



DEEP  
LEARNING  
INSTITUTE



DLI Accelerated Data Science Teaching Kit

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

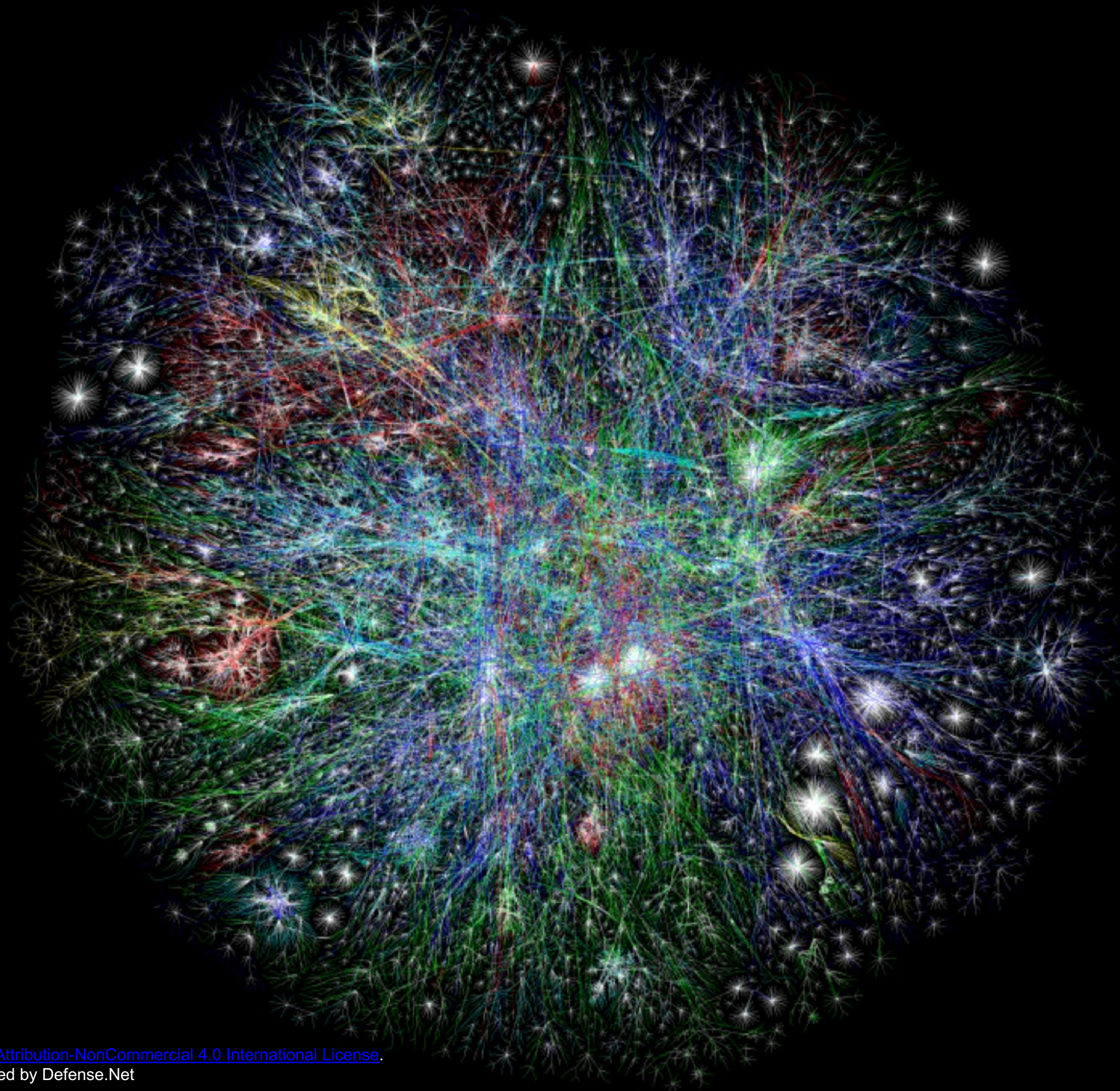


The Accelerated Data Science Teaching Kit is licensed by NVIDIA, Georgia Institute of Technology, and Prairie View A&M University under the [Creative Commons Attribution-NonCommercial 4.0 International License](https://creativecommons.org/licenses/by-nc/4.0/).

