V2-25: High-Productivity Analytics

Mission-Critical Computing
NSF CENTER FOR SPACE, HIGH-PERFORMANCE, AND RESILIENT COMPUTING (SHREC)

SHREC Annual Workshop (SAW24-25)

University of Pittsburgh
BYU BRIGHAM YOUNG UNIVERSITY
VIRGINIA TECH
UF UNIVERSITY of FLORIDA

January 14-15, 2025

Faculty: Chris North, Doug Bowman
(Professors of Computer Science at Virginia Tech)

Students: Sungwon In, Eric Rippey, Ibrahim Tahmid, Xuxin Tang, Frank Wanye

Number of requested memberships ≥ 5

V2-25 Tasks

- High-Performance Analytics =
  Immersive Analytics + Human-AI Interaction + Parallel Computation

- **Task 1: Immersive Data Science (Sungwon In) [presented by Chris]**
- **Task 2: Immersive Semantic Interaction (Ibrahim Tahmid)**
- **Task 3: Interactive LLM for High-Performance Sensemaking (Xuxin Tang)**
- **Task 4: Parallel Computational Analytics (Frank Wanye)**

Task 1: Immersive Data Science

Mission-Critical Computing
NSF CENTER FOR SPACE, HIGH-PERFORMANCE, AND RESILIENT COMPUTING (SHREC)

SHREC Annual Workshop (SAW24-25)

University of Pittsburgh
BYU BRIGHAM YOUNG UNIVERSITY
VIRGINIA TECH
UF UNIVERSITY of FLORIDA

