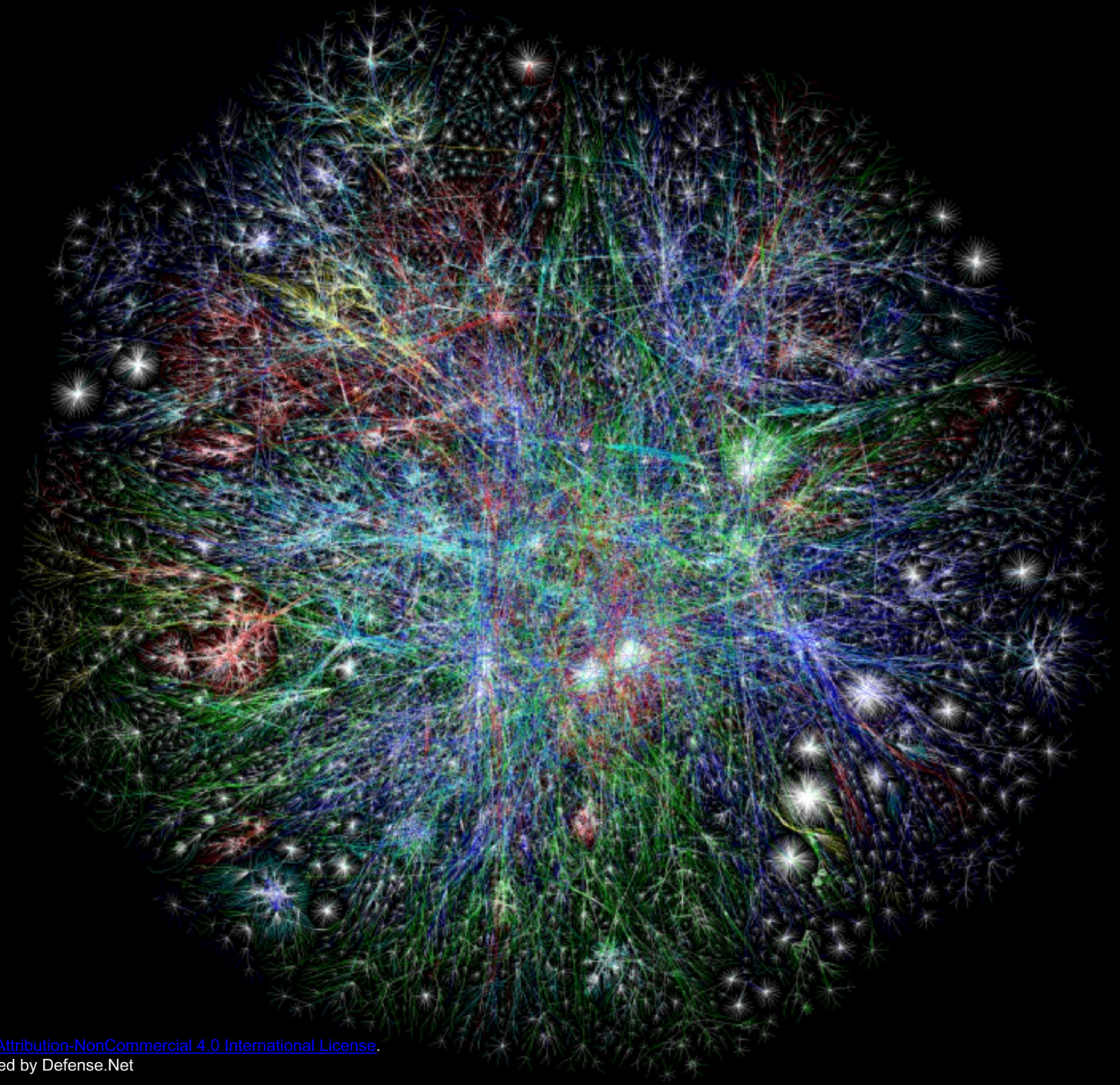

DLI Accelerated Data Science Teaching Kit

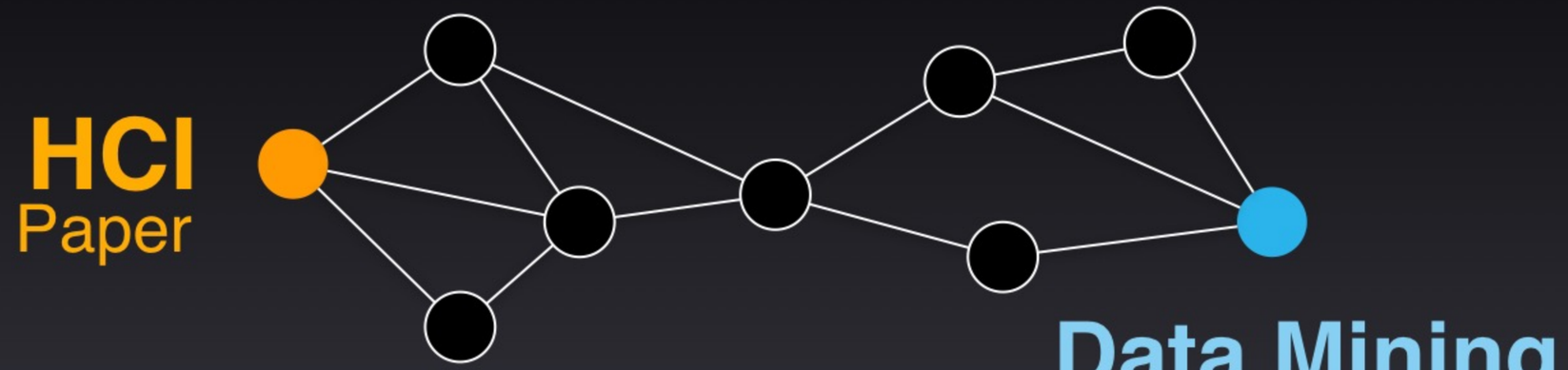
# Lecture 1.6 - Example Data Science Project 1: Apolo Graph Exploration

# **Apolo** Graph Exploration:
# Machine Learning + Visualization

**Apolo: Making Sense of Large Network Data by Combining Rich User Interaction and Machine Learning**.
Duen Horng (Polo) Chau, Aniket Kittur, Jason I. Hong, Christos Faloutsos. CHI 2011.