# 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.





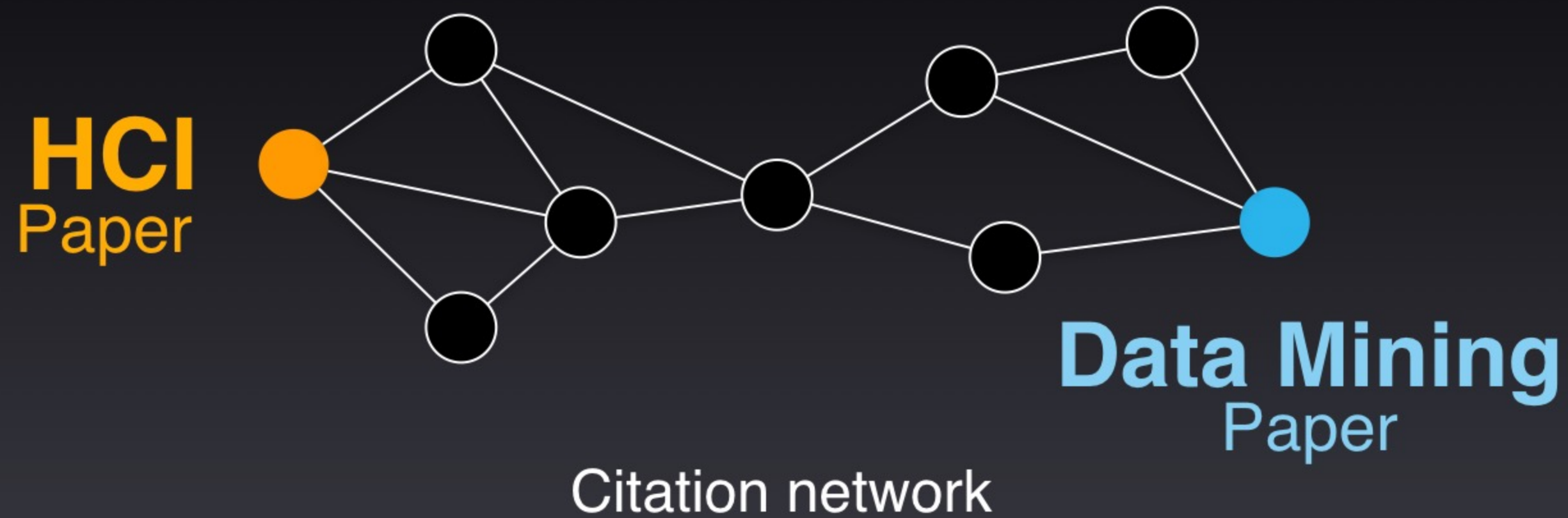




# BEAUTIFUL HAIRBALL DEATH STAR SPAGHETTI



# Finding **More** Relevant Nodes



# Finding **More** Relevant Nodes



# Finding **More** Relevant Nodes



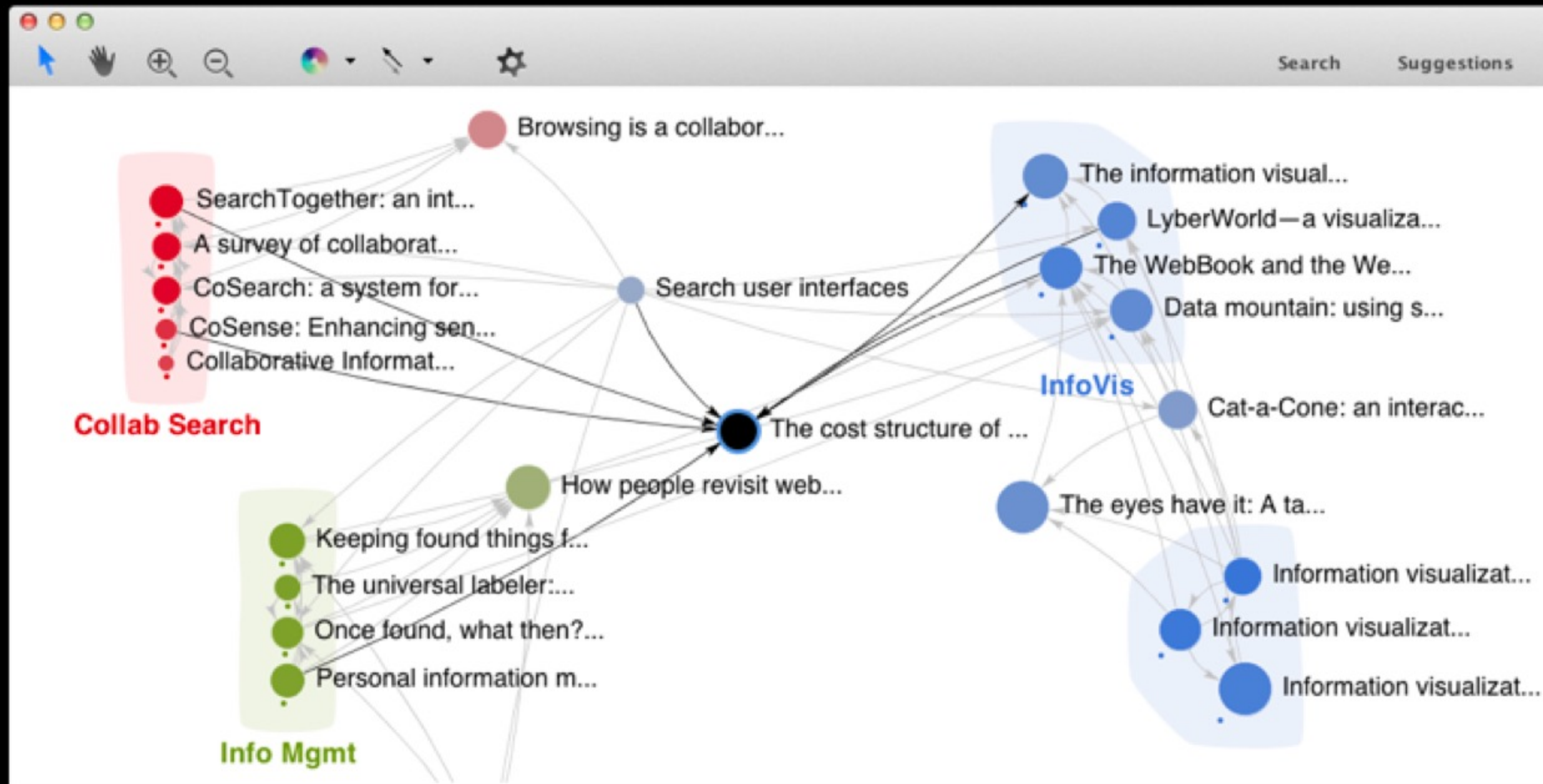
Apolo uses **guilt-by-association**  
(Belief Propagation)



# Demo: Mapping the Sensemaking Literature

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

**Edges:** 150k citations



**The cost structure of sensemaking**

Russell, D.M. and Stefik, M.J. and Pirolli, P. and Card, S.K.

