **Novel techniques suitable for exploring networks and their clusterings in the presence of uncertainties**
   - Novel tool for Visual Parameter Space Exploration

❖ Foster the transfer of exploration methodologies when data is uncertain (faulty, incomplete, etc.) between different application areas of visualization research (e.g., humanities and brain science)

❖ Refine evaluation strategies for visual decision-making tools by taking into account the degree of trust and confidence users have on the underlying algorithms

❖ Shorten the development cycle for visual decision-making tools, e.g. by developing "meta"tools for such applications and through participatory design

Possible Roadmap

❖ **More user experiments with our tools (slowed-down by COVID-19)**

❖ **Our tools are continuing to be expanded and improved**
  ✔ Additional funding to continue development of our tools

❑ **Most helpful aspects**

    ❑ Focus of the thematic call

    ❑ Multidisciplinary Collaboration between the partners

    ❑ Transnational dimension

❑ **Inherent challenges**

    ❑ Can be tricky to align four national projects under one umbrella

        ❑ **Would be good to keep additional national overhead to a minimum (such as additional reporting, proposal phase, etc.)**

    ❑ ANR personnel turnover was high

❖ **Gender balance**

✔ **PROVIDEDH** Gender-balanced but...

▪ **CS: all men, Humanities: all women**

✔ **IVAN** CS: 5 men / 1 woman

❖ **Uncertainty representation helps mitigate biases**

✔ By increasing algorithmic literacy and critical thinking towards algorithms

❖ **Bridging humanities and experimental sciences**

✔ Interdisciplinary work is challenging: researchers should take this into account and plan accordingly when making proposals.

❖ **PROVIDEDH**
  ✔ Code and datasets:
    ▪ **https://github.com/providedh**
    ▪ **https://visusal.github.io/pilaster**
  ✔ OA Publications: https://providedh.eu/dissemination/

❖ **IVAN**
  ✔ https://aviz.fr/paohvis
  ✔ https://aviz.fr/pkclustering
  ✔ https://c4science.ch/source/guidedGSE/
  ✔ OA Publications: https://ivan.ai/#publications

❖ Several of our tools are available on the web, open source, and will remain usable by researchers and inspirational for industry

❖ **PROVIDEDH**
  ✔ Test possible applications of the platforms in a **learning environment** at high school/university level.
  ✔ Focused on poetry analysis.
  ✔ Interesting application scenario because:
    ▪ **It incorporates state-of-the-art linguistic algorithms which may produce multiple interpretations of the same corpus.**
    ▪ **Datasets and algorithms complex enough to drive the development of novel progressive visual analytics techniques.**
  ✔ In collaboration with experts outside the consortium: POSTDATA project (ERC Starting Grant 2015 – Horizon 2020)

❖ **IVAN**
  ✔ Included network analysis in university courses
  ✔ Outreach and collaboration with partners outside the grant (e.g. Austrian Center for Digital Humanities)

Questions ?

PROVIDEDH:
Poetry Analysis and Teaching

- Conduct a **new**, **full evaluation of the system** on the use-case of **poetry analysis (scansion)** in collaboration with DH (non-VIS) scholars and professors.
- **Good application scenario** because:
  - Incorporates **state-of-the-art linguistic models** that are relevant in a **specific DH context and set of concrete tasks**.
  - Allows testing **how collaboration between experts and the algorithms** is enabled in the system in a new, untested domain.
    - to produce **multiple interpretations** of a poetry corpus.
    - to **refine the underlying analysis** algorithm.
  - **Datasets** and **algorithms** are **complex enough** to test **progressive visual analytics approaches** in conjunction with **uncertainty visualization techniques**.
- Test **possible applications** for **teaching poetry analysis** at high school/university level.
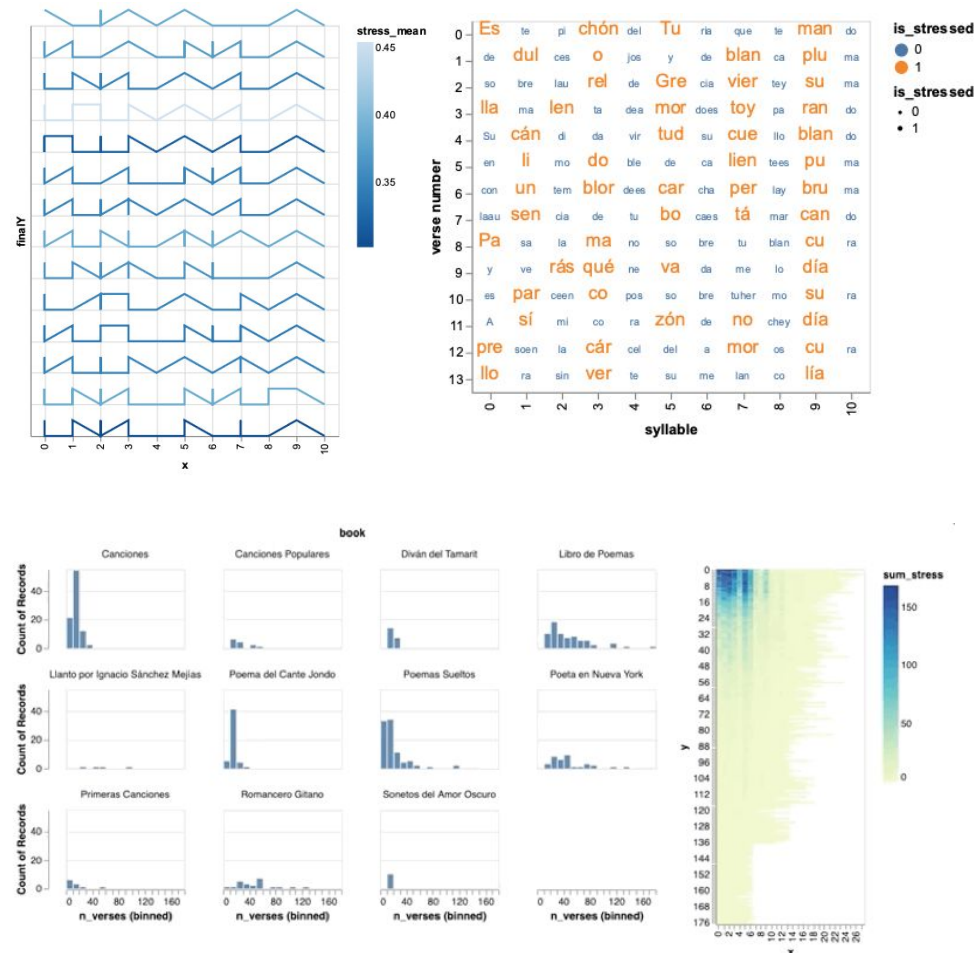
When data has uncertainty and decisions should be made, ICT is not enough. Therefore the VADMU projects focus on the following:

❑ **Exploration, interaction, visualization, to understand the results of analyses and nature of uncertainty**

❑ **Low latency for algorithms to fit the human capabilities**

  ❑ Fast algorithms, or approximate results, or progressive results

❑ **New lines or methods for data analysis at scale under uncertainty**

PROVIDEDH: Project Focus

❖ **Two key scientific questions**
  ✓

❖ **Key challenges**
  ✓ Categorise all sources of uncertainty that can affect DH
  ✓ Develop a set of metrics for the degree of uncertainty propose a (software) framework that makes use it
  ✓ Clarify what DH tasks need decision-making and build a pipeline oriented towards the handling of uncertainty

- University of Salamanca (SPAIN)
  — Visual Analytics Leader
  Project Coordinator: Roberto Therón

- Austrian Academy of Sciences (AUSTRIA)
  — Open Innovation in Digital Humanities Leader
  PI: Eveline Wandl-Vogt

- Trinity College Dublin, Faculty of Arts Humanities and Social Sciences (IRELAND)
  — Engagement and Exploitation Leader
  PI: Jennifer Edmond

- Poznan Supercomputing and Networking Center (POLAND)
  — Infrastructures and Platform Leader
  PI: Cezary Mazurek

PROVIDEH GANTT		MONTH
		1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36
WP1	Defining and managing Uncertainty in DH	
WP2	Open innovation in the Humanities	
WP3	User interfaces and Evaluation	
WP4	Collaborative platform development	
WP5	Dissemination and Exploitation of Results	
WP6	Project Management	
Milestones		X X X X X X X X X X X X X X X
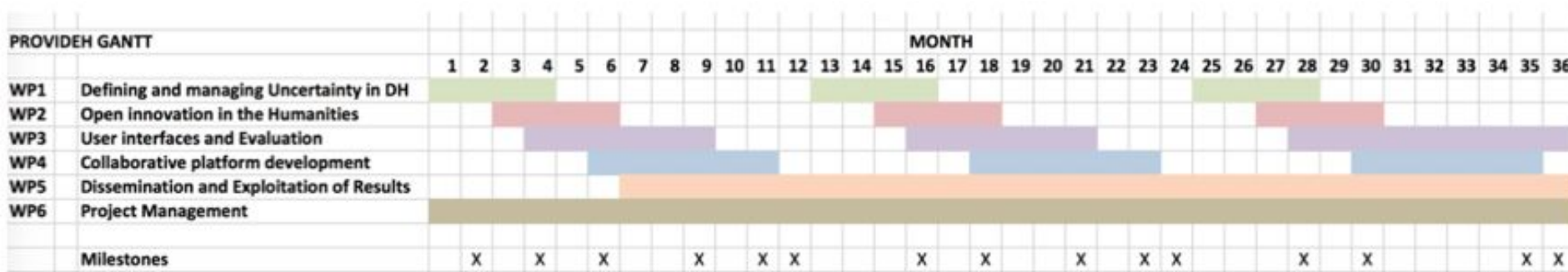
PROVIDEDH:
Expert-Driven Uncertainty Taxonomy

- Enabled by several **evaluations** that took place during the course of the project.
  - **Ethnographic studies** organized as workshops with experts
  - **Quantitative and qualitative pilot studies**
    - Interview sessions
    - Questionnaires
- Different tasks:
  - Name/Entity normalization
    - Cooking recipes
    - 1641 Depositions
  - Literary analysis
    - Beckett's texts

Epistemic Uncertainty
(Human-generated, Discrete PDF)

Aleatoric Uncertainty
(Machine-generated, continuous PDF)

Ignorance
(lack of context)

Non-credibility (error or bias seen to be entering system)

Variation
(differing values within a contested category)

Gaps
(outlier record where information is missing)

Algorithmic uncertainty
(The results of the computation are probabilistic)

$0 <= f(x) <= 1$

{very low, low, medium, high, very high, unknown}

Human-Induced Uncertainty (Epistemic)

Ignorance   Non-credibility   Variation   Gaps   User-Defined #1   User-Defined #2