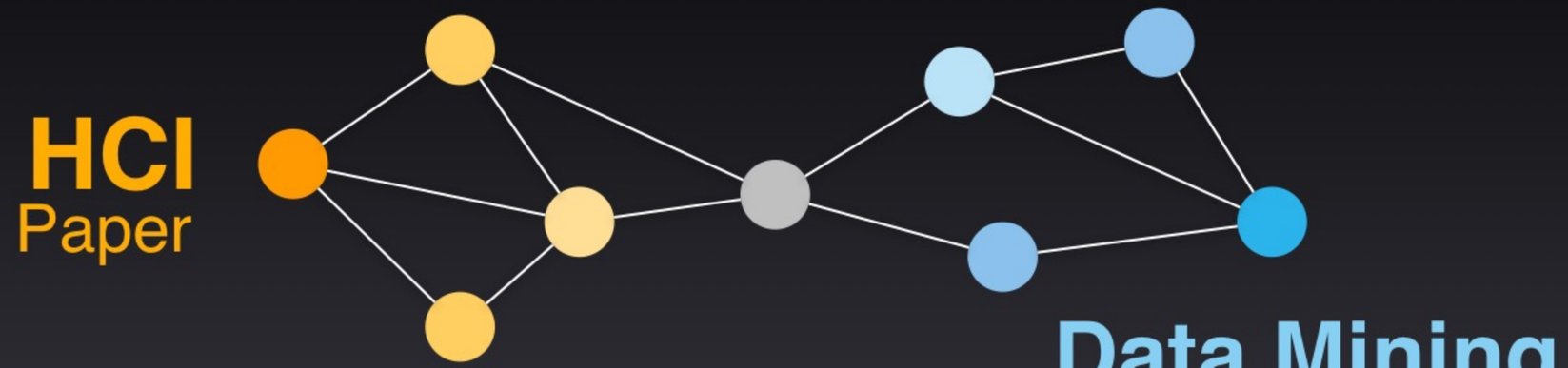


# Finding More Relevant Nodes



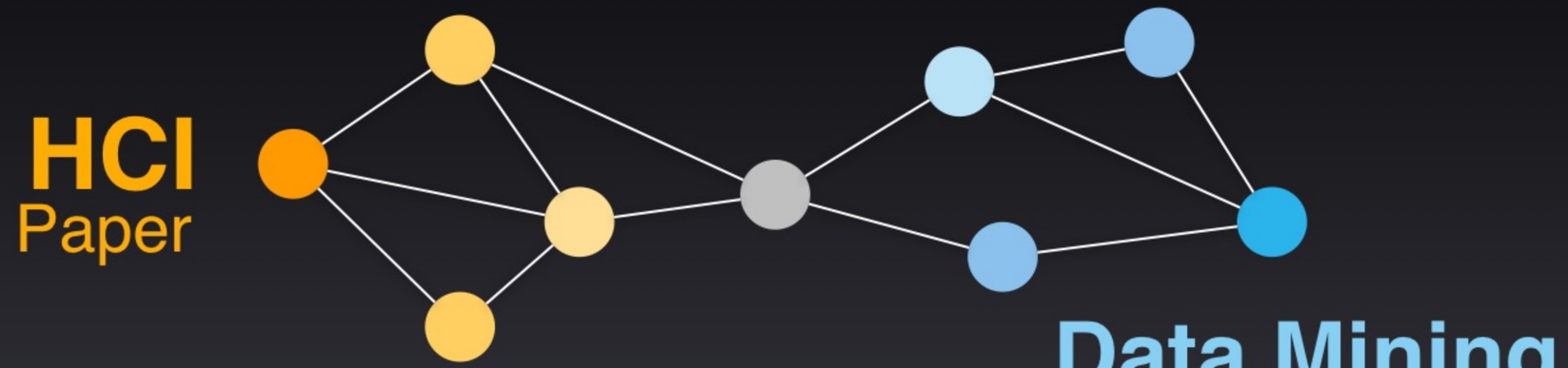

