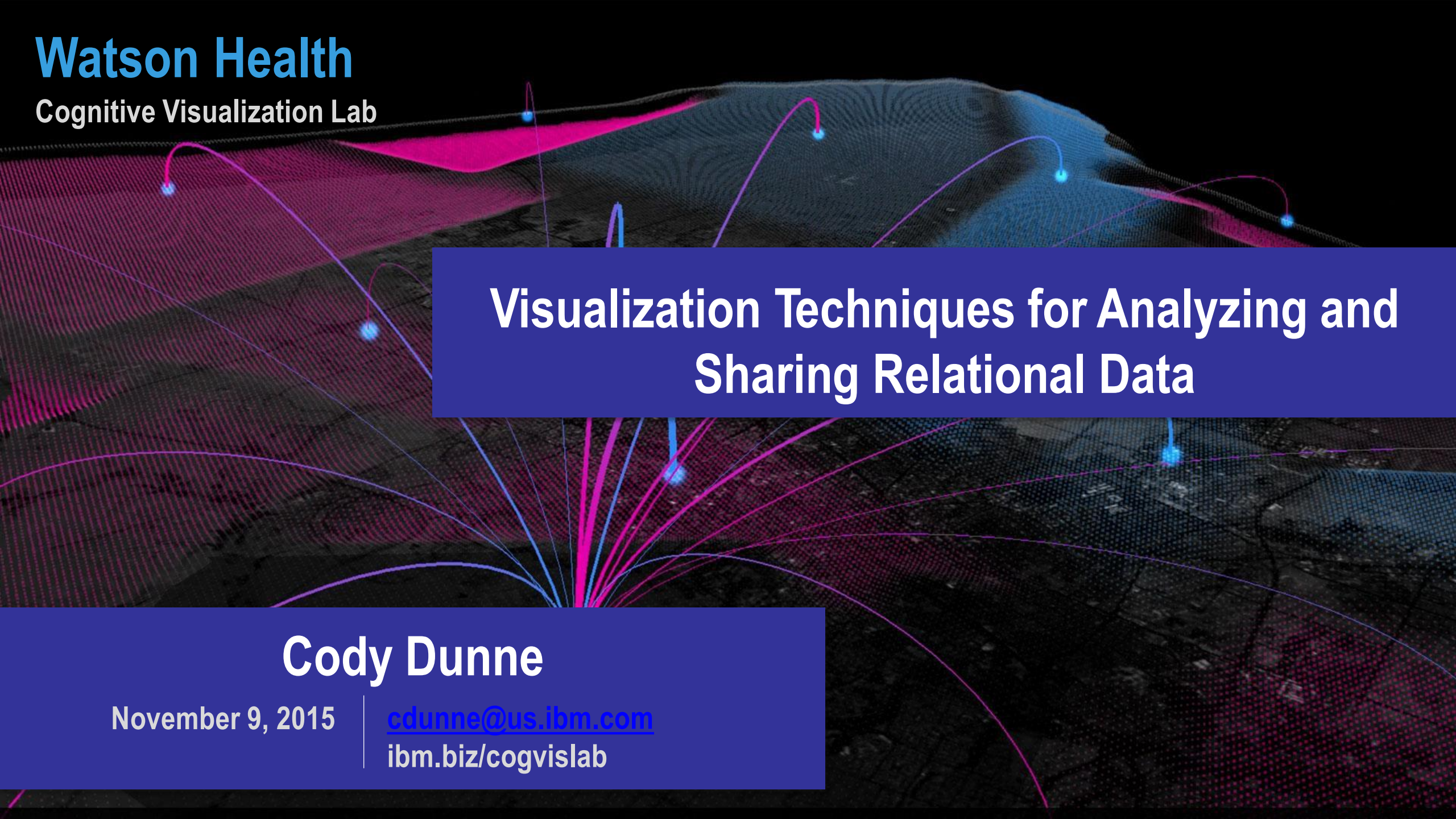


Watson Health

Cognitive Visualization Lab

Visualization Techniques for Analyzing and Sharing Relational Data

Cody Dunne

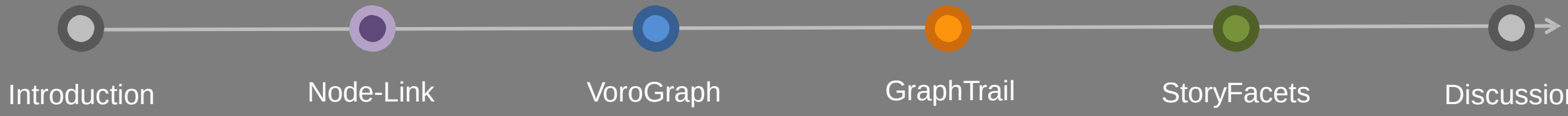
November 9, 2015

cdunne@us.ibm.com

ibm.biz/cogvislab

Visualization Techniques …. Relational Data

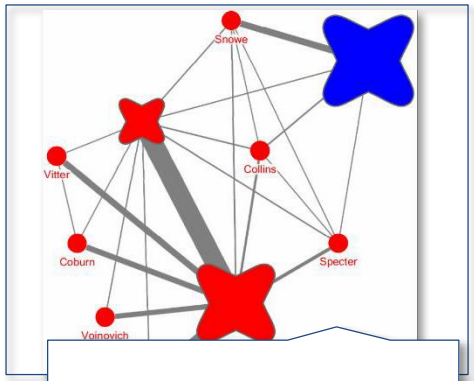
Agenda



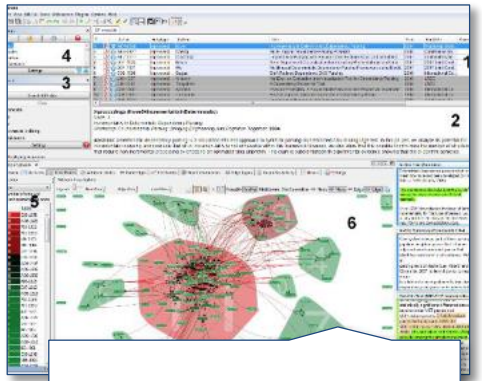
Cody Dunne, PhD – Cognitive Visualization RSM