Sungwon In

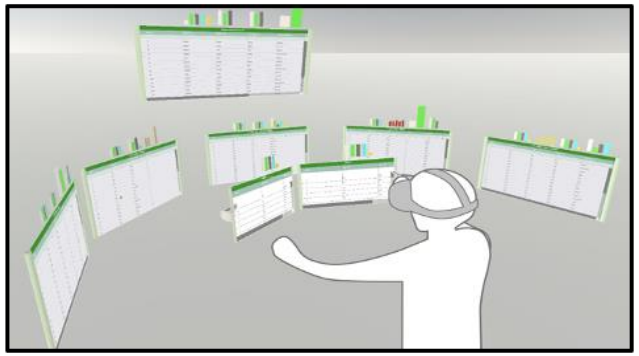
January 14-15, 2025

ICoN: Immersive Computational Notebook for Data Science

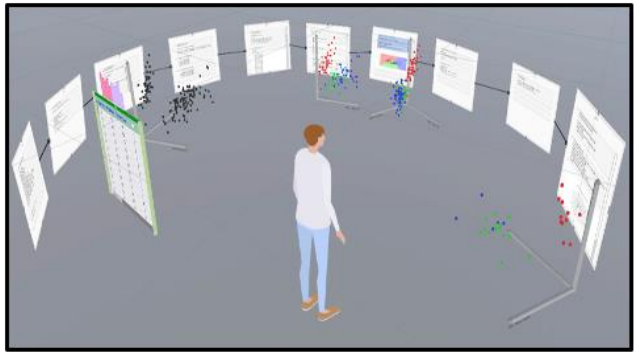
General Limitations

- Scalability
- Optimizing Interaction
- Target Users
- Broader Applicability

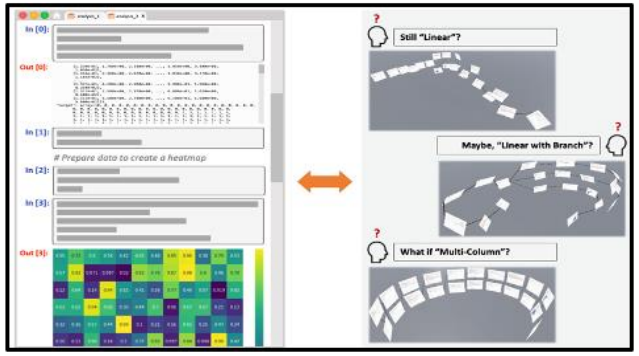
Initial Exploration of ICoN



Immersive Data Transformation



Evaluation of ICoN



Organizations in ICoN

ICoN: Immersive Computational Notebook for Data Science

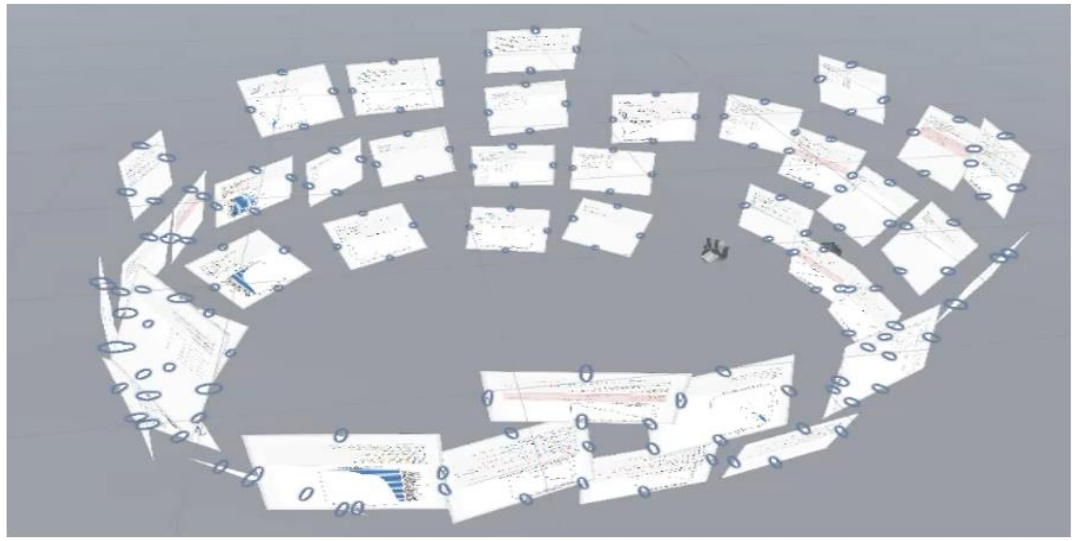
- **What Specifically?**

Scalability

- Why does it matter?