Machine-Induced Uncertainty (Aleatoric)

Algorithm #1   Algorithm #2   ...   Algorithm #N
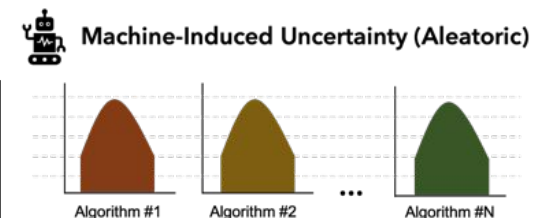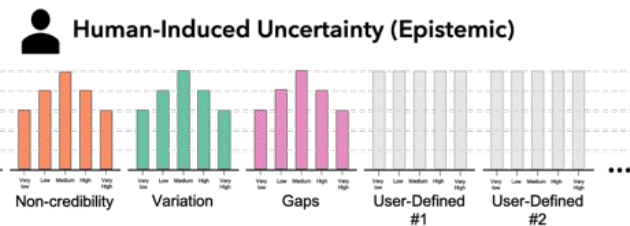
- Consider robots as "first-class citizens"
- E.g.: In topic models, probability of a word to belong to a certain topic

21

Certainty taxonomy

Certainty tags in TEI allow to annotate missing or incorrect information, specify your confidence on a modification, and collaborate with other people's work through nested annotations. Choose what different sources of uncertainty you will use to describe your annotations and their color and icon scheme.

Category name

ignorance

"Write a description for ignoran

credibility

"Write a description for credibil

imprecision

"Write a description for impreci

incompleteness

"Write a description for incomp
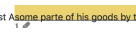
Add
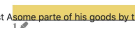
"Write a description for the nev

Example text with current scheme

in some ignorance sed,urna mauris, dapibus some credibility.dapibusmetus. Integer urna some imprecision metus. Integer urna mauris,mauris, dapibus in lacus sed, some incompleteness sed, dictum porttitor metus.

damnified the sume of fiftie fortie fiue pounds. Of prouision in the house to the value of three pounds. The totall Of turffe & fuell fortie shillinges Of garden fruites to the value of fortie shillings The totall of his losses amounts to foure score pounds two shillings. The deponent saith that he lost A some parte of his goods by the meanes of JohnCondon fitz dauid of Bally mc Patricke in the said County gentleman
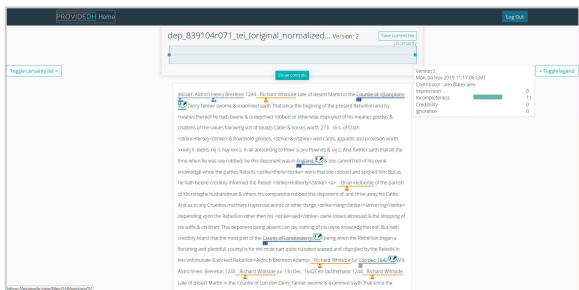
Variation   Non-credibility

Ignorance   Variation

22
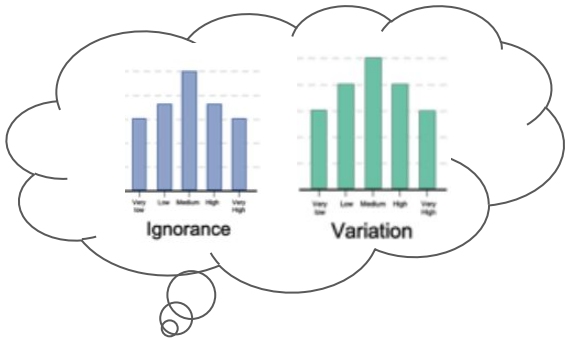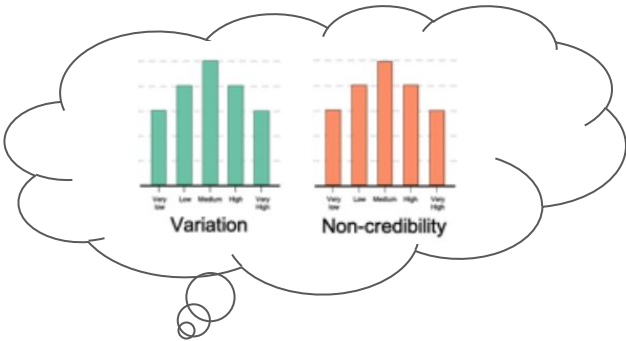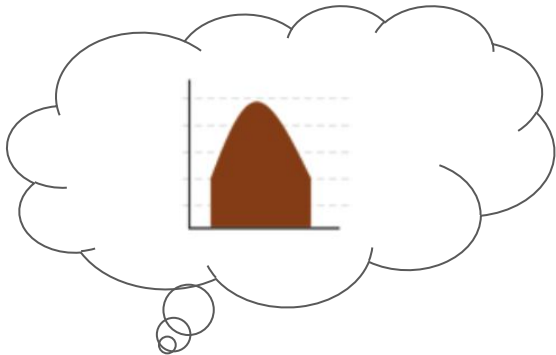
PROVIDEDH: Publication & Dissemination