Web: ibm.biz/codydunne

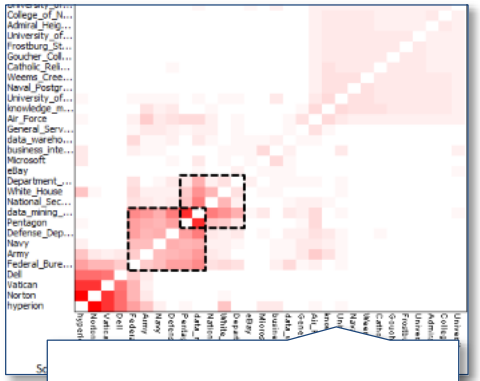
Email: cdunne@us.ibm.com



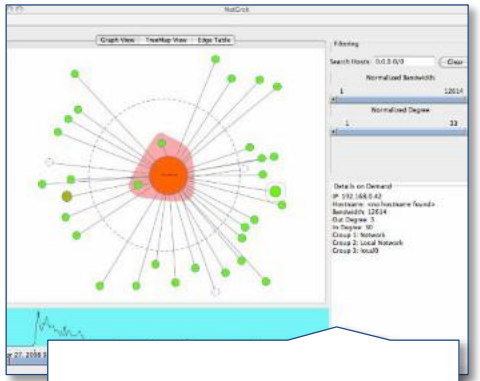
aggregation techniques



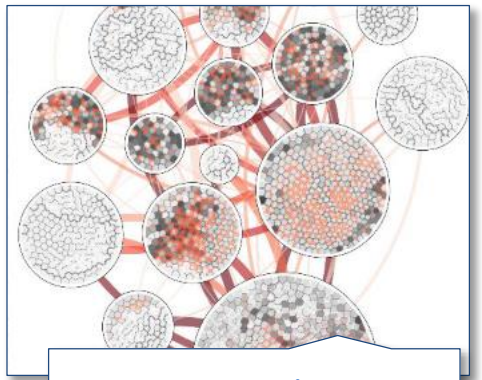
literature exploration



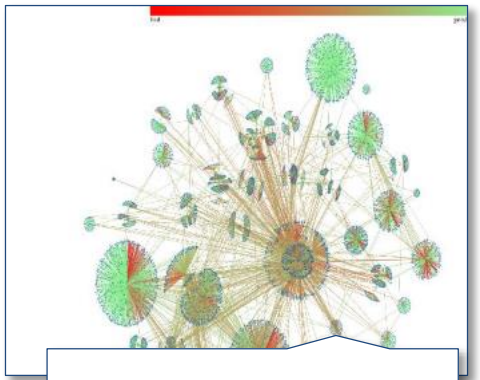
news term occurrence



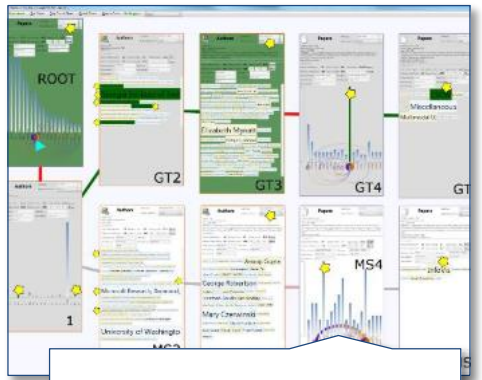
computer network traffic



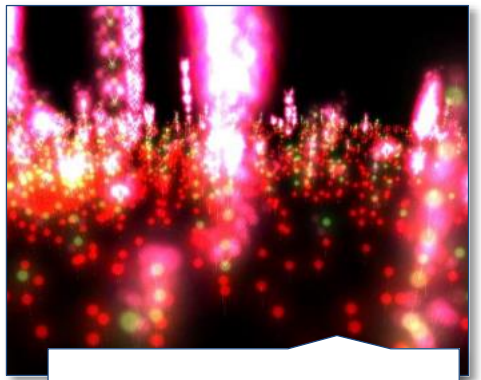
epidemiology/dynamic networks



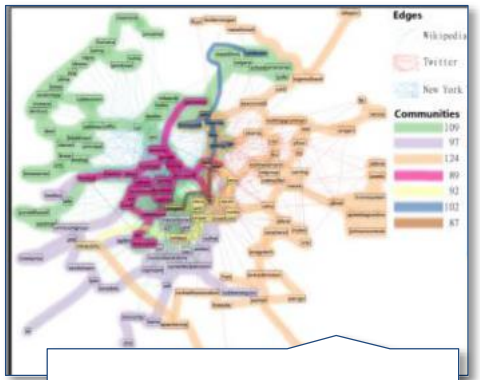
layout readability



exploration provenance



network type overviews



group/set visualization

The Data Problem

Why Visualization?

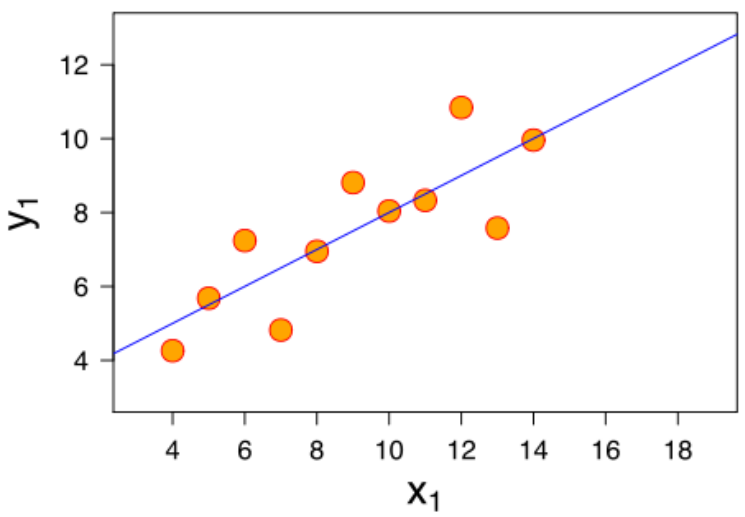
Anscombe's quartet – Table

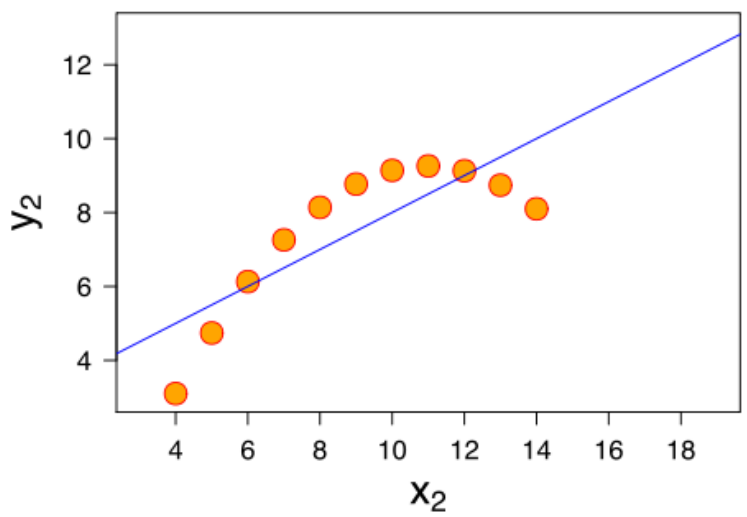
I		II		III		IV	
x	y	x	y	x	y	x	y
10.00	8.04	10.00	9.14	10.00	7.46	8.00	6.58
8.00	6.95	8.00	8.14	8.00	6.77	8.00	5.76
13.00	7.58	13.00	8.74	13.00	12.74	8.00	7.71
9.00	8.81	9.00	8.77	9.00	7.11	8.00	8.84
11.00	8.33	11.00	9.26	11.00	7.81	8.00	8.47
14.00	9.96	14.00	8.10	14.00	8.84	8.00	7.04
6.00	7.24	6.00	6.13	6.00	6.08	8.00	5.25
4.00	4.26	4.00	3.10	4.00	5.39	19.00	12.50
12.00	10.84	12.00	9.13	12.00	8.15	8.00	5.56
7.00	4.82	7.00	7.26	7.00	6.42	8.00	7.91
5.00	5.68	5.00	4.74	5.00	5.73	8.00	6.89

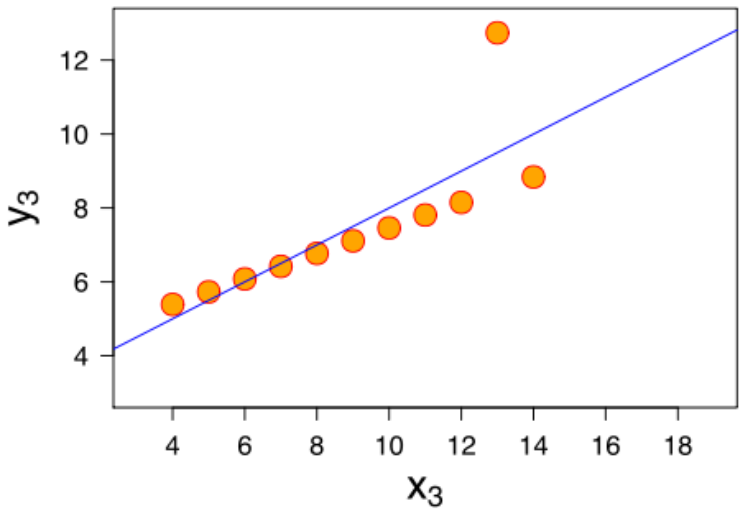
Why Visualization?

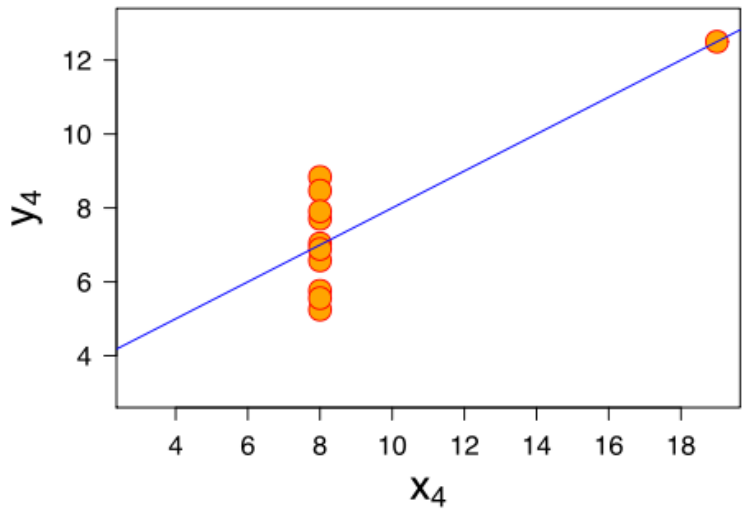
Anscombe's quartet – Statistics & Visualization

Property in Each Case	Value	Equality
Mean of x	9	Exact
Variance of x	11	Exact
Mean of y	7.50	2 decimal places
Variance of y	4.122 or 4.127	3 decimal places
Correlation between x & y	0.816	3 decimal places
Linear regression line	y = 3.00 + 0.500x	2 & 3 decimal places









Why Visualization?

