



Untangling Influence and Desire

Visual analysis of massive graph data

David Jonker, Scott Langevin



uncharted

Strata + Hadoop World, London 2015

“The answer to life, the universe and everything...”

A screenshot of a Google search page. The search bar contains the text "the answer to life the universe and everything". Below the search bar, there are navigation tabs for "Web", "Images", "Videos", "Maps", "News", "More", and "Search tools". The search results show "About 60,500,000 results (0.57 seconds)". A prominent result is a calculator interface with the text "The answer to life the universe and everything =" above it. The calculator's display shows the number "42". The calculator has various buttons including "Rad", "Inv", "π", "e", "Ans", "EXP", "x!", "ln", "tan", "x'", "(", ")", "7", "8", "9", "4", "5", "6", "1", "2", "3", "0", ".", "=", and "+". A "More info" link is visible at the bottom right of the calculator interface.

When is a simple
answer too simple?

Why Seeing the Universe Matters

Believability / Trust



Perception of Truth

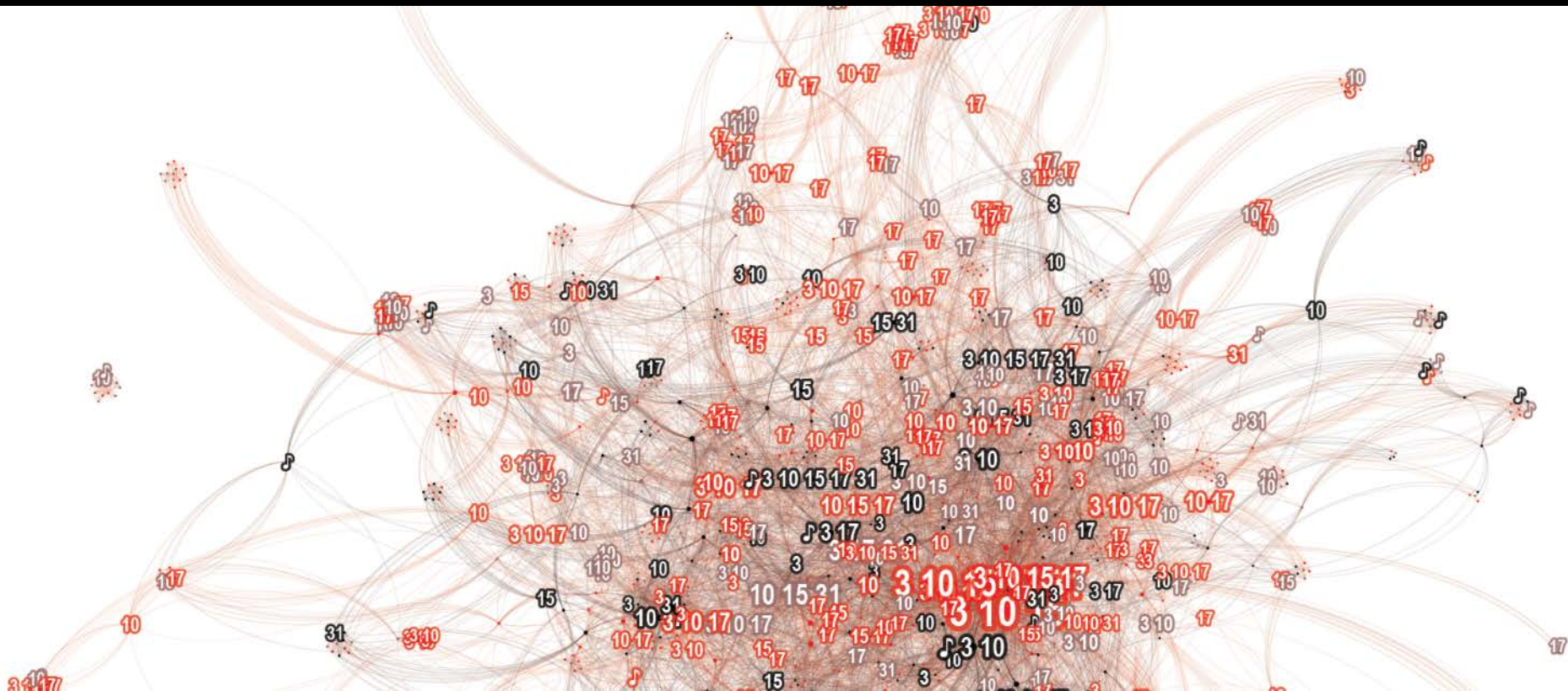


Pattern and Anomaly Detection

Effective visual analytics combine the best of both human and computational capabilities.