- 19 Journal Articles
- 9 publications in Conference Proceedings
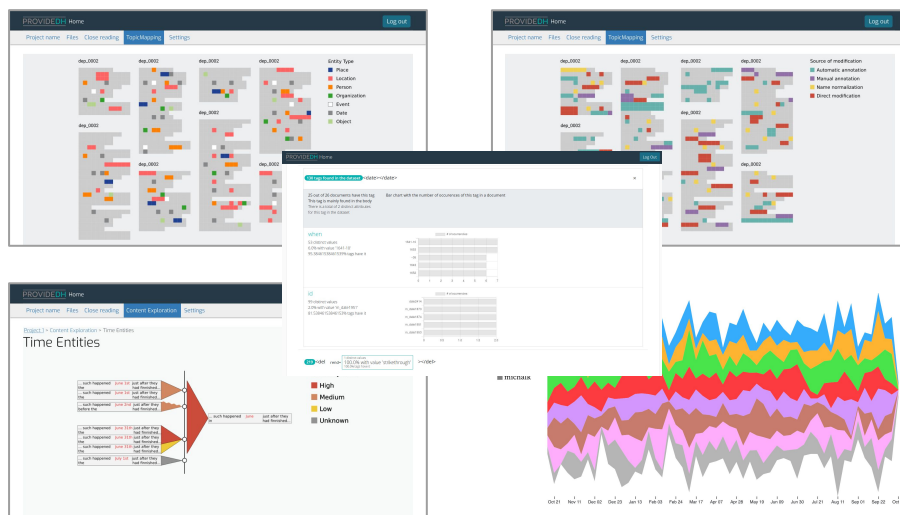- 3 conference workshops

Collaborative Platform (https://providedh.ehum.psnc.pl)

Close Reading and Annotation

Distant Reading

IVAN: Interactive Visual Analysis of Networks

Torsten Möller,
torsten.moeller@univie.ac.at
Christian Knoll
christian.knoll@univie.ac.at
Expertise in exploring the issues of modeling under uncertainty and decision making

Jean-Daniel Fekete, Paola Valdivia
Expertise in HCI, network visualization, and evaluation methods for visualization

Dimitri Van De Ville, Raphaël Liégeois
Graph signal processing expertise including spectral/wavelet approaches
with application to brain networks
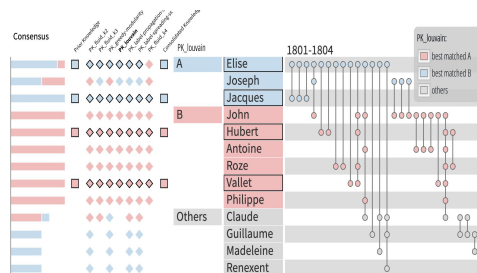
Project start: April 2018

IVAN: Interactive Visual Analysis of Networks

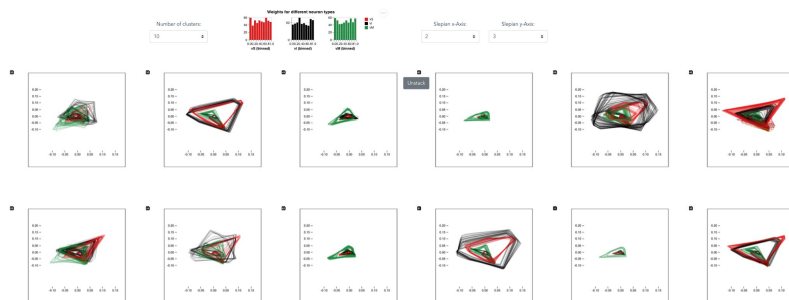Our contributions are in three principle areas:

1. **Novel algorithms for network clustering**
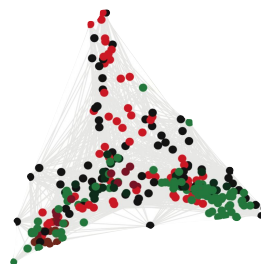
2. **Comparing multiple network clusterings**

3. **Novel techniques suitable for exploring networks and their clusterings in the presence of uncertainties**

IVAN: Interactive Visual Analysis of Networks

❑ Requirement analysis: needs and user tasks

❑ Two full-day Creative Visualization-Opportunities (CVO) Workshops*

❑ Results used as a foundation for the development of visual analysis systems

Paris Workshop, 2019

Geneva Workshop, 2020

- Eight DH Researchers
- Historic trading networks

- Eight Medical Researchers
- EEG data of Epilepsy patients

* *A Framework for Creative Visualization-Opportunities Workshops*
E. Kerzner, S. Goodwin, J. Dykes, S. Jones and M. Meyer
in IEEE Transactions on Visualization and Computer Graphics, vol. 25, no. 1, pp. 748-758, Jan. 2019,
doi: 10.1109/TVCG.2018.2865241.

□ **Zooming into network to reveal "deep" organization**

□ New algorithm based on concept of graph Slepians that generalizes Laplacian embedding

□ Visual tool for parameter space exploration in development

connectome of *C. Elegans*

sensory   inter   motor

❑ **Integrating Prior Knowledge in Mixed Initiative Social Network Clustering**

■ Integrates PK from the expert into the clustering process and keeps him/her in the analysis loop

■ Ensemble of clustering algorithms (including Slepian algorithm) is performed on the data

System:

Workflow:

2
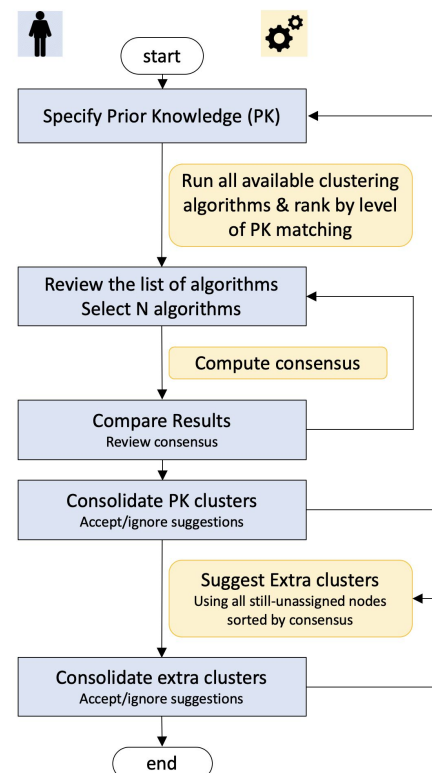
IVAN: Interactive Visual Analysis of Networks

❏ **Visual Parameter Space Exploration of input space for Slepian algorithm**
- Clustering of the output of the algorithm
- First prototype currently in development