Citation network

# Finding More Relevant Nodes




Citation network

# Finding More Relevant Nodes

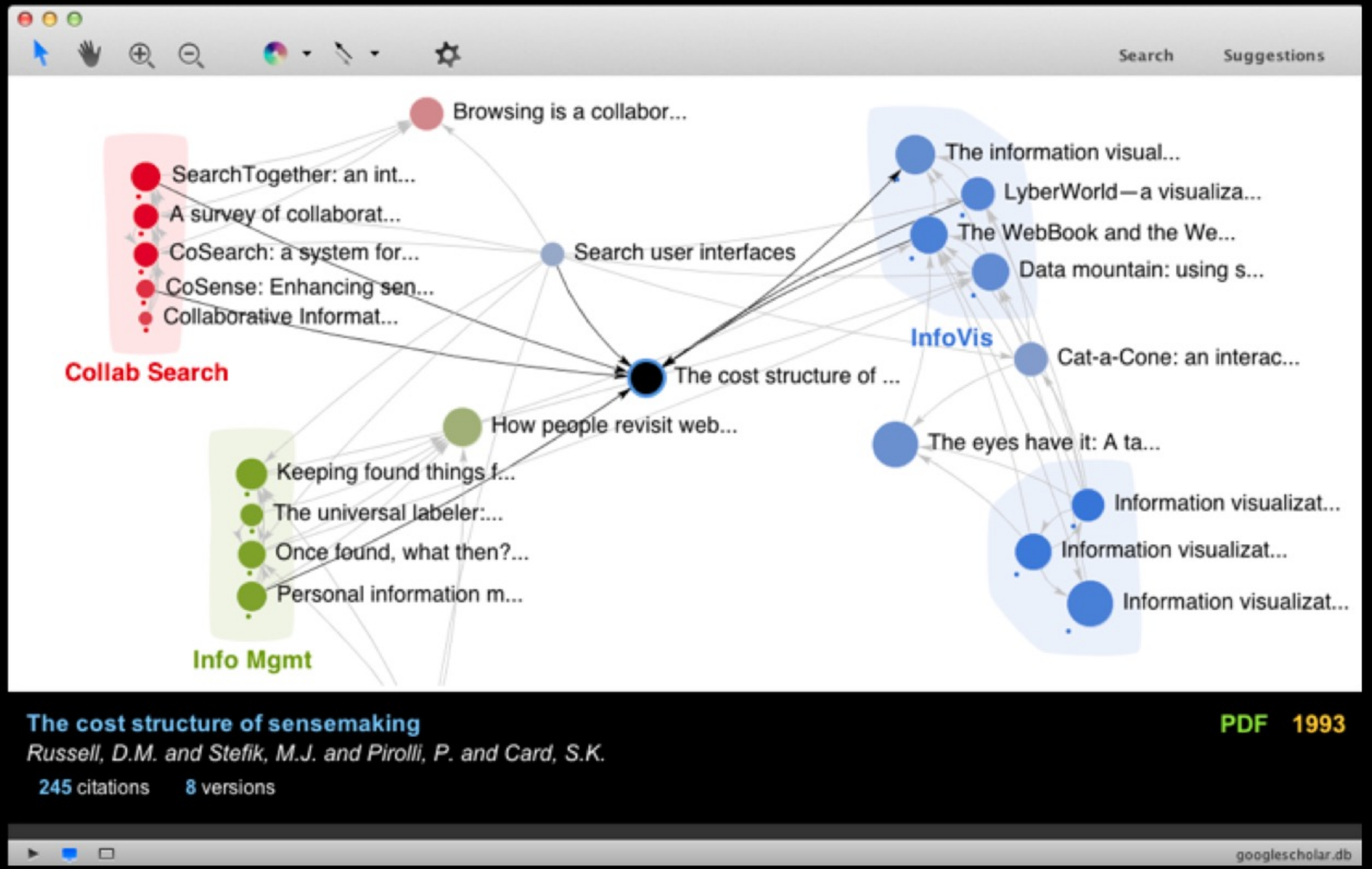