Tukey

No catalogue of techniques can convey a willingness to look for what can be seen, whether or not anticipated. Yet this is at the heart of exploratory data analysis. ... the picture-examining eye is the best finder we have of the wholly unanticipated.

– Tukey, 1980

Design Choices

Charm vs. clarity

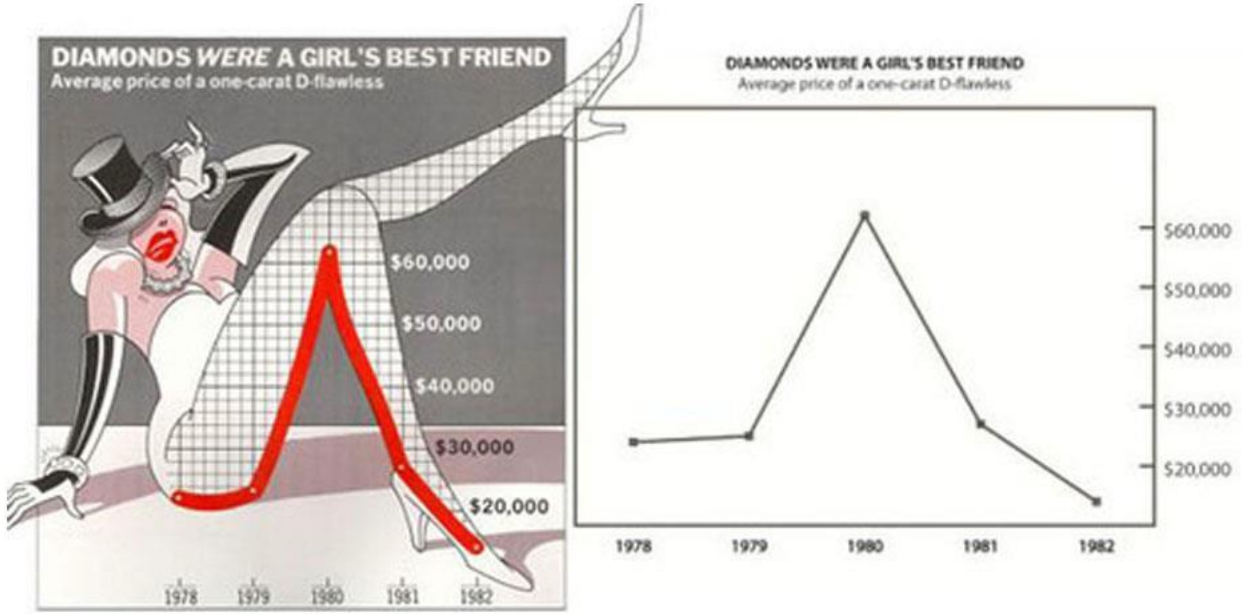
Charm and Direct Appeal

Clarity and Power

Design Choices

Charm vs. clarity

Nigel Holmes

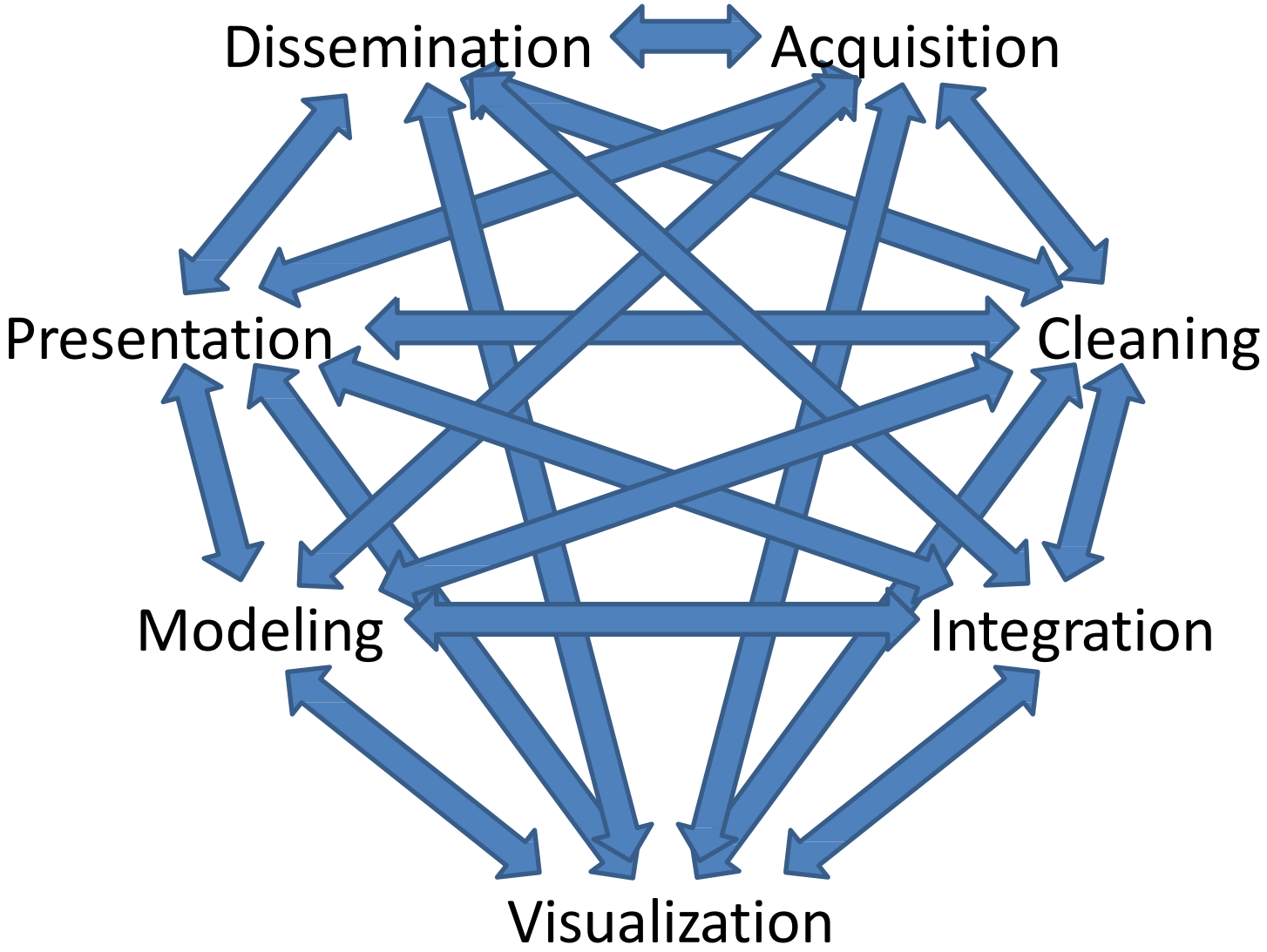


Edward Tufte

Data Analysis Process

Kandel et al., 2012	Heer, 2013 (unpublished)
Discovery	Acquisition
Wrangling	Cleaning Integration
Profiling	Visualization
Modeling	Modeling
Reporting	Presentation Dissemination