Example of a Small Universe

Sample of NBA Fan Community Chatter, Toronto Raptors

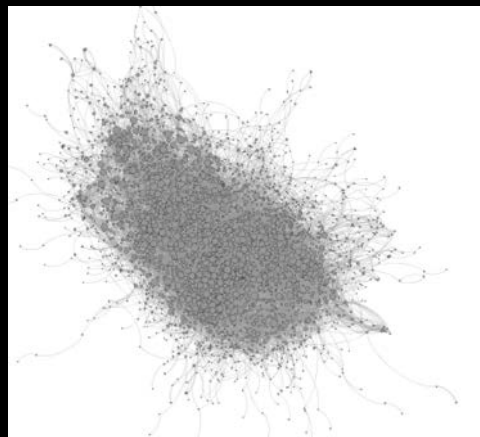


Challenges of Big Graphs as Maps

Should form a map of communities and relationships

BUT

- Often prohibitively costly to compute layout
- Often an unresolved hairball that reveals little
- Often prohibitively costly to render in real-time
- Ineffective labeling impedes understanding
- How to drill down from a 10,000 foot view?



Chelsea FC Fan Communities

Twitter mentions graph amongst accounts that have tweeted about Chelsea FC team or players, Aug 2014 – Feb 2015

248,747,072 tweets

554,430 account nodes

100,700 mention links

Demo
[skip](#)