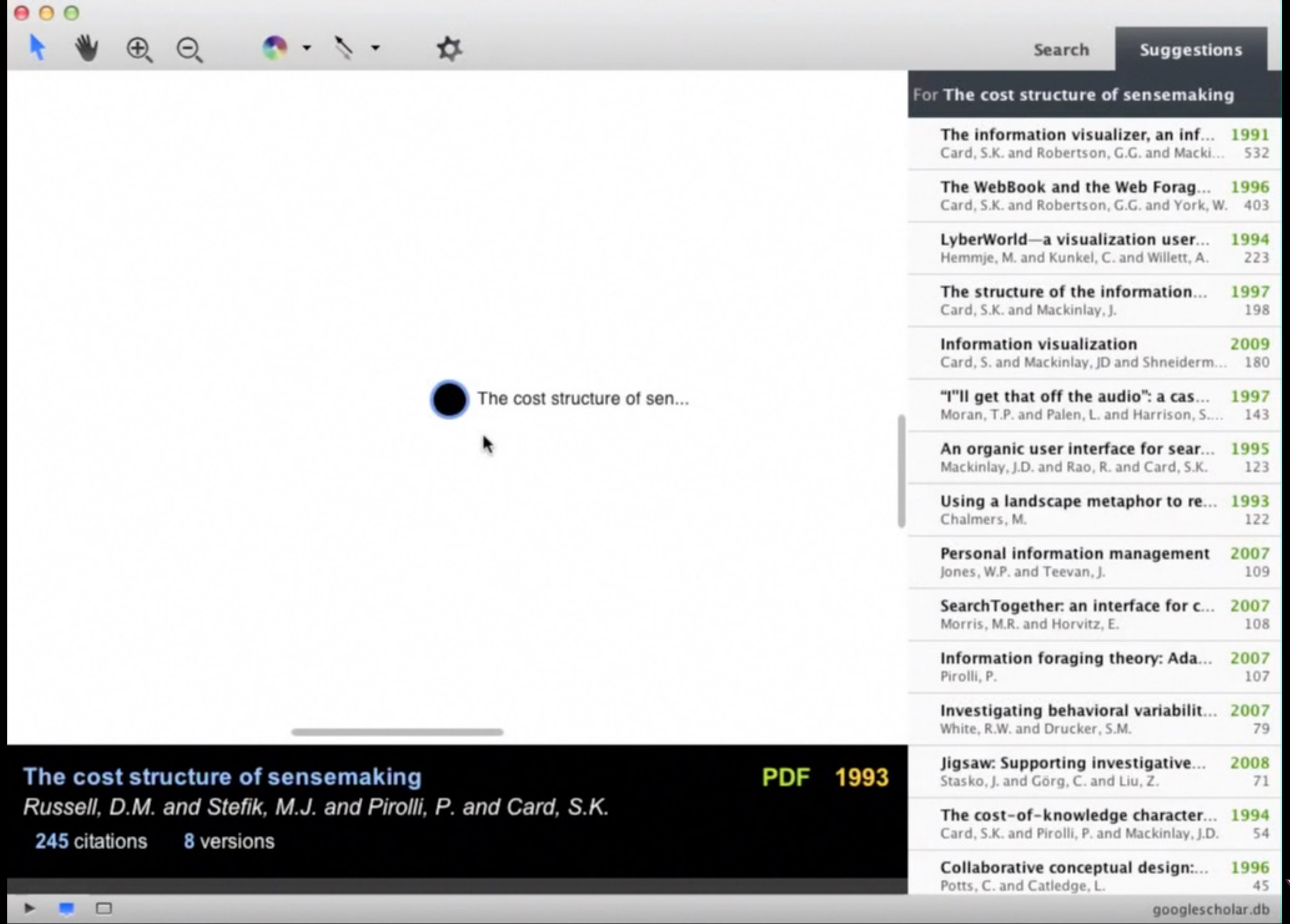

Citation network

Apolo uses **guilt-by-association**
(Belief Propagation)