245 citations 8 versions

PDF 1993

For The cost structure of sensemaking

<b>The information visualizer, an inf...</b>	<b>1991</b>
Card, S.K. and Robertson, G.G. and Macki...	532
<b>The WebBook and the Web Forag...</b>	<b>1996</b>
Card, S.K. and Robertson, G.G. and York, W.	403
<b>LyberWorld—a visualization user...</b>	<b>1994</b>
Hemmje, M. and Kunkel, C. and Willett, A.	223
<b>The structure of the information...</b>	<b>1997</b>
Card, S.K. and Mackinlay, J.	198
<b>Information visualization</b>	<b>2009</b>
Card, S. and Mackinlay, JD and Shneiderm...	180
<b>"I'll get that off the audio": a cas...</b>	<b>1997</b>
Moran, T.P. and Palen, L. and Harrison, S...	143
<b>An organic user interface for sear...</b>	<b>1995</b>
Mackinlay, J.D. and Rao, R. and Card, S.K.	123
<b>Using a landscape metaphor to re...</b>	<b>1993</b>
Chalmers, M.	122
<b>Personal information management</b>	<b>2007</b>
Jones, W.P. and Teevan, J.	109
<b>SearchTogether: an interface for c...</b>	<b>2007</b>
Morris, M.R. and Horvitz, E.	108
<b>Information foraging theory: Ada...</b>	<b>2007</b>
Pirolli, P.	107
<b>Investigating behavioral variabilit...</b>	<b>2007</b>
White, R.W. and Drucker, S.M.	79
<b>Jigsaw: Supporting investigative...</b>	<b>2008</b>
Stasko, J. and Görg, C. and Liu, Z.	71
<b>The cost-of-knowledge character...</b>	<b>1994</b>
Card, S.K. and Pirolli, P. and Mackinlay, J.D.	54
<b>Collaborative conceptual design:...</b>	<b>1996</b>
Potts, C. and Catledge, L.	45

The cost structure of sen...

**The cost structure of sensemaking**  
*Russell, D.M. and Stefik, M.J. and Pirolli, P. and Card, S.K.*  
245 citations 8 versions