Geo Communities

Fans Mention Graph



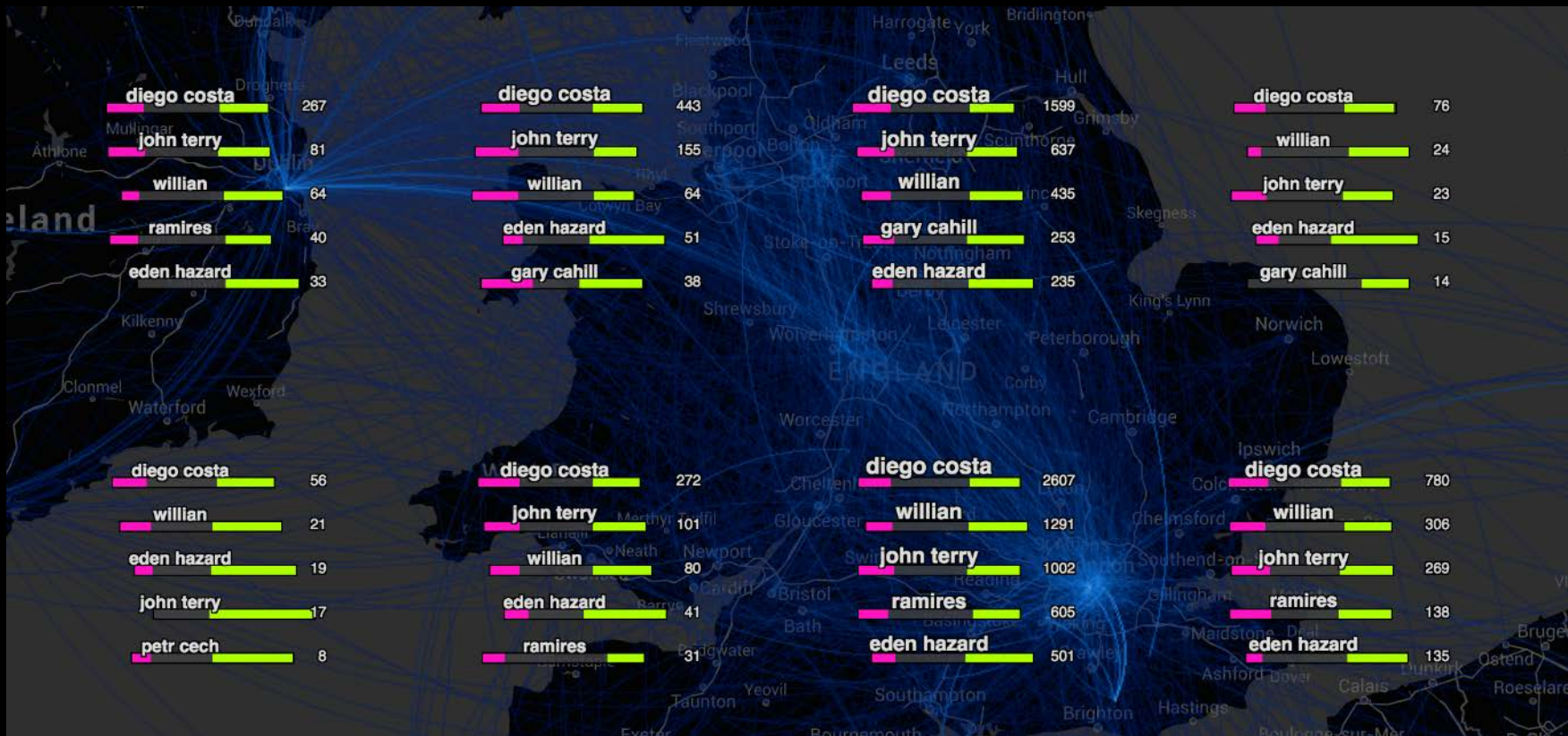
Geo Communities

Top Fans



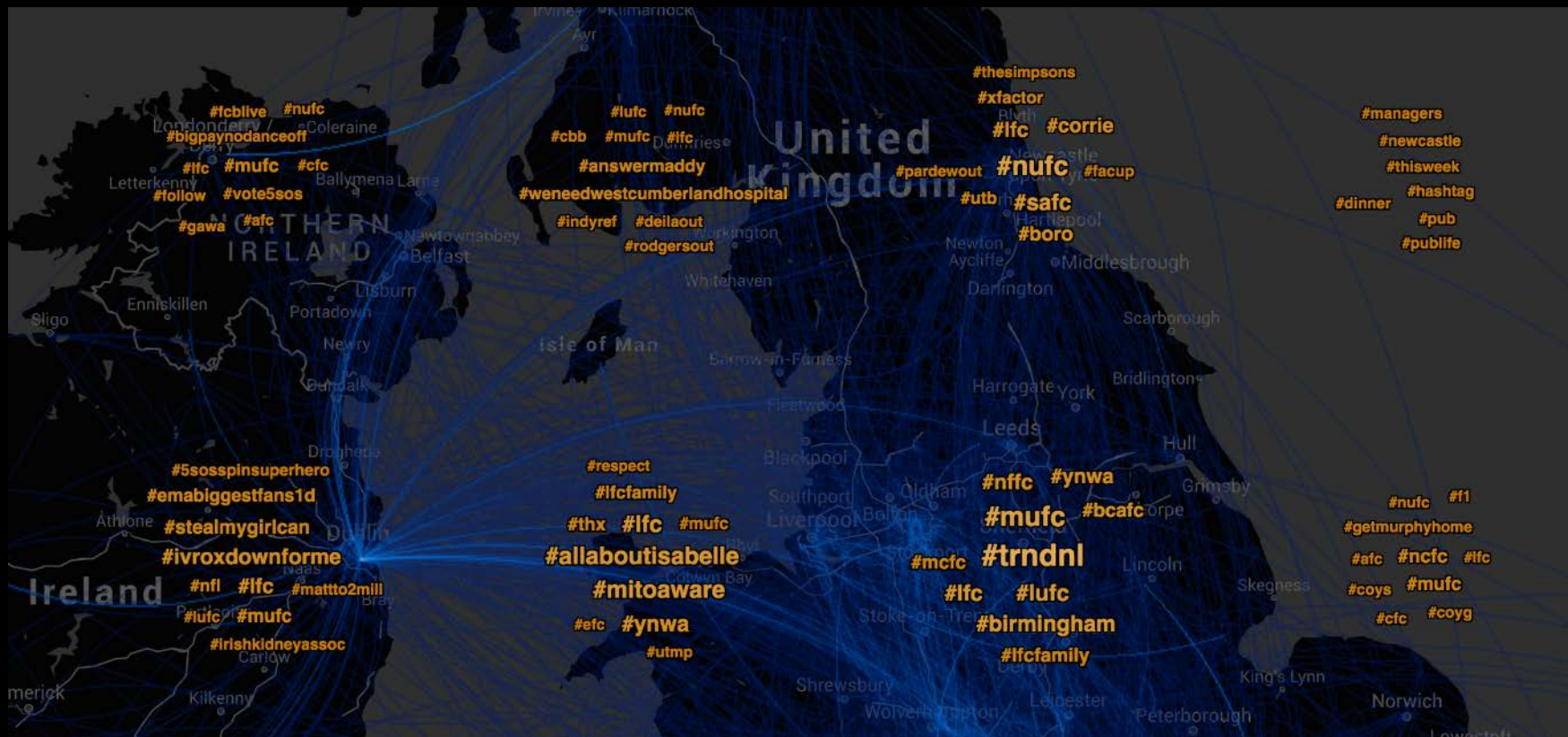
Geo Communities

Player Sentiment



Geo Communities

Top Hashtags



Community Graph



Technical Approach

Tile-Based Community Graph Visualization

- Distributed hierarchical community clustering.
- Hierarchical community-based graph layout.
- Multi-scale node and link tile-based rendering with LOD.
- Layered community analytics.