# *Demo*: Mapping the Sensemaking Literature

**Nodes**: 80k papers from Google Scholar (node size: #citation)
**Edges**: 150k citations

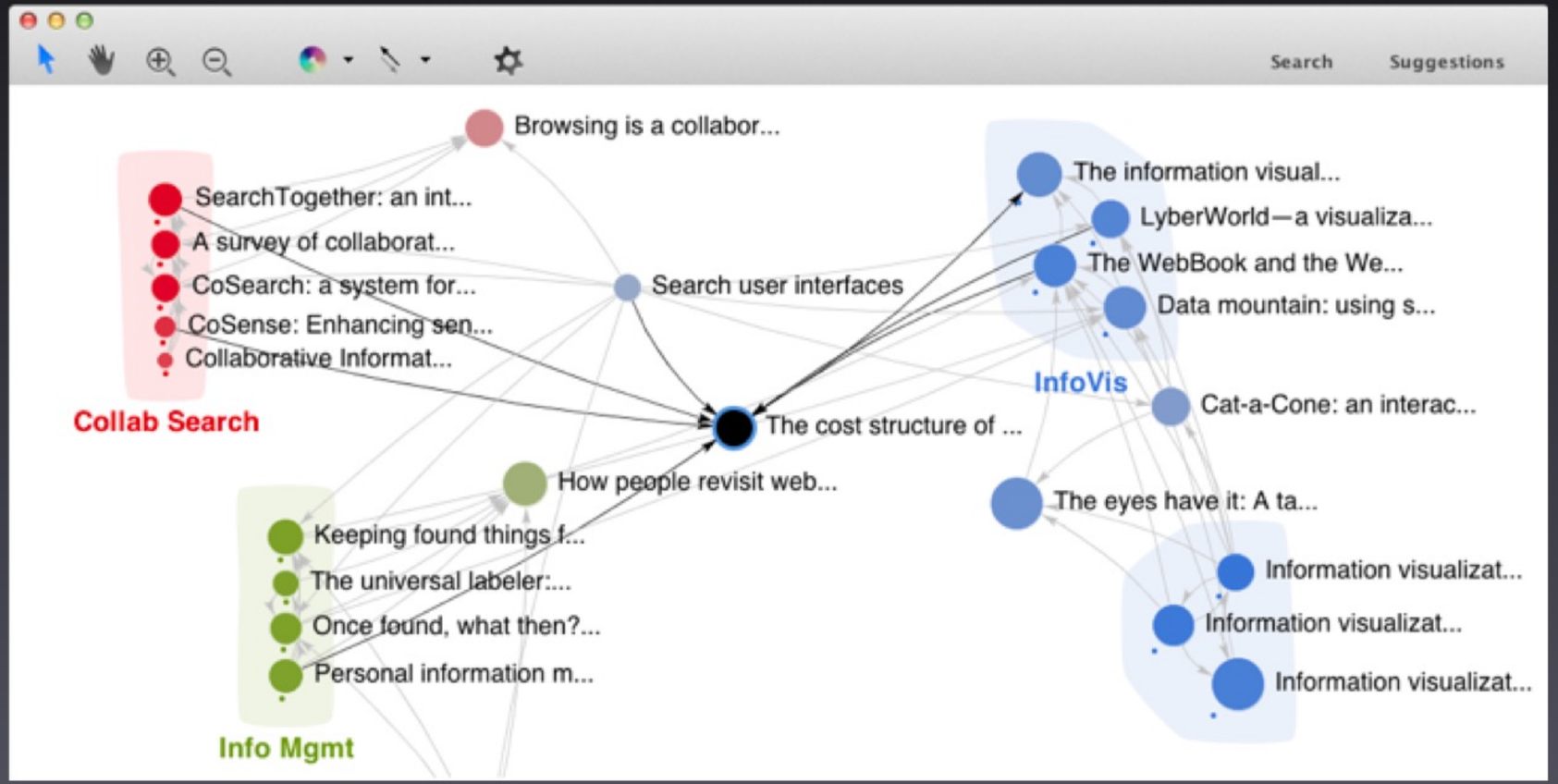




# Key Ideas (Recap)

Specify exemplars

Find other relevant nodes (BP)



# What did **Apolo** go through?

| | |
|---|---|
| Collection | Scrape Google Scholar. No API. 😩 |
| Cleaning | |
| Integration | |
| Analysis | Design inference algorithm<br>(Which nodes to show next?) |
| Visualization | Interactive visualization you just saw |
| Presentation | Paper, talks, lectures |
| Dissemination | Working on re-creating Apolo as web app |

# Apolo: Making Sense of Large Network Data by Combining Rich User Interaction and Machine Learning

**Duen Horng (Polo) Chau, Aniket Kittur, Jason I. Hong, Christos Faloutsos**
School of Computer Science
Carnegie Mellon University
Pittsburgh, PA 15213, USA
{dchau, nkittur, jasonh, christos}@cs.cmu.edu

**ABSTRACT**

Extracting useful knowledge from large network datasets has become a fundamental challenge in many domains, from scientific literature to social networks and the web. We introduce Apolo, a system that uses a mixed-initiative approach—combining visualization, rich user interaction and machine learning—to guide the user to incrementally and interactively explore large network data and make sense of it. Apolo engages the user in bottom-up sensemaking to gradually build up an understanding over time by starting small, rather