  - If we are working on huge projects, More:
    1. Codes (cells/notebooks)
    2. Data Artifacts
    3. Line Indicators

50 notebooks

100 notebooks



100 notebooks From inside

ICoN: Immersive Computational Notebook for Data Science

How can we improve manageability for large-scale data science tasks?

Task 2: Immersive Semantic Interaction

Mission-Critical Computing
NSF CENTER FOR SPACE, HIGH-PERFORMANCE, AND RESILIENT COMPUTING (SHREC)

SHREC Annual Workshop (SAW24-25)

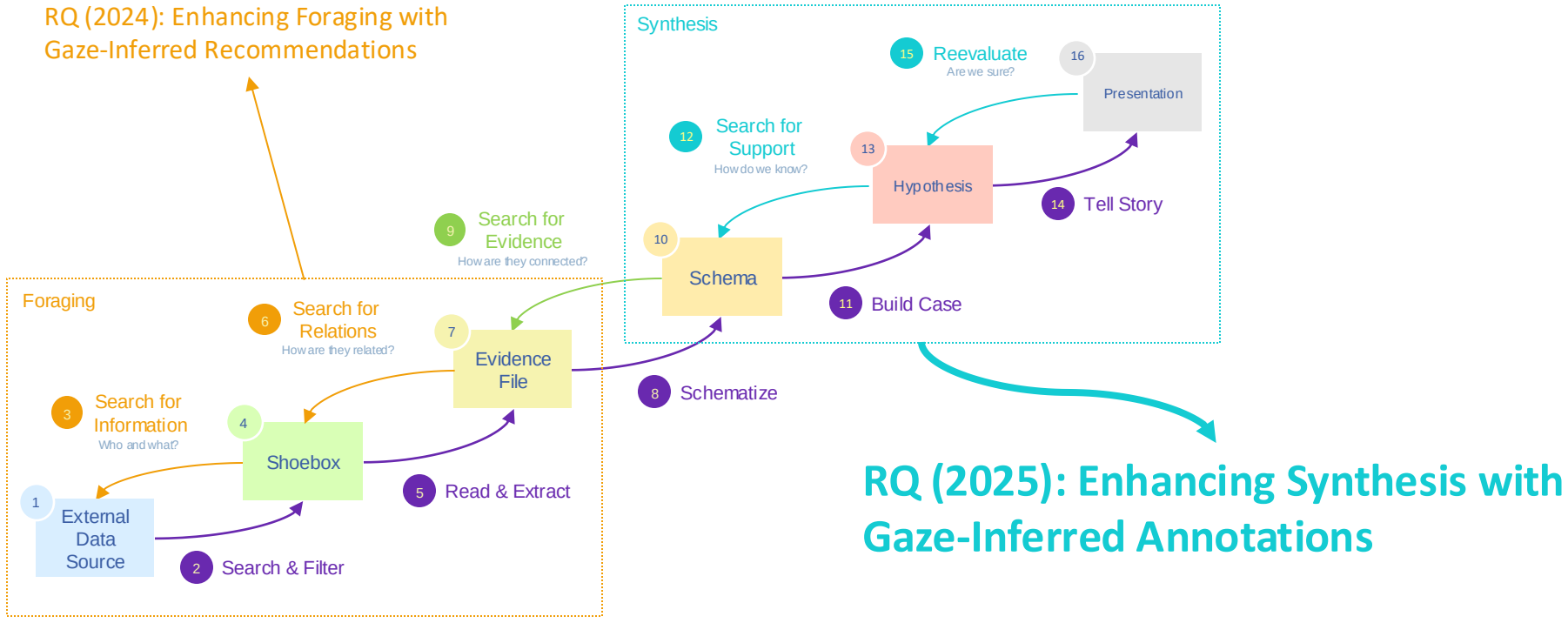
University of Pittsburgh
BYU BRIGHAM YOUNG UNIVERSITY
VIRGINIA TECH
UF UNIVERSITY of FLORIDA