Aperture Tiles

TILE-BASED VISUAL ANALYTICS

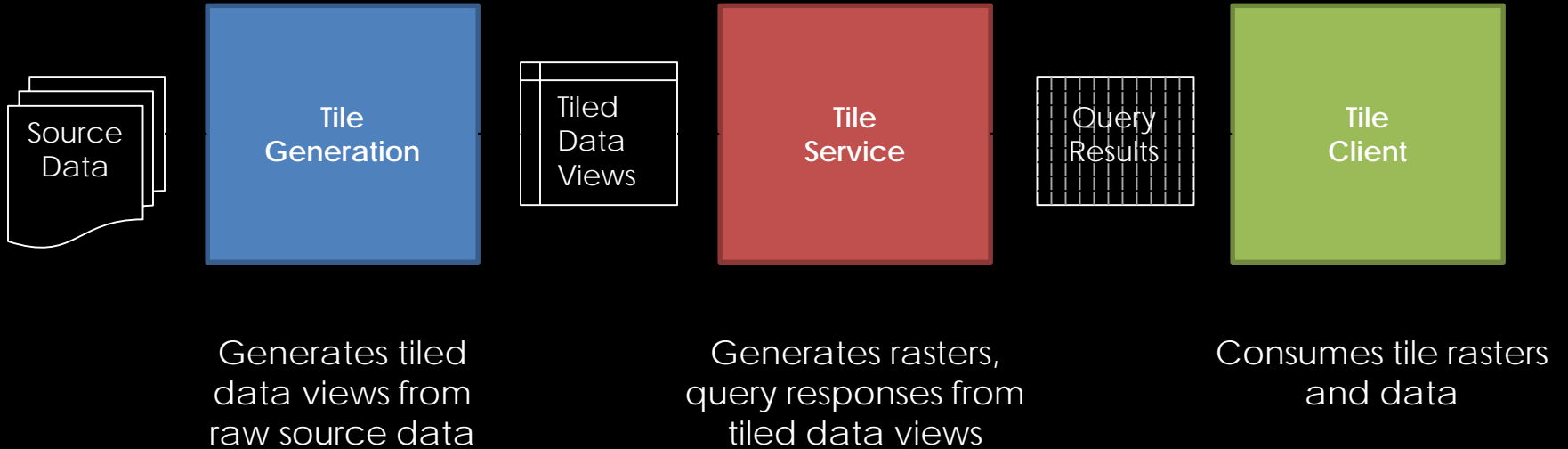
- Hierarchical data tiling using cluster computing.
- Interactive on-demand image tile generation.
- Layers of raw data and derivative analytics.

OPEN SOURCE

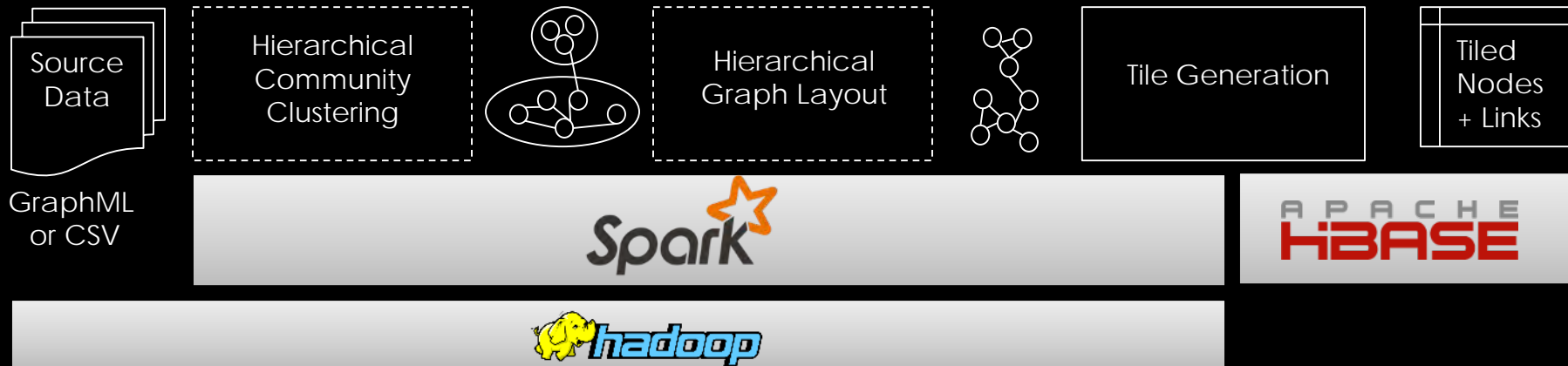
- Uncharted research product.
- Built on Apache Spark and Hadoop.
- Cross Plots, Geospatial, Time-series, **Graph analysis**



Three Tiers

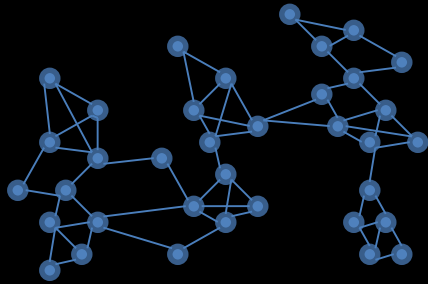


Graph Tiling Pipeline

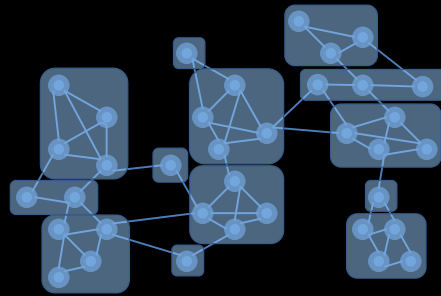


 Optional

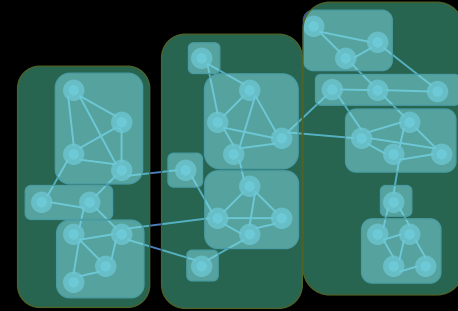
Hierarchical Community Clustering



Raw Graph

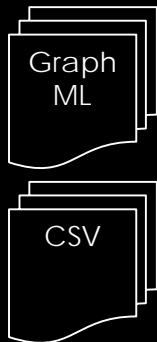


Level 1 Communities



Level 2 Communities

...