Data Analysis Process



Node-Link Network Visualizations

Node-Link Graph Visualization

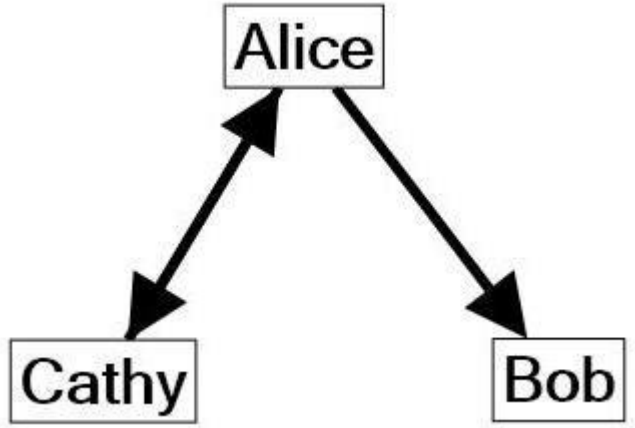
General

Graph ≈ Network

Node ≈ Vertex ≈ Entity

Edge ≈ Link ≈ Relationship ≈ Tie

Node 1	Node 2
Alice	Bob
Alice	Cathy
Cathy	Alice



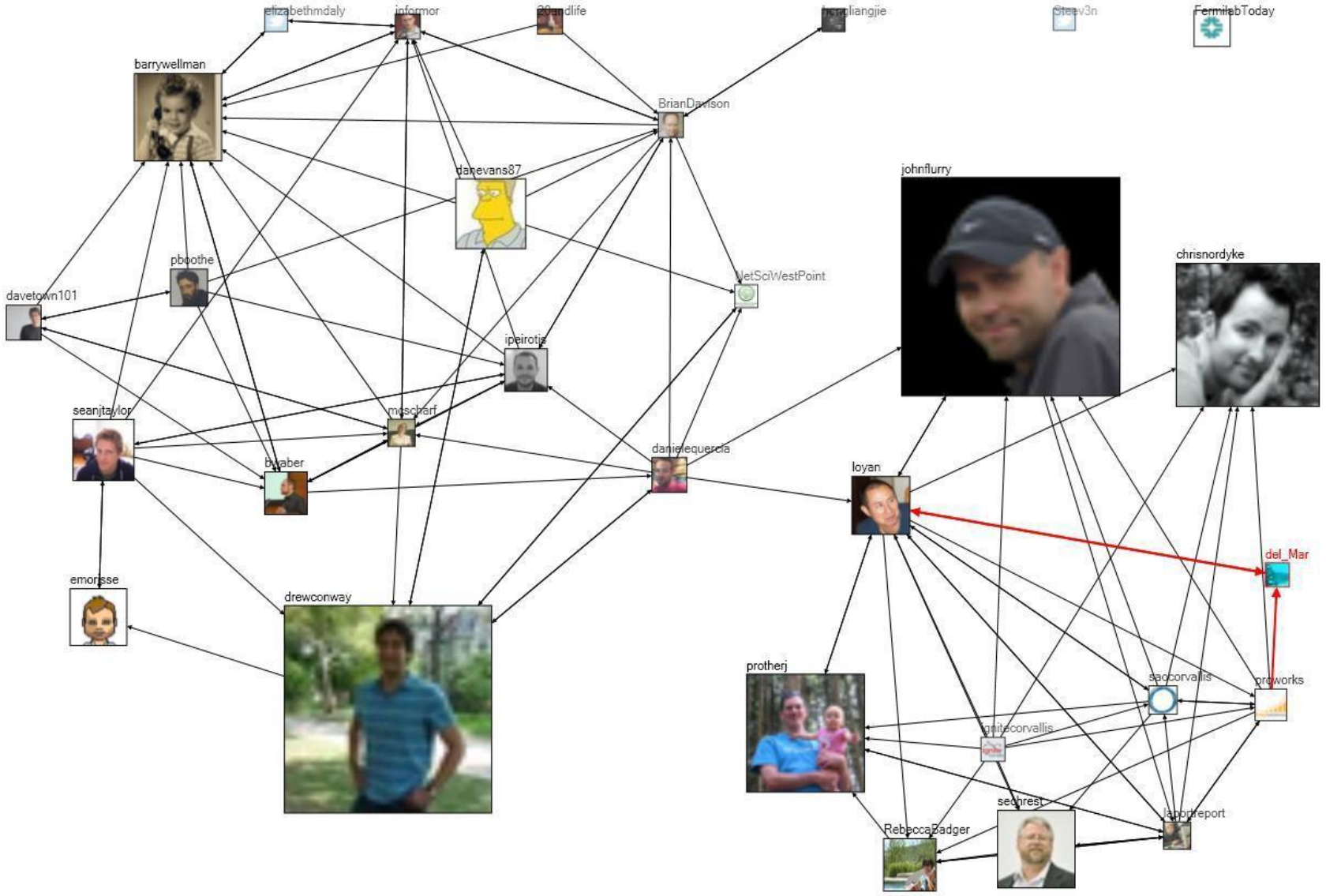
Node-ation

Tweets of the #Win09 Workshop

#	User 1	User 2	#	User 1	User 2
1	20andlife	barrywellman	15	danevans87	informor
2	20andlife	BrianDavidson	16	danevans87	NetSciWestPoint
3	barrywellman	elizabethmdaly	17	danielequercia	BrianDavidson
4	barrywellman	informor	18	danielequercia	drewconway
5	BrianDavidson	hcraygliangjie	19	danielequercia	ipeirotis
6	BrianDavidson	informor	20	danielequercia	johnflurry
7	BrianDavidson	NetSciWestPoint	21	danielequercia	loyan
8	byaber	barrywellman	22	danielequercia	loyan
9	byaber	danielequercia	23	danielequercia	mcscharf
10	byaber	mcscharf	24	danielequercia	NetSciWestPoint
11	chrisnordyke	RebeccaBadger	…	…	…
12	danevans87	barrywellman	106	sechrest	Japportreport
13	danevans87	BrianDavidson	107	sechrest	loyan
14	danevans87	drewconway	108	sechrest	RebeccaBadger

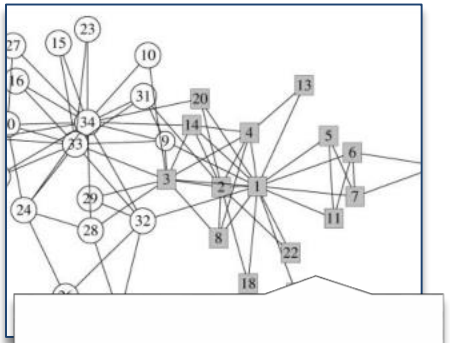
Node-Link Network Visualization

Tweets of the #Win09 Workshop

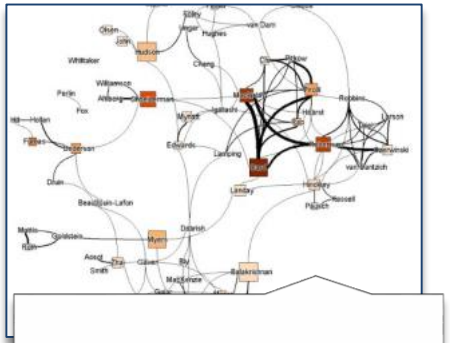


Node-Link Network Visualization

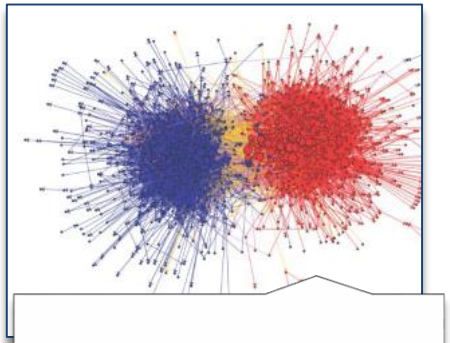
Who uses network analysis?



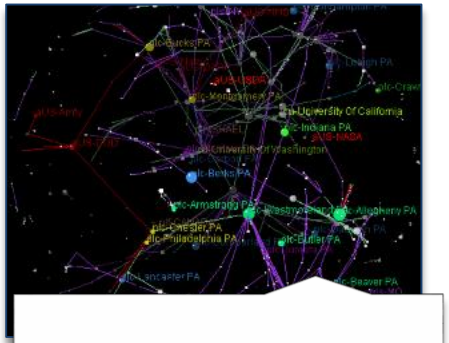
Sociology



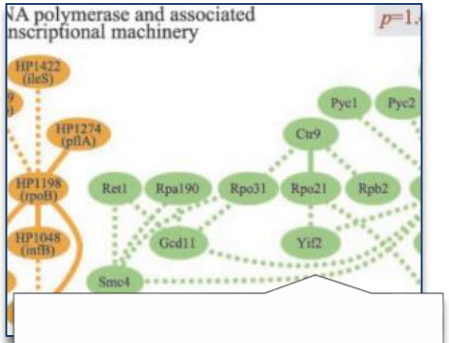
Scientometrics



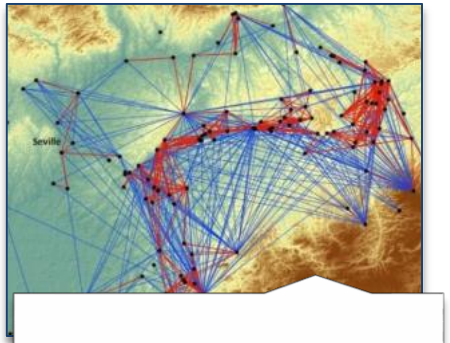
Politics



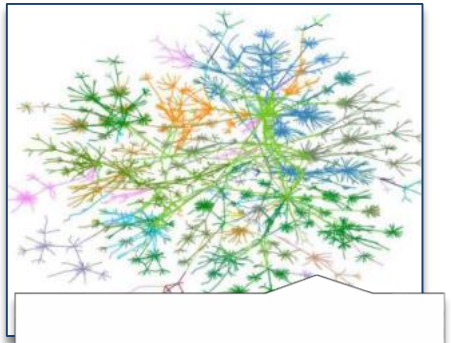
Urban Planning



Biology



Archaeology



WWW

Network Visualization in IBM

Services exposing network data (many more coming)