Ibrahim Tahmid

January 14-15, 2025

V2

Goal: Enhancing Synthesis with Gaze-Derived Annotations



V2

Study Plan

Study 1: Gaze Data Collection

- Professional analysts complete sensemaking while their gaze is tracked
- Help us isolate the foraging steps and ensure gaze data is a decent representation of the ground truth

Data Preparation
Use gaze data to infer information relevance
Annotate documents based on their relevance

Study 2: Annotation Evaluation

- Condition 1: Novice analysts synthesize information from the annotated documents
- Condition 2: Novice analysts synthesize information without any annotation

Expected Outcome

- Understand the benefits and challenges of gaze-derived annotations in synthesis

Task 3: Interactive LLM for High-Performance Sensemaking

Mission-Critical Computing
NSF CENTER FOR SPACE, HIGH-PERFORMANCE, AND RESILIENT COMPUTING (SHREC)

SHREC Annual Workshop (SAW24-25)

University of Pittsburgh
BYU BRIGHAM YOUNG UNIVERSITY
VIRGINIA TECH
UF UNIVERSITY of FLORIDA

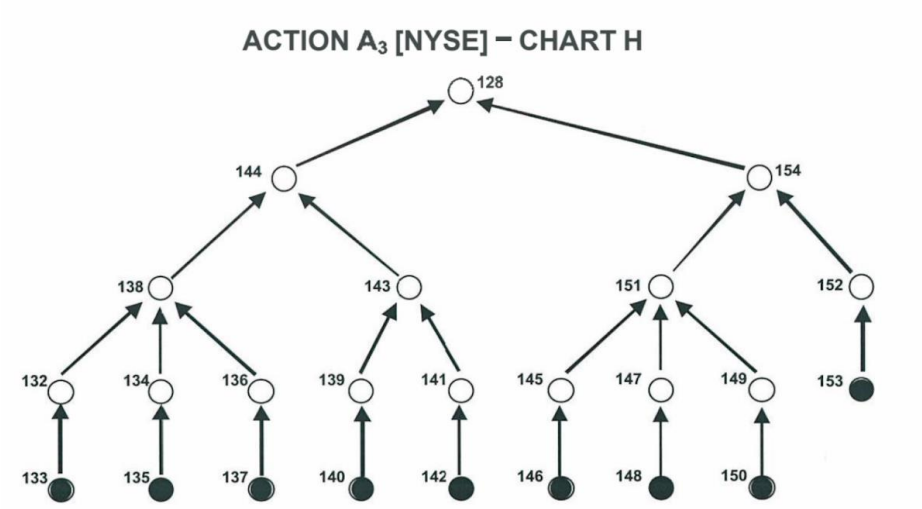
Xuxin Tang

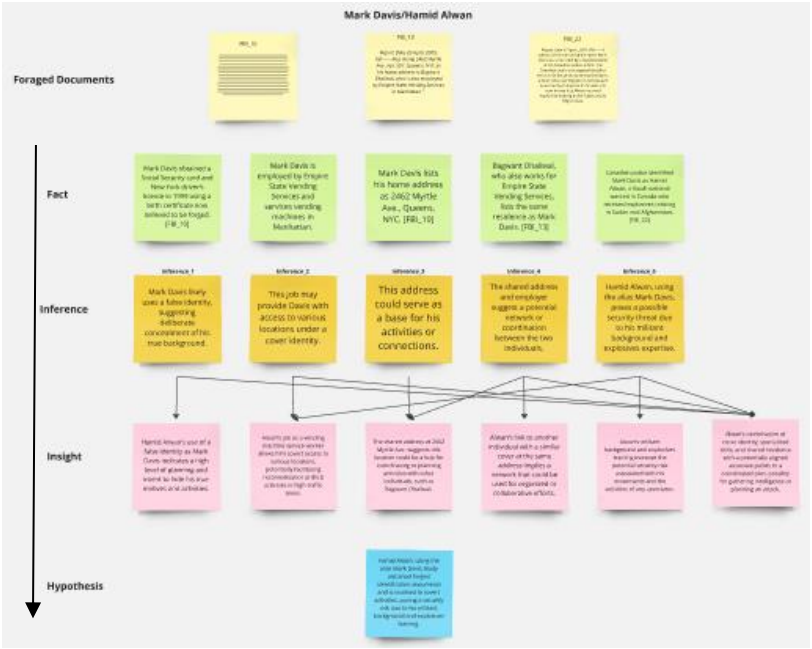
January 14-15, 2025

V2

In-Context Human-LLM Visual Conversation in Workspace

How to save users from intensive document reading for higher-performance sensemaking?





Plan

V2

- Preliminary Study:
  - Is it possible to use LLMs and workspace to solve a sensemaking task without document reading?
  - What will users' interactions and solving processes be?

- Interface Design:
  - How can we design effective visualizations for graphical conversations by incorporating external data sources?

Task 4: Parallel Computational Analytics

Mission-Critical Computing
NSF CENTER FOR SPACE, HIGH-PERFORMANCE, AND RESILIENT COMPUTING (SHREC)

SHREC Annual Workshop (SAW24-25)

University of Pittsburgh
BYU BRIGHAM YOUNG UNIVERSITY
VIRGINIA TECH
UF UNIVERSITY of FLORIDA

Eric Rippey, Frank Wanye

January 14-15, 2025

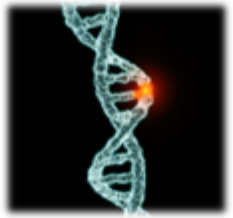
V2

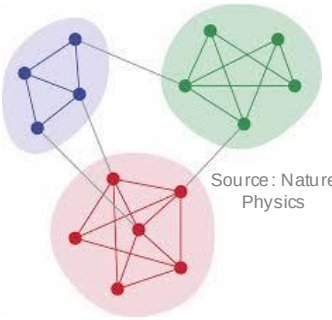
Parallel Computational Analytics:

Parallel Graph Clustering

Motivation

- Graph clustering → use cases across many domains ✓

Networking	Finance	Bioinformatics	Social media
			
Intrusion detection	Fraud detection	Epidemiology & Drug discovery	Recommendation systems



- Accurate graph clustering → computationally expensive ✗
  - Impractical runtime (order of hours) for large graphs with millions of vertices/edges

Approach

Focus on statistically robust stochastic block partitioning (SBP) algorithm

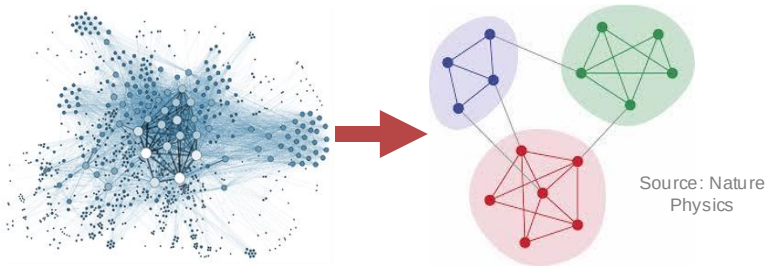
- Accelerate and democratize access to accelerated SBP
- Perform visualization and analysis tasks using accelerated SBP

Milestones

V2

Memberships:
(Mandatory+Optional), e.g., (2+3)

Parallel Computation



Task 4A (1 + 0) Graph Clustering on heterogenous platforms

- *Leverage GPUs and/or FPGAs to accelerate and/or optimize graph clustering via SyCL/OpenCL/Chapel*

Task 4B (1 + 0) Democratizing access to cutting-edge graph clustering research

- *Incorporate prior years' research artifacts into graph-tool library*

Visualization & Analysis

Task 4C (0 + 1) *Rigorous performance vs. accuracy evaluation of graph clustering algorithms via parallel computing (Louvain, Leiden, GossipMap, ..., SBP)*

Task 4D (0 + 1) *Effective visualization of graph clustering output for large graphs (2D or 3D via AR/VR)*

Task 4E (0 + 1) *Explore challenges in applying graph clustering to real-world applications (anomaly detection, bioinformatics, etc…)*