PDF 1993

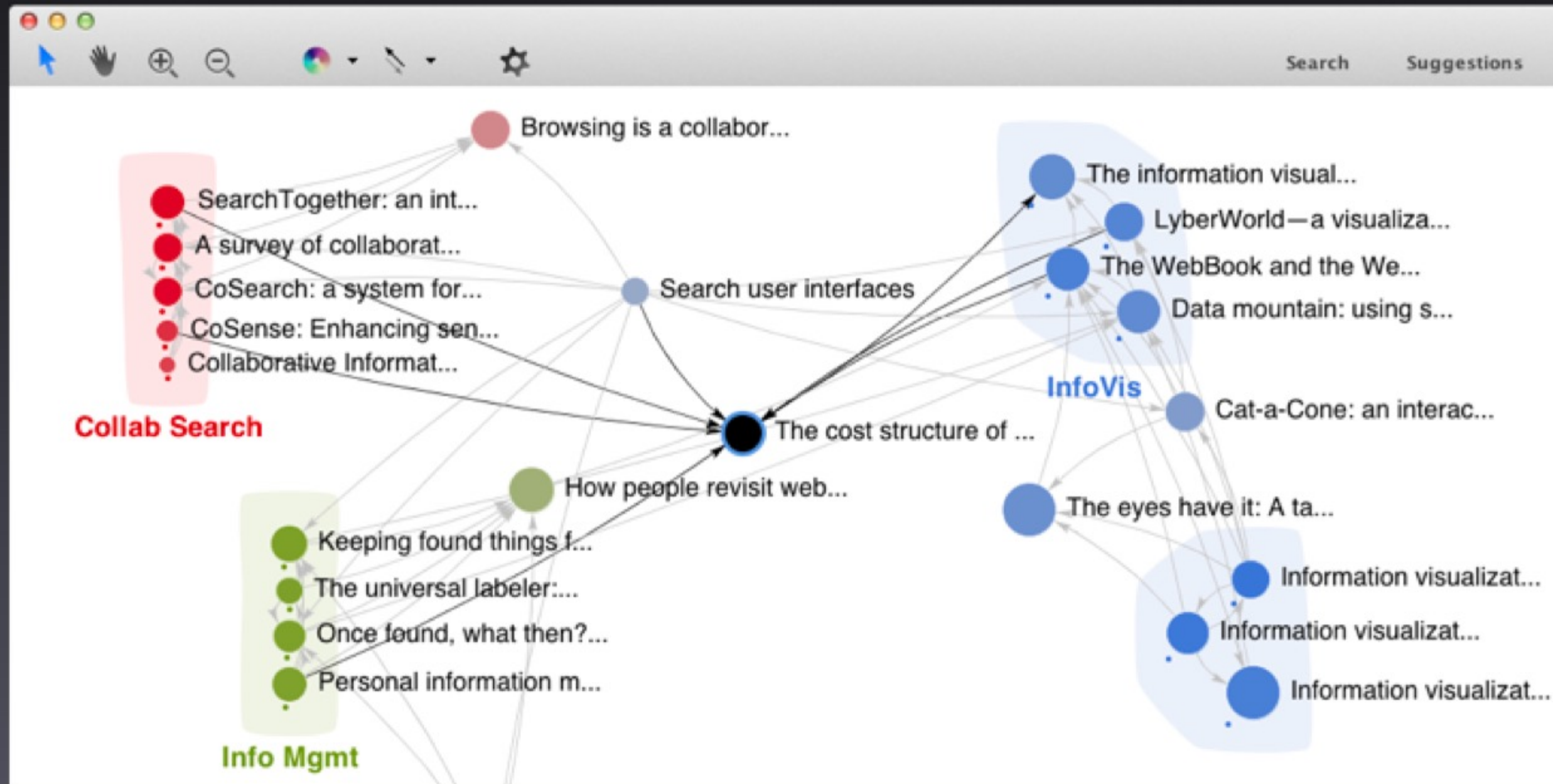


# Key Ideas (Recap)



Specify **exemplars**

Find **other** relevant nodes (BP)



# What did **Apolo** go through?

Collection

Scrape Google Scholar. No API. 🙄

Cleaning

Integration

Analysis

Design inference algorithm  
(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.

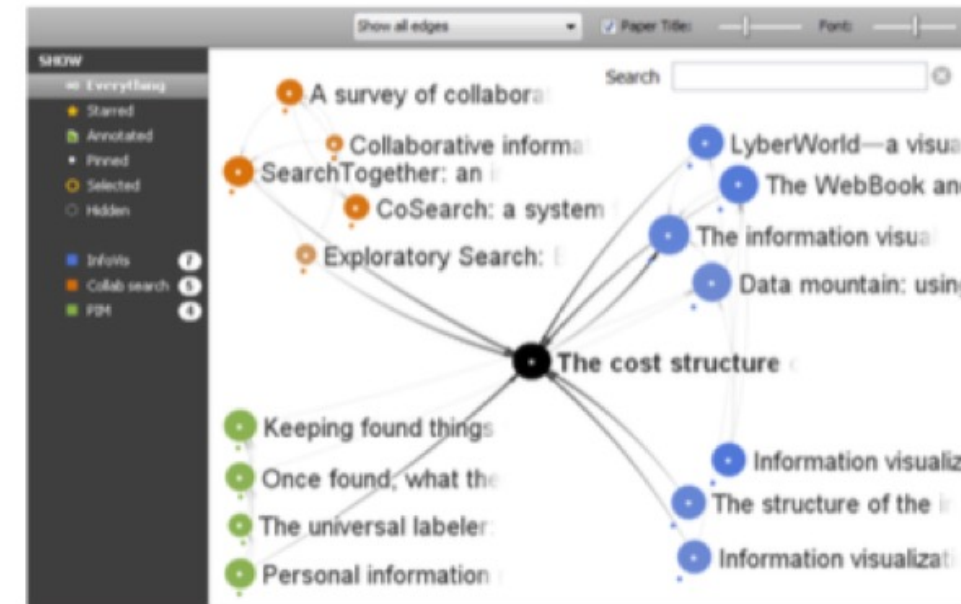


Figure 1. Apolo displaying citation network data around the article *The Cost Structure of Sensemaking*. The user gradually builds up a mental model of the research areas around the article by manually inspecting some neighboring articles in the visualization and specifying them as exemplar articles (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