Watson

Build cognitive apps that help enhance, scale, and accelerate human expertise

AlchemyAPI
IBM

Concept Expansion
IBM *BETA*

Concept Insights
IBM *BETA*

Language Identification
IBM *BETA*

Machine Translation
IBM *BETA*

Personality Insights
IBM

Question and Answer
IBM *BETA*

Relationship Extraction
IBM *BETA*

Speech To Text
IBM *BETA*

Text to Speech
IBM *BETA*

Tradeoff Analytics
IBM

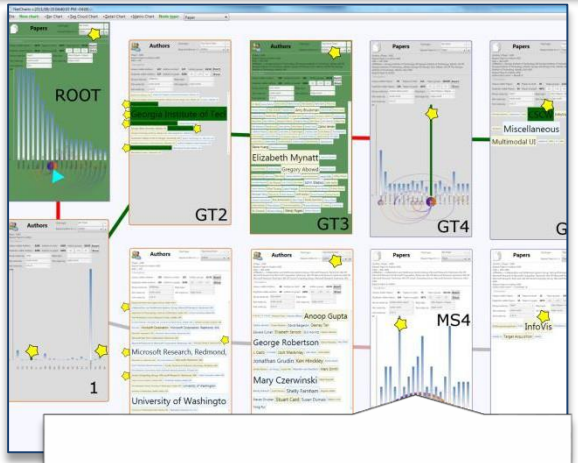
Visual Recognition
IBM *BETA*

Cognitive Commerce™
Third Party