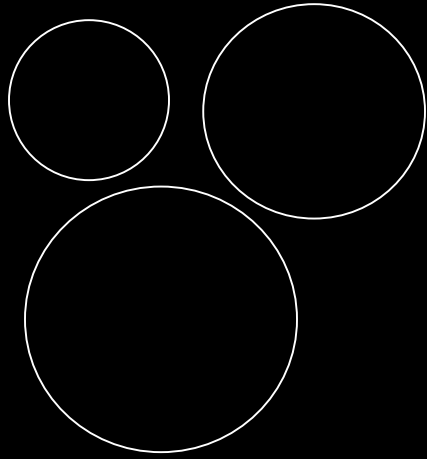


Distributed Louvain Modularity

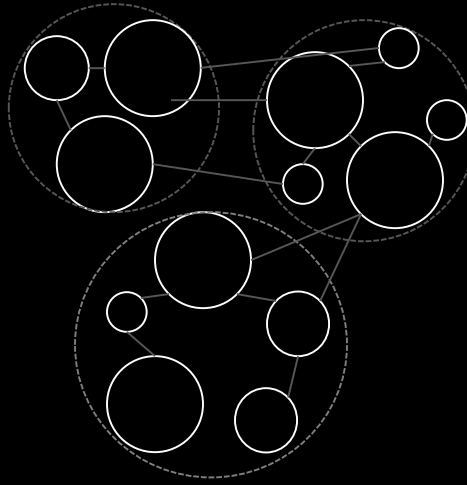


Hierarchical Communities

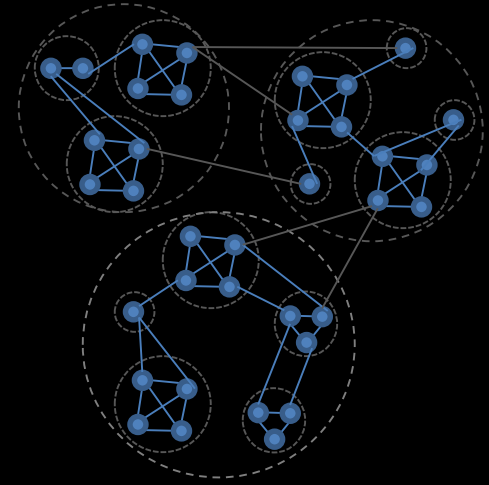
Hierarchical Graph Layout



Level 2 Communities Layout



Level 1 Communities Layout



Raw Nodes Layout

Hierarchical
Communities

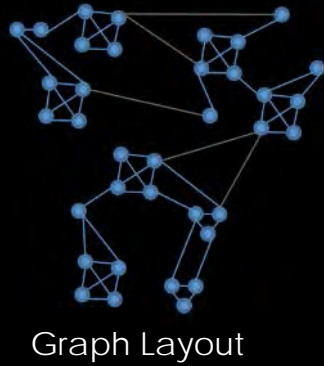


Hierarchical
Force Directed Layout



Graph Layout

Hierarchical Tile Generation



- Aggregated Nodes + Edges
- Community Analytics



Product Affinity

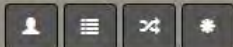
Stanford collection of anonymized Amazon product reviews and
"customers who bought this also bought..." over 9 years