- Allow researchers from other domains to use algorithm on their data
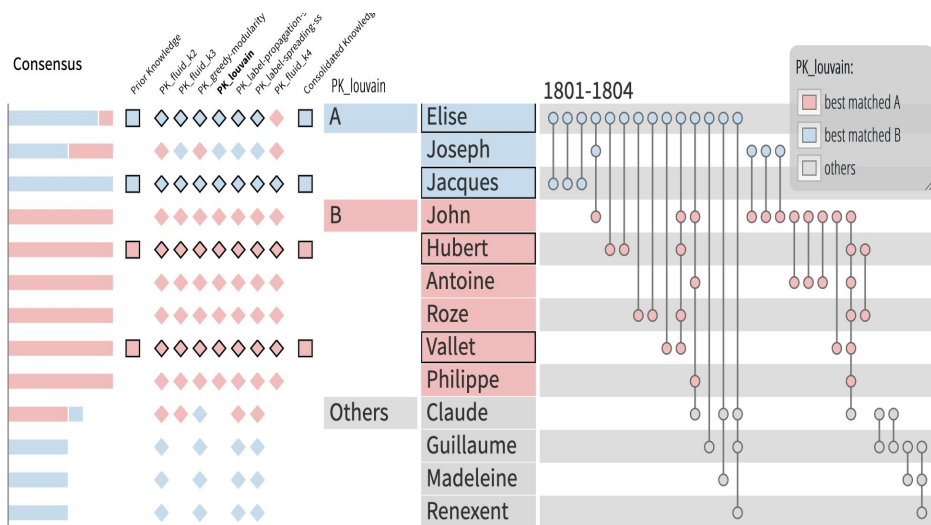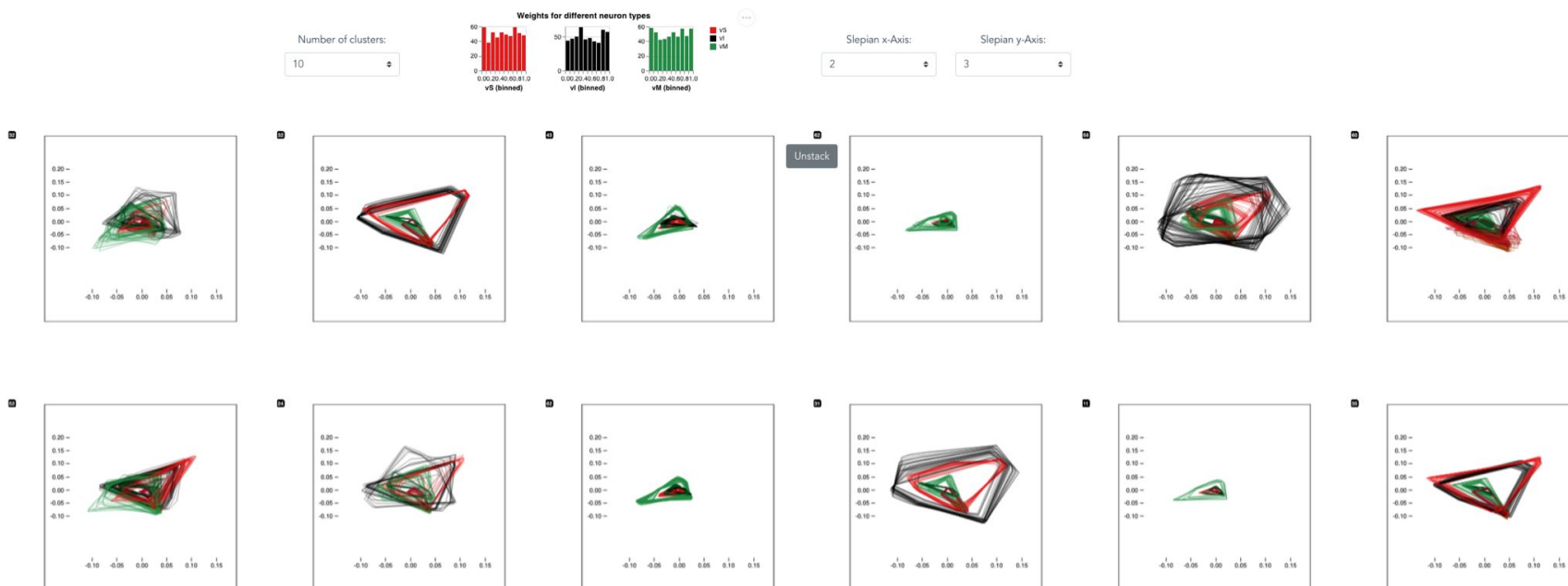
❑ **Taking into account fluctuations of networks (dynamics)**
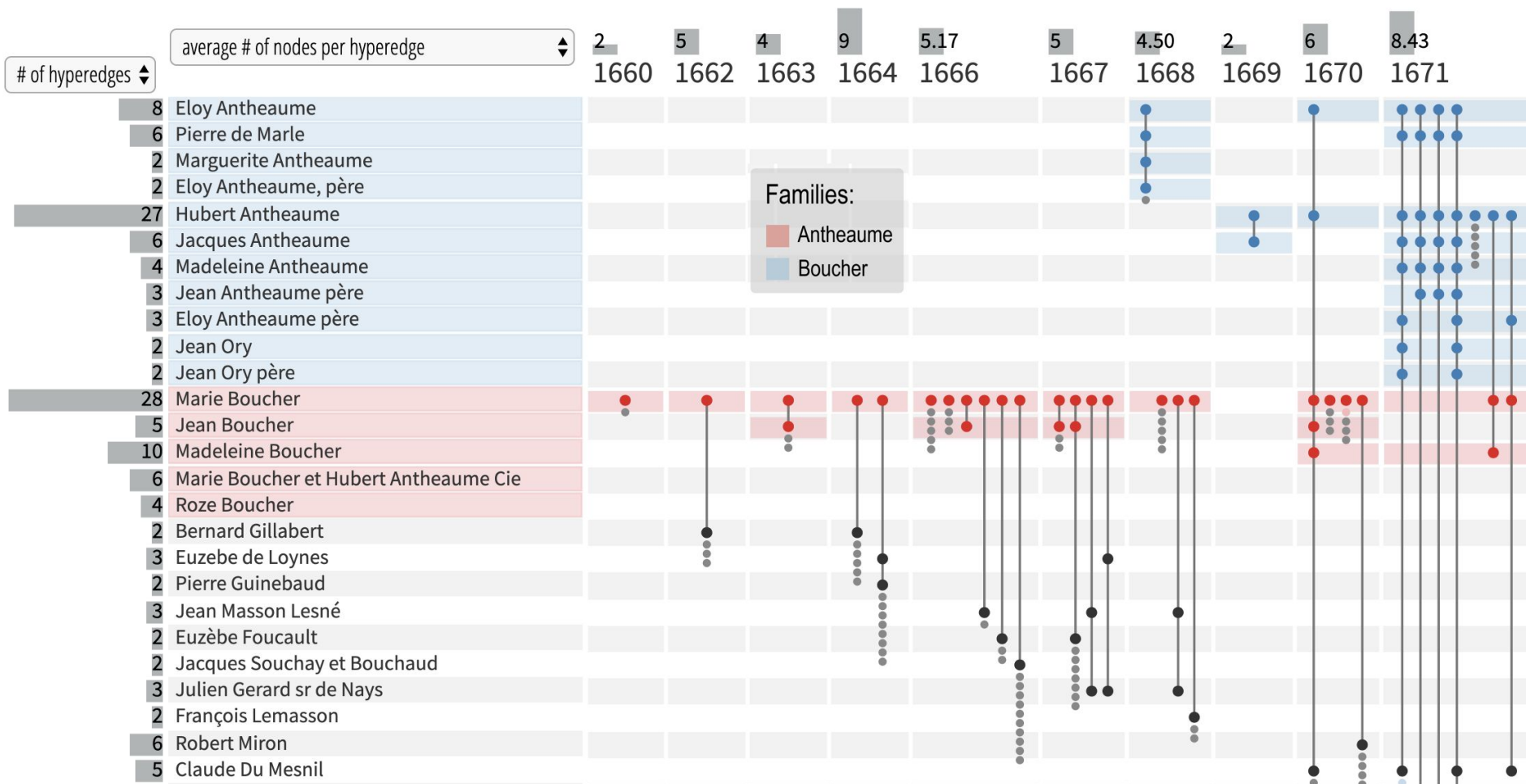
3