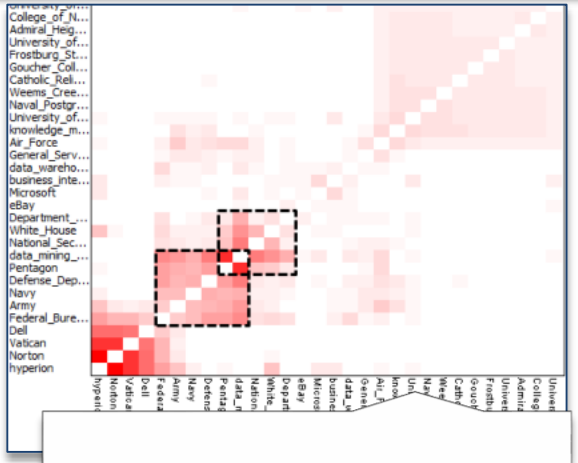
Cognitive Graph
Third Party

Cognitive Insights™
Third Party

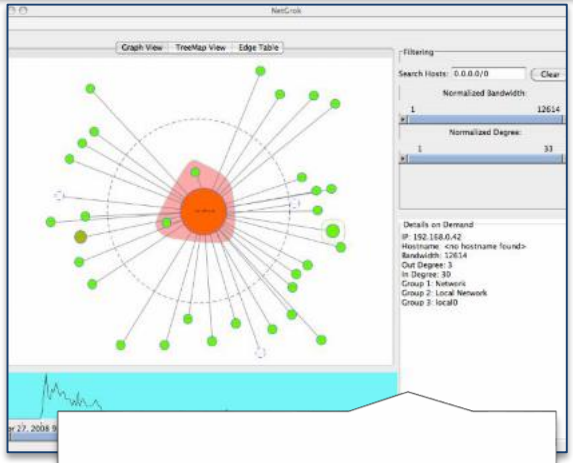
Alternate Network Visualizations



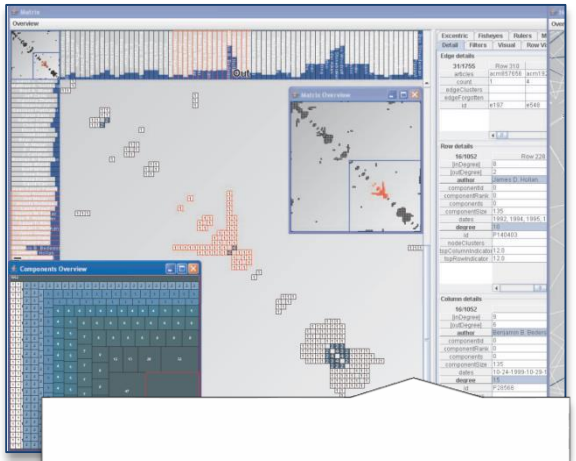
Dunne et al., 2012



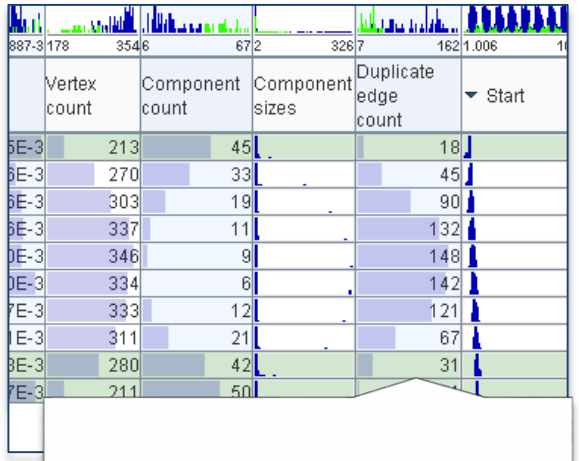
Gove et al., 2011



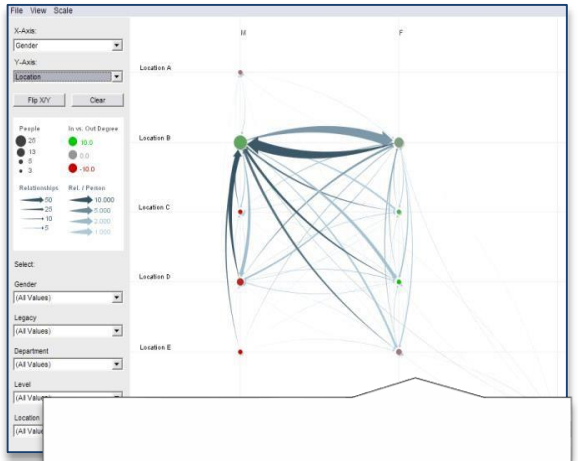
Blue et al., 2008



Henry & Fekete, 2006



Freire et al., 2010



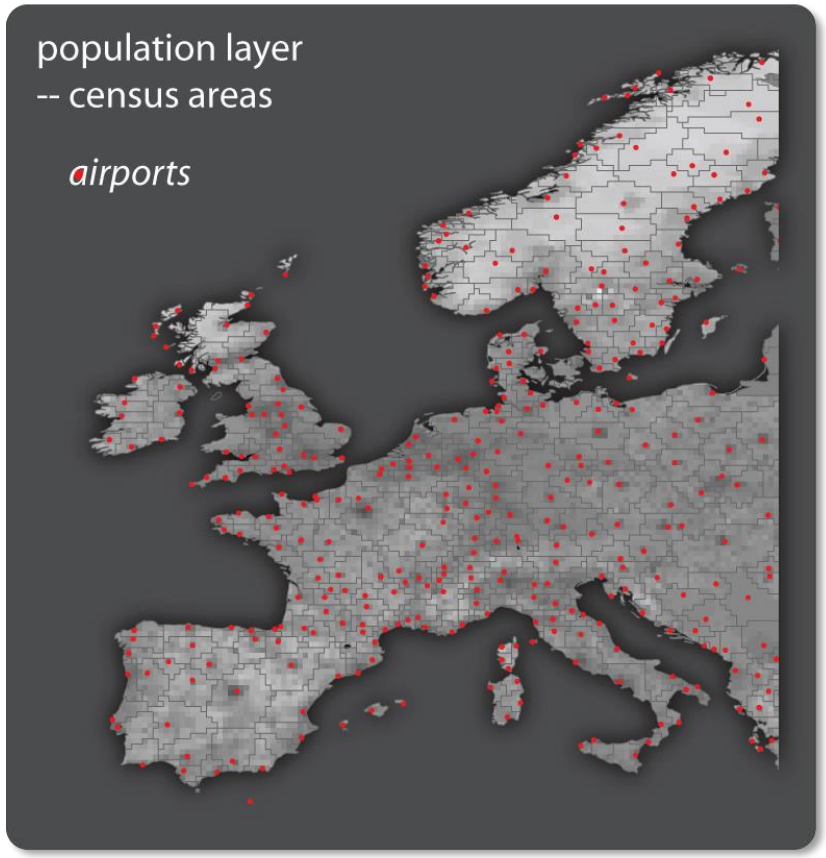
Wattenberg, 2006

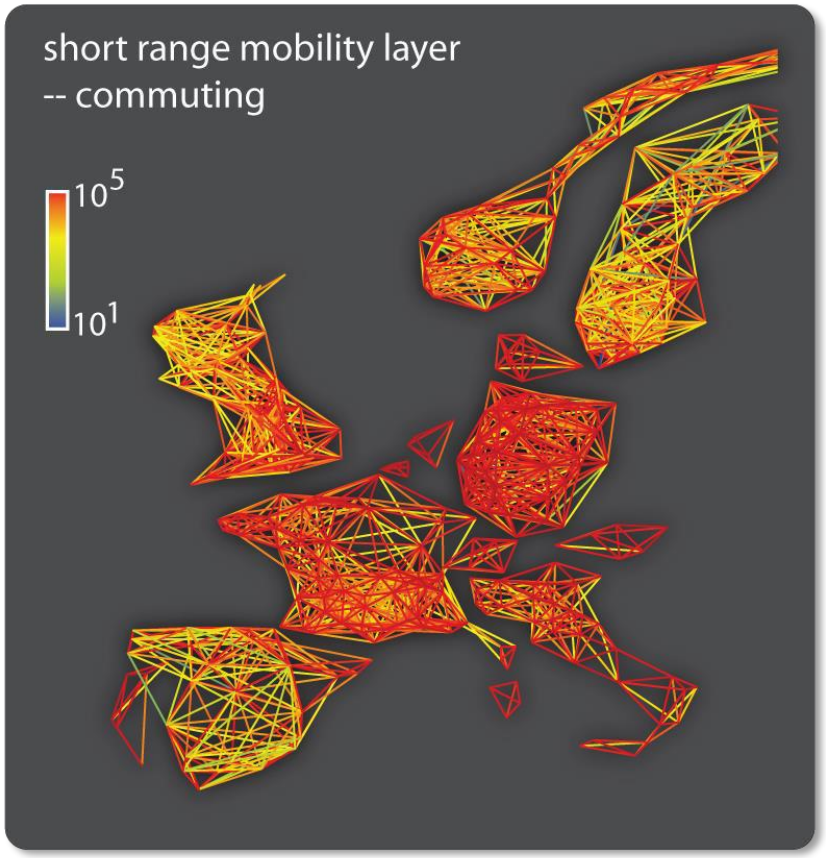