2,372,409 product and customer nodes

9,909,551 review and co-purchase links


Demo
skip





Showing 11 of 11 results, ordered by Entity Class, Popularity Percentile

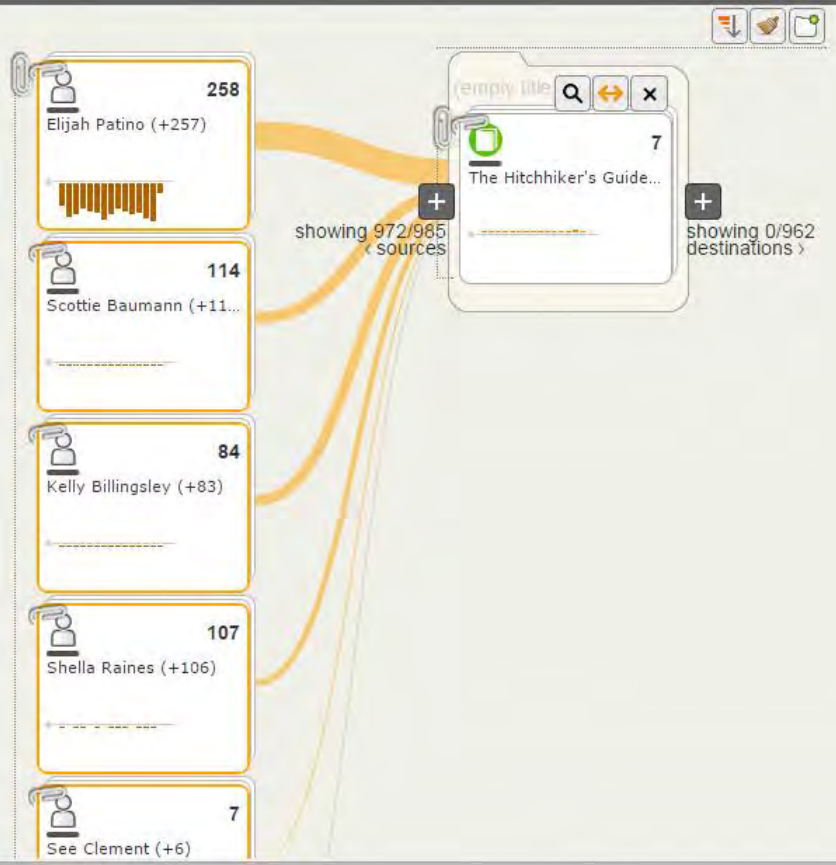
<input type="checkbox"/>	Name	Entity Class 	 Popularity Percentile	Reviews	Avg Rating
--------------------------	------	--	---	---------	------------

Entity Class: Book 

<input checked="" type="checkbox"/>	The Hitchhiker's Guide to the Galaxy	 Book	100	579	4.5
	▶ ID: 1906711 • Entity Type: product • Product No: 0345391802 • Status: active • First Review: Jan 1, 1970 12:00am • Last Review: Jul 24, 2005 3:05am •				
<input checked="" type="checkbox"/>	The Ultimate Hitchhiker's Guide to the Galaxy	 Book	100	73	4.5
	▶ ID: 1281544 • Entity Type: product • Product No: 0345453743 • Status: active • First Review: Oct 27, 2002 3:34am • Last Review: Jul 25, 2005 2:34am •				
<input checked="" type="checkbox"/>	The Hitchhiker's Guide to the Galaxy	 Book	99	171	4.5
	▶ ID: 539090 • Entity Type: product • Product No: 0345418913 • Status: active • First Review: Oct 4, 1996 2:15am • Last Review: Jun 8, 2005 2:15am •				
<input checked="" type="checkbox"/>	The Hitchhiker's Guide to the Galaxy : A Novel	 Book	96	171	4.5
	▶ ID: 1234786 • Entity Type: product • Product No: 0517542099 • Status: active • First Review: Jan 1, 1970 12:02am • Last Review: Jun 8, 2005 2:32am •				
<input checked="" type="checkbox"/>	The Hitchhiker's Guide to the Galaxy	 Book	96	171	4.5
	▶ ID: 1185757 • Entity Type: product • Product No: 1590072561 • Status: active • First Review: Jan 1, 1970 12:00am • Last Review: Jun 8, 2005 2:30am •				
<input checked="" type="checkbox"/>	The Hitchhiker's Guide to the Galaxy	 Book	95	171	4.5