IVAN: Interactive Visual Analysis of Networks

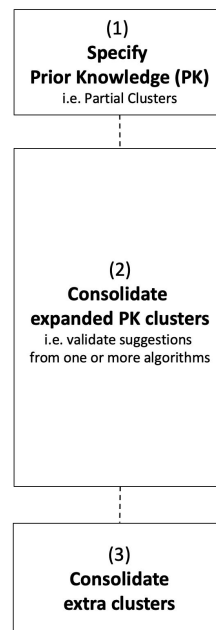❑ **Integrating Prior Knowledge in Mixed Initiative Social Network Clustering**

3
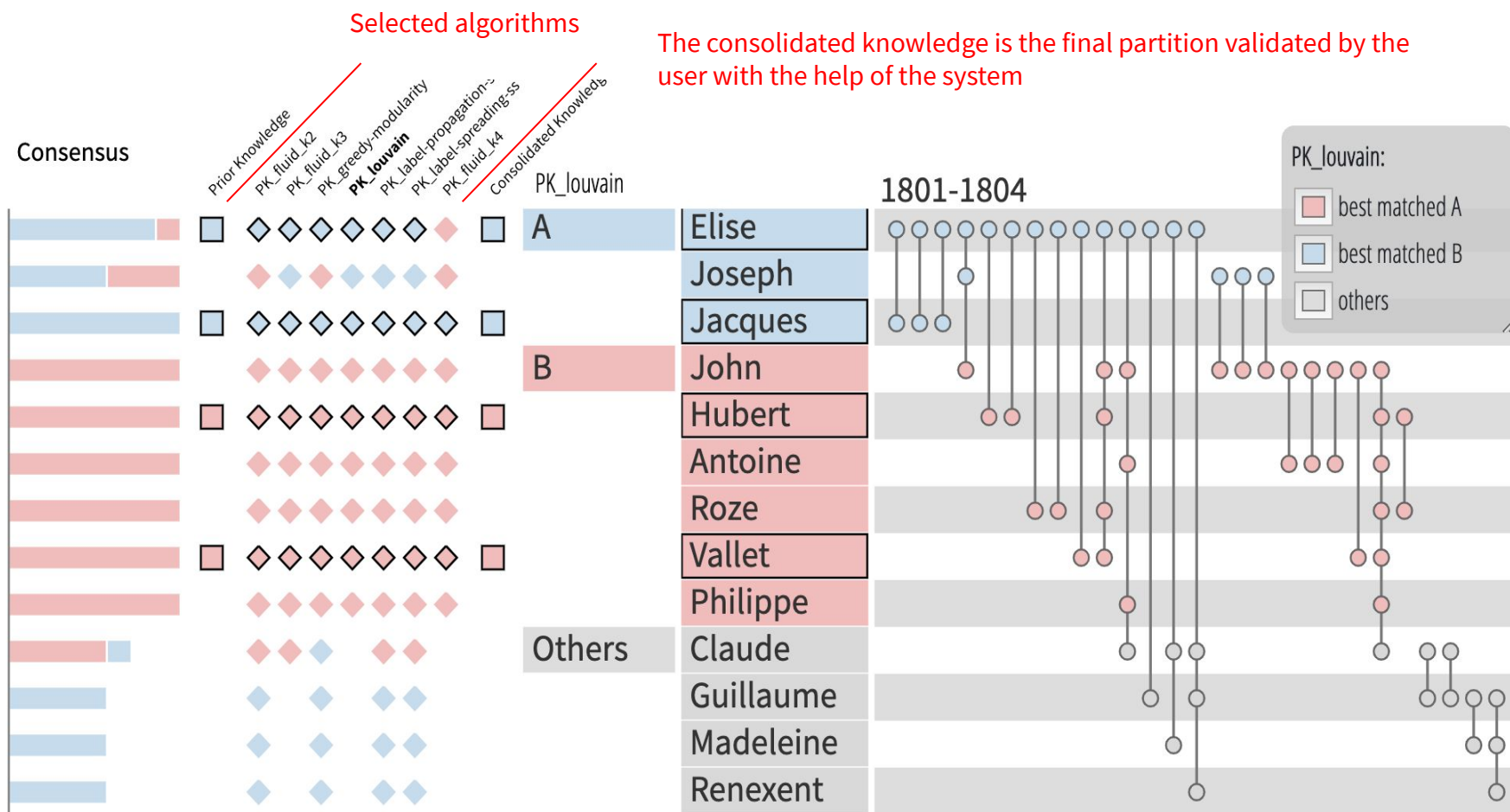
IVAN: Significant Results and Dissemination to Date

❑ **Dissemination**

    ❑ Prototype code on git

        ▪ **https://aviz.fr/paohvis**

        ▪ **https://aviz.fr/pkclustering**

    ❑ Interactions and collaborations with user groups from neuroscience and social sciences

❑ **Academic output**

    ❑ 10 articles (7 journal, 3 conference)

Integrating Prior Knowledge in Mixed Initiative Social Network Clustering. Alexis Pister, Paolo Buono, Jean-Daniel Fekete, Catherine Plaisant, Paola Valdivia, IEEE Transactions on Visualization and Computer Graphics, IEEE, ⟨10.1109/TVCG.2020.3030347⟩

Interpreting Null Models of Resting-State Functional MRI Raphaël Liégeois, B. T. Thomas Yeo, Dimitri Van De Ville bioRxiv 2021.03.30.437514; https://doi.org/10.1101/2021.03.30.437514

Extending Recommendations for Creative Visualization-Opportunities Workshops. C. Knoll, A. Çetin, T. Möller and M. Meyer IN: 2020 IEEE Workshop on Evaluation and Beyond - Methodological Approaches to Visualization (BELIV), Salt Lake City, UT, USA, 2020, pp. 81-88 https://doi.org/10.1109/BELIV51497.2020.00017

Guided graph spectral embedding: Application to the C. elegans connectome. Miljan Petrovic, Thomas A. W. Bolton, Maria Giulia Preti, Raphaël Liégeois and Dimitri Van De Ville IN: Network Neuroscience, Volume 3 | Issue 3 | 2019, p.807-826 https://doi.org/10.1162/netn_a_00084

Dynamic mode decomposition of resting-state and task fMRI. J. Casorso, X. Kong, W. Chi, D. Van De Ville, T. Yeo, and R. Liégeois. IN: Neuroimage 194, pp. 42-54, 2019 https://doi.org/10.1016/j.neuroimage.2019.03.019

Resting brain dynamics at different timescales capture distinct aspects of human behavior. R. Liégeois, J. Li, R. Kong, D. Van De Ville, T. Ge, M. Sabuncu and T. Yeo. IN: Nature Communications 10 (1), 2317, 2019. https://doi.org/10.1038/s41467-019-10317-7

Analyzing Dynamic Hypergraphs with Parallel Aggregated Ordered Hypergraph Visualization. P. Valdivia, P. Buono, C. Plaisant, N. Dufournaud and J.-D. Fekete. IN: IEEE Transactions on Visualization and Computer Graphics. In Press. https://doi.org/10.1109/TVCG.2019.2933196

Time-resolved analysis of dynamic graphs: an extended Slepian design. R. Liégeois, I. Merad, and D. Van De Ville.