VoroGraph

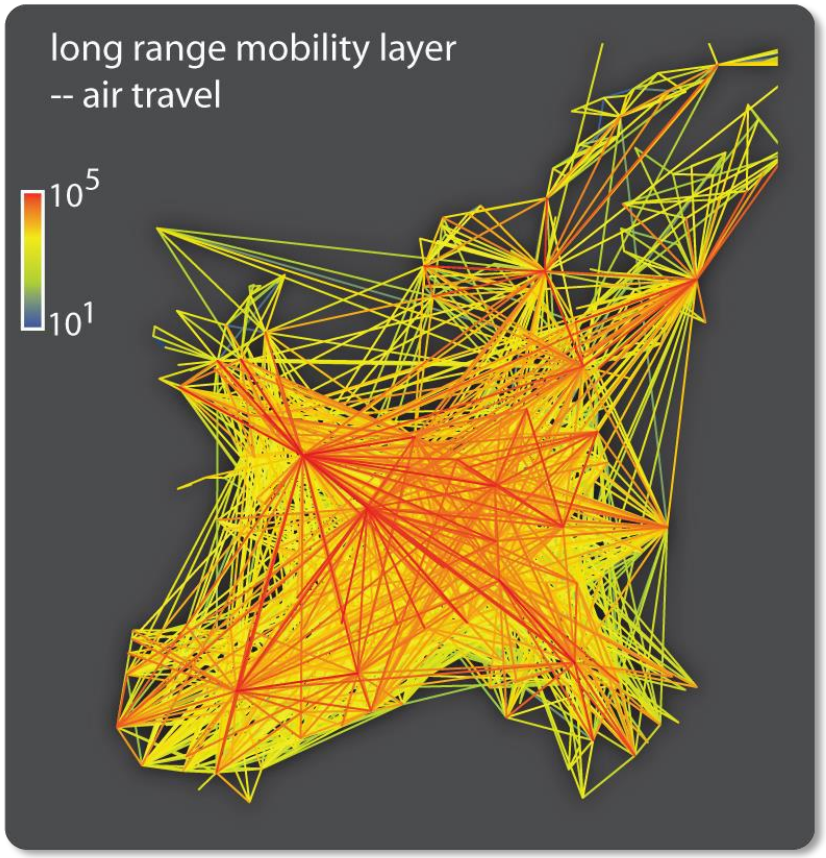

Introduction
Node-Link
VoroGraph
GraphTrail
StoryFacets
Discussion

GLEAM Epidemic Model

Population basins, local commuting, and global flights

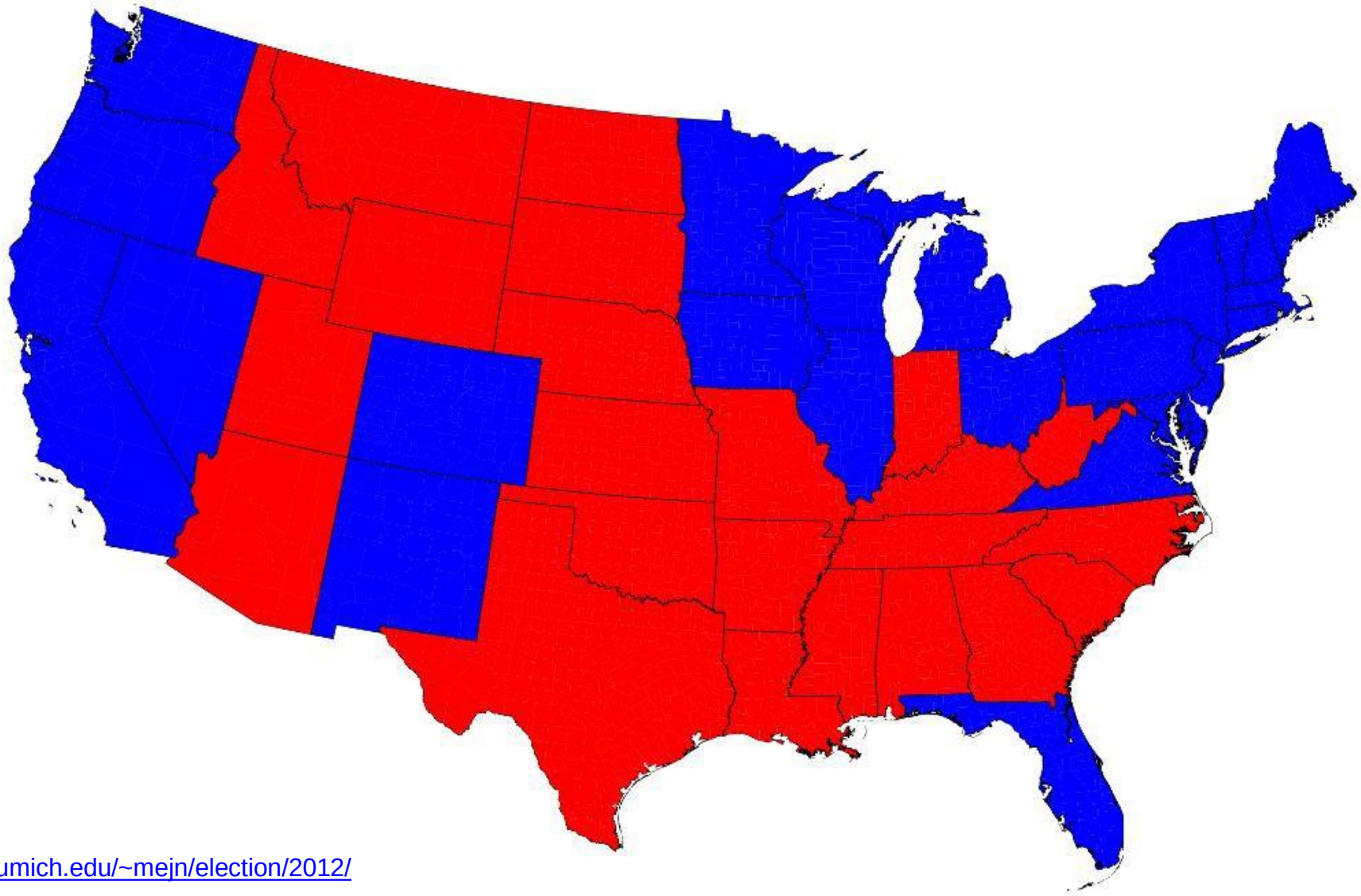






Choropleth Maps

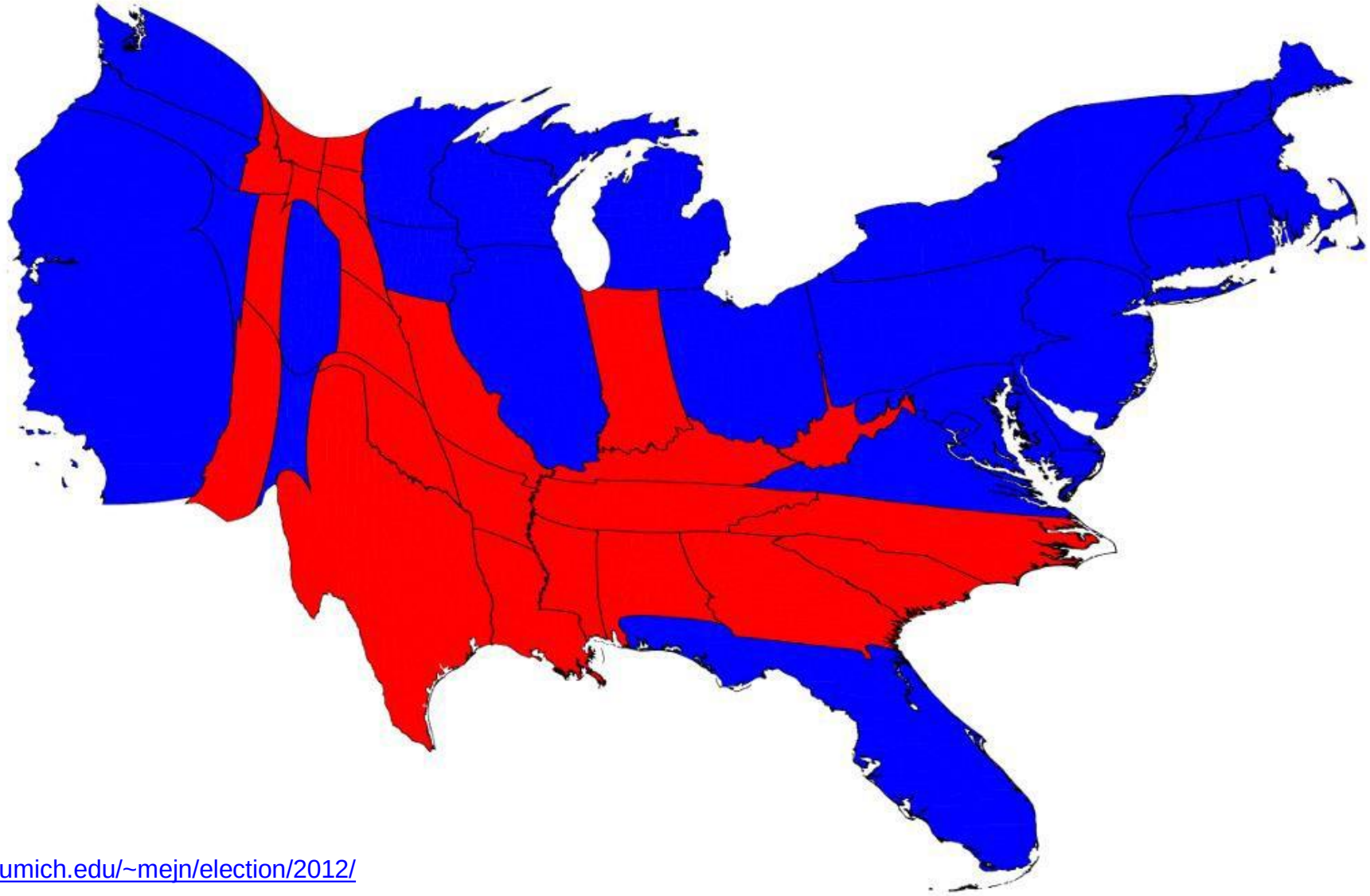
Gastner and Newman 2012 Election (State)



http://www-personal.umich.edu/~mejn/election/2012/

Area Morphing

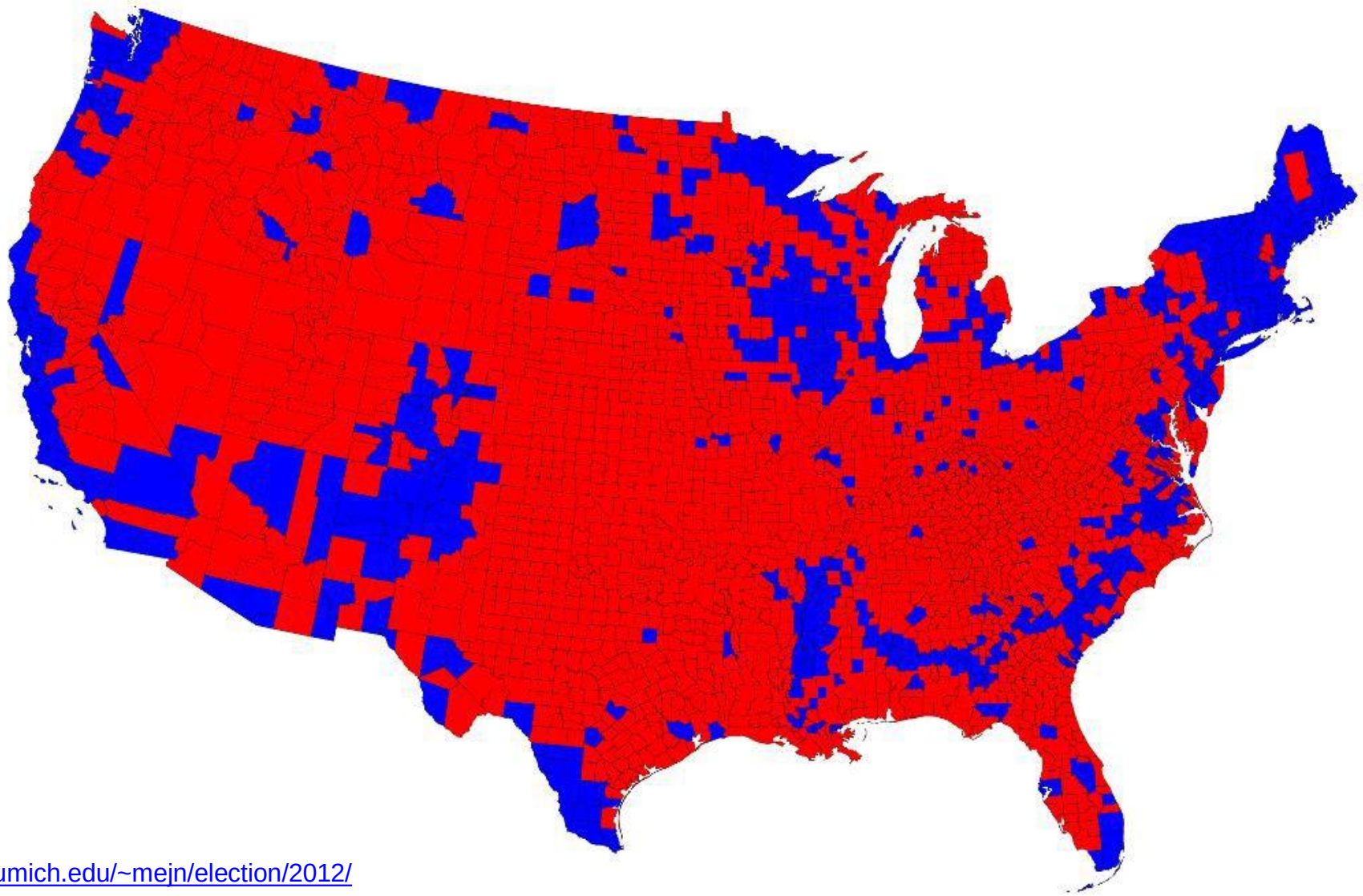
Gastner and Newman 2012 Election (State)