Influential Customers

Jan 1, 2002 to Jan 1, 2006



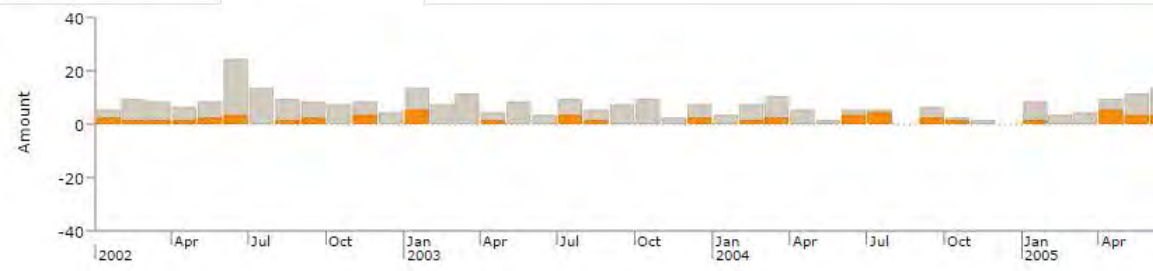
Jan 1, 2002 to Jan 1, 2006

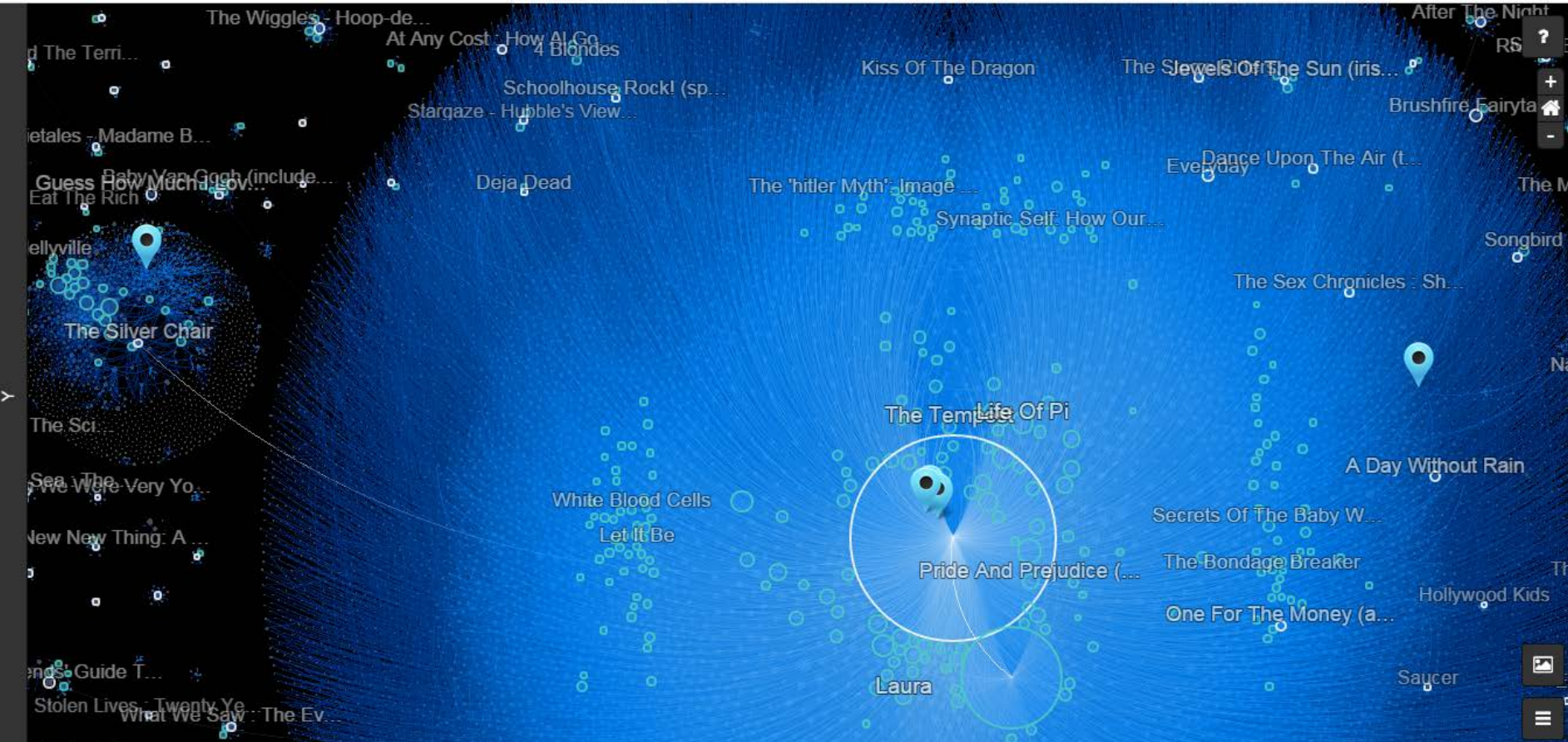


Cluster: So Long, and Thanks for All the Fish (+8)

Transaction Table Transaction Chart

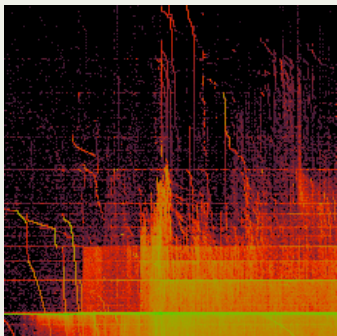
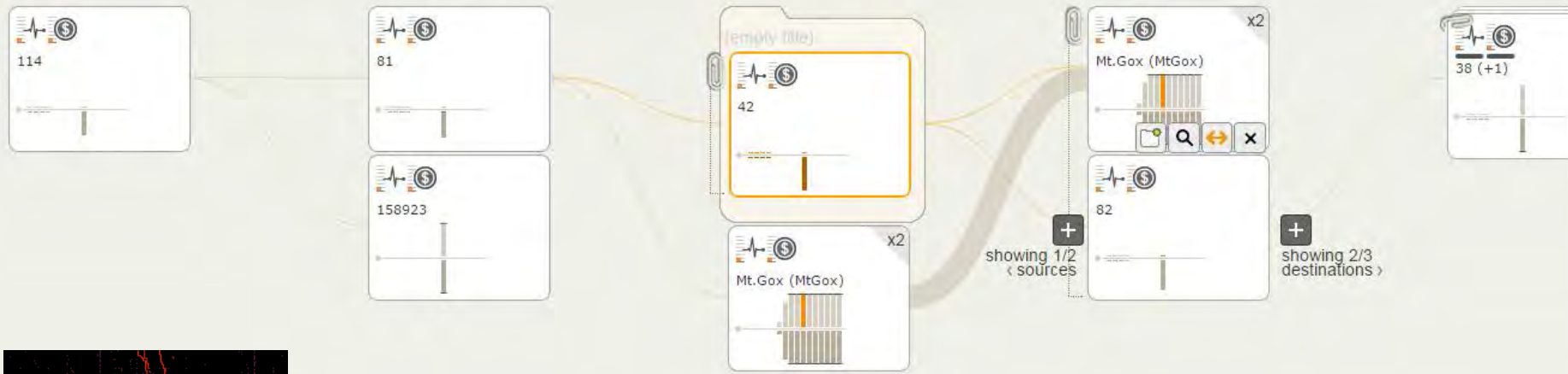
EntityClass
book (9)
100%