Wavelets and Sparsity XVIII 1113810, 2019.

Using Dynamic Hypergraphs to Reveal the Evolution of the Business Network of a 17th Century French Woman Merchant.
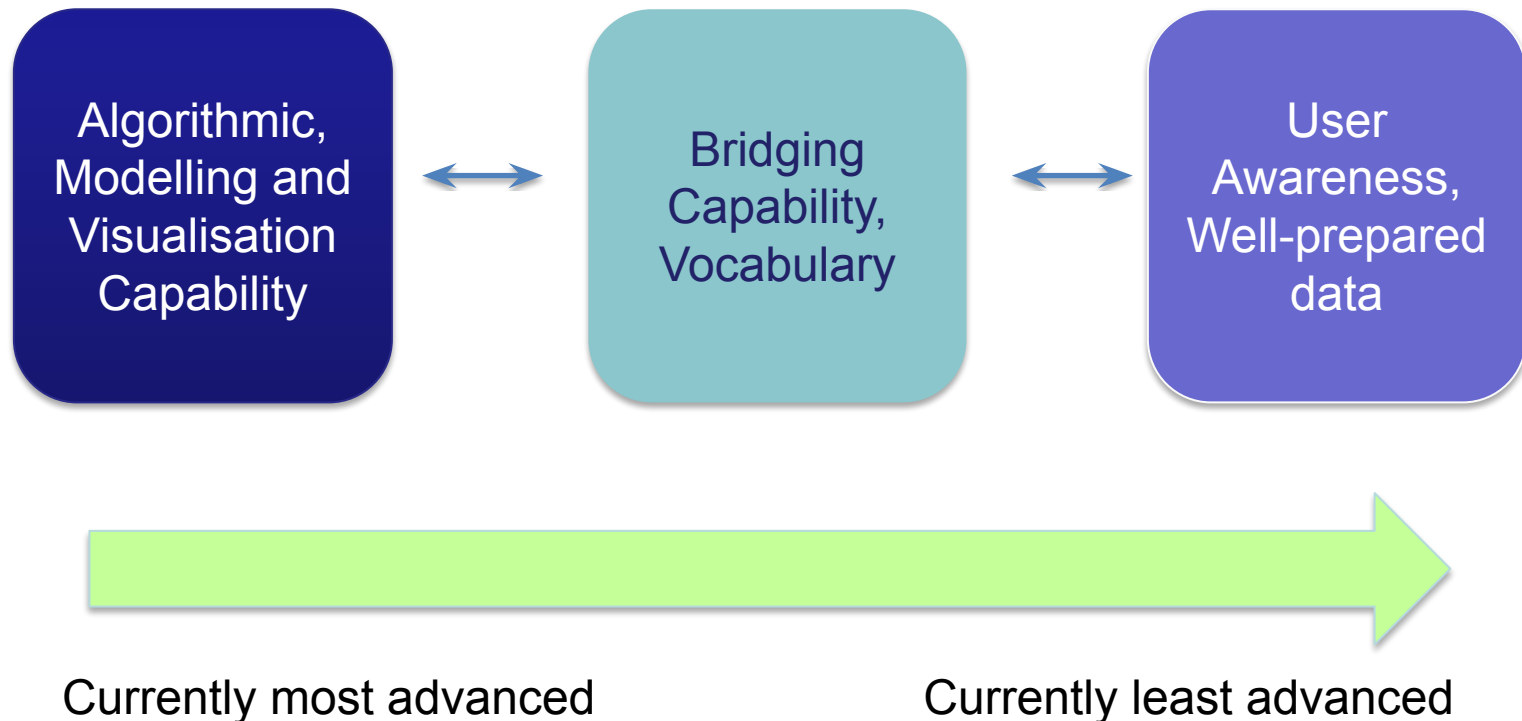
Paola Valdivia, Paolo Buono, Catherine Plaisant, Nicole Dufournaud, Jean-Daniel Fekete. VIS 2018 - 3rd Workshop on Visualization for the Digital Humanities, Oct 2018, Berlin, Germany.

Topic Challenges and Needs

- **Scalability, Efficiency**
- **Interpretability**
  - Lack of 'ground truth'
- **Evaluation**
  - Limited value of existing benchmarks
- **Role and needs of the human agents**
  - Challenges of new user communities
- **Using the language of the domain**

Possible Roadmap

Three areas of interlinked challenges and opportunities that need to develop in concert

| Algorithmic, Modelling and Visualisation Capability | ← → | Bridging Capability, Vocabulary | ← → | User Awareness, Well-prepared data |

Currently most advanced   Currently least advanced

Role of the CHIST-ERA Support

❑ **Most helpful aspects**

   ❑ Focus of the thematic call

   ❑ Multidisciplinary Collaboration between the partners

   ❑ Transnational dimension

❑ **Inherent challenges**

   ❑ Can be tricky to align four national projects under one umbrella

      ❑ **Would be good to keep additional national overhead to a minimum (such as additional reporting, proposal phase, etc.)**

Questions ?