than starting big and drilling down. Apolo also helps users find relevant information by specifying exemplars, and then using a machine learning method called Belief Propagation to infer which other nodes may be of interest. We evaluated Apolo with twelve participants in a between-subjects study, with the task being to find relevant new papers to update an existing survey paper. Using expert judges, participants using Apolo found significantly more relevant papers. Subjective feedback of Apolo was also very positive.

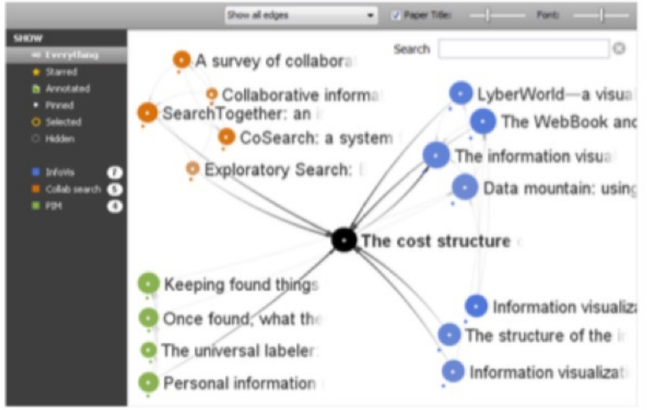



**Figure 1. Apolo displaying citation network data around the article *The Cost Structure of Sensemaking*. The user gradually builds up a mental model of the research areas around the article by manually inspecting some neighboring articles in the visualization and specifying them as exemplar articles (with colored dots underneath) for some ad hoc groups, and instructs Apolo to find more articles relevant to them.**

**Apolo: Making Sense of Large Network Data by Combining Rich User Interaction and Machine Learning**. Duen Horng (Polo) Chau, Aniket Kittur, Jason I. Hong, Christos Faloutsos. *ACM Conference on Human Factors in Computing Systems (CHI) 2011*. May 7-12, 2011.

back; H.5.2 Information Interfaces and Presentation: User

the new domain to understand and contribute to it


DEEP
LEARNING
INSTITUTE


PRAIRIE VIEW
A&M UNIVERSITY

DLI Accelerated Data Science Teaching Kit

# Thank You