?
+
-
[Home icon]
[Image icon]
[Menu icon]

Jul 1, 2009 to Jul 1, 2013

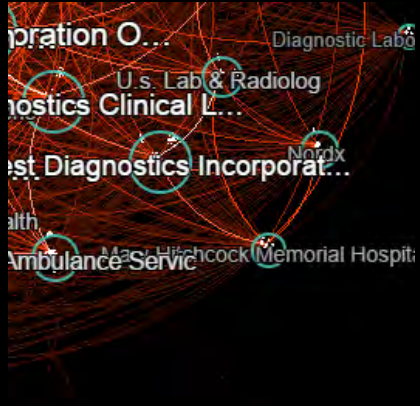


Bitcoin Transaction Correlations

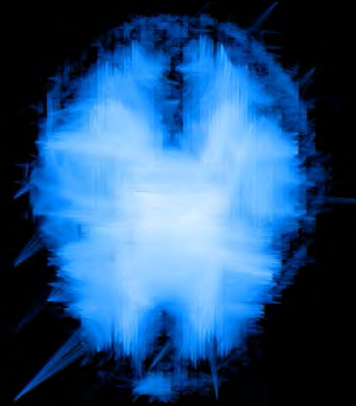


Silk Road Bitcoin Sources within Global Communities

Other Applications



Doc Graph



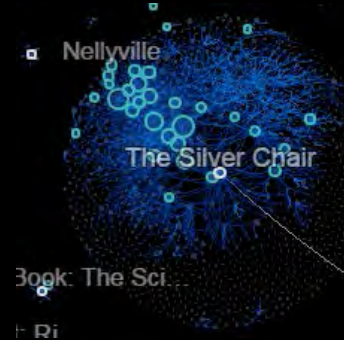
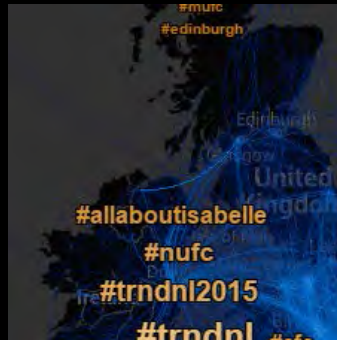
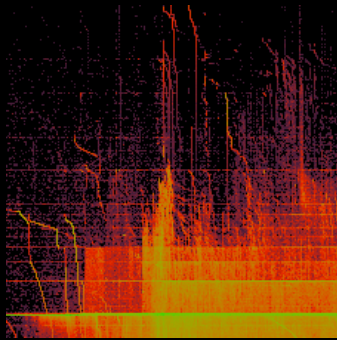
Neural Networks

more
to come

Summary

How much of the big picture is missed with simple charts?

Combining computational analytics with highly expressive
Interactive visualization affords deeper understanding and trust.
Map techniques and cluster computing are highly applicable.

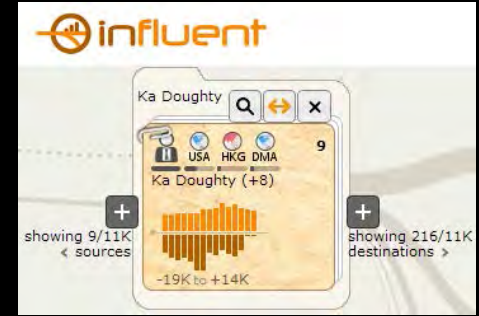
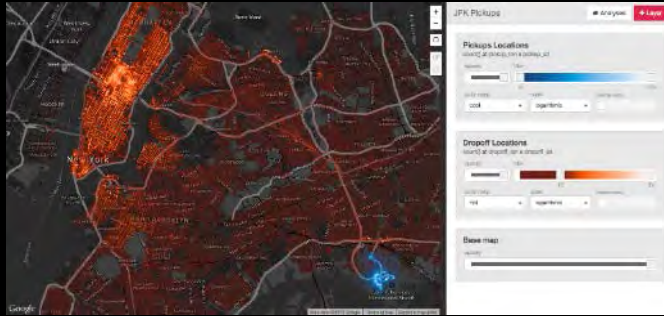


Behavior and influence moves beyond What, to **Why** and **How**.

What's Next?

Products

PanTera™
databricks™



Graph Tiling Pipeline

- Enhanced analytic layers for communities.
- Integration of runtime graph analytics.
- Further optimization.



Questions?