http://www-personal.umich.edu/~mejn/election/2012/

Area Morphing

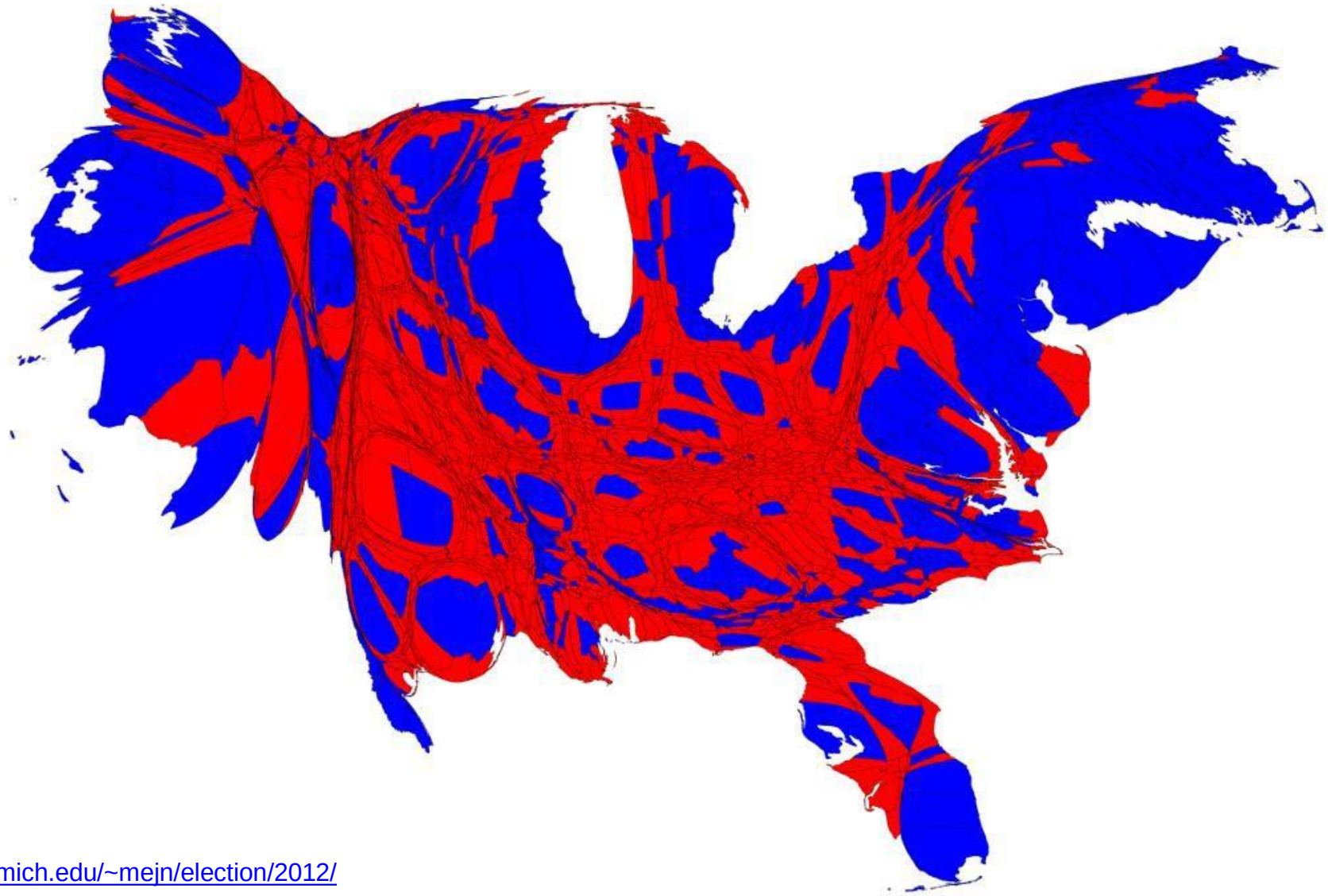
Gastner and Newman 2012 Election (County)



http://www-personal.umich.edu/~mejn/election/2012/

Contiguous Density Equalizing Cartograms

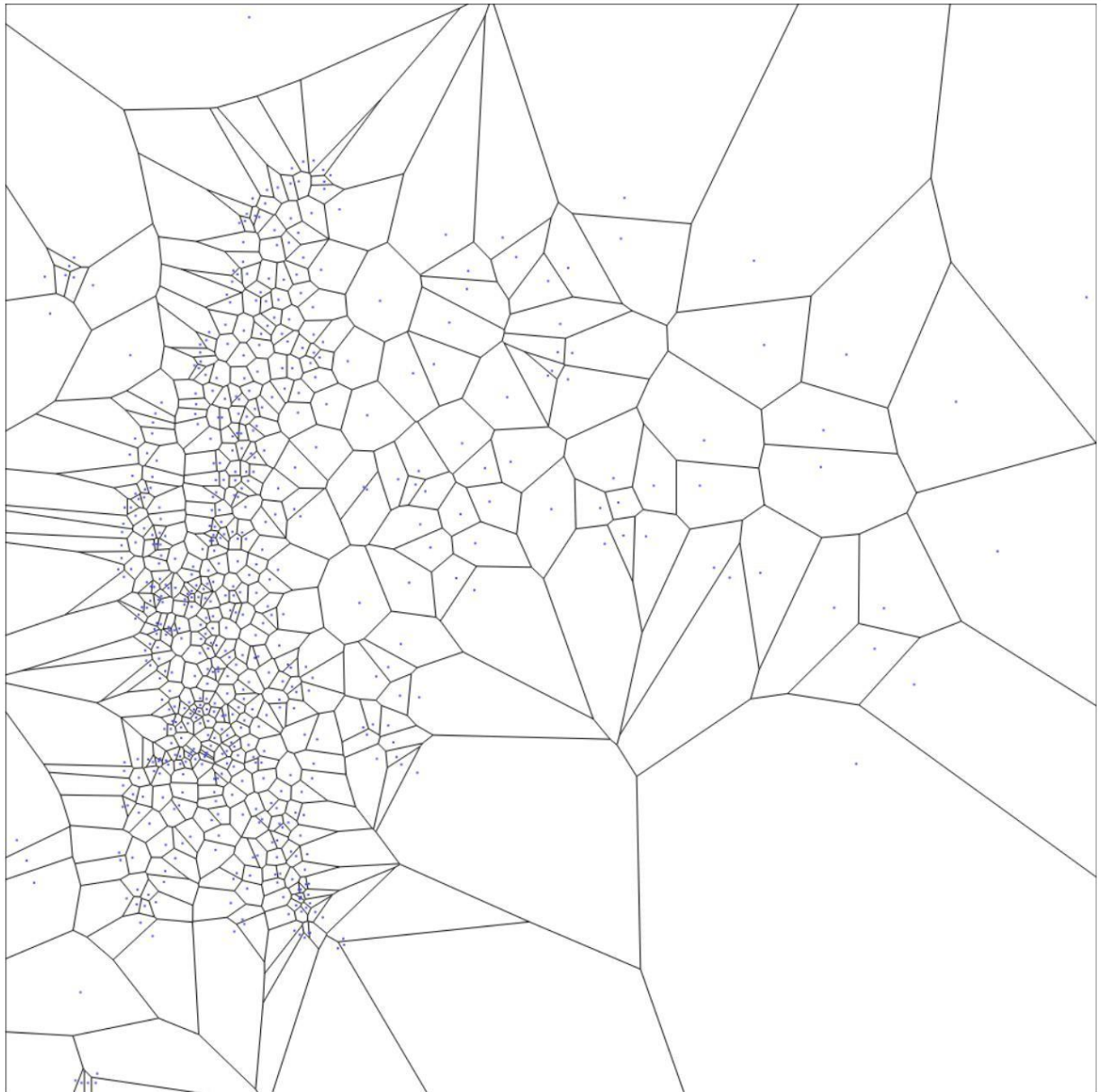
Gastner and Newman 2012 Election (County)



http://www-personal.umich.edu/~mejn/election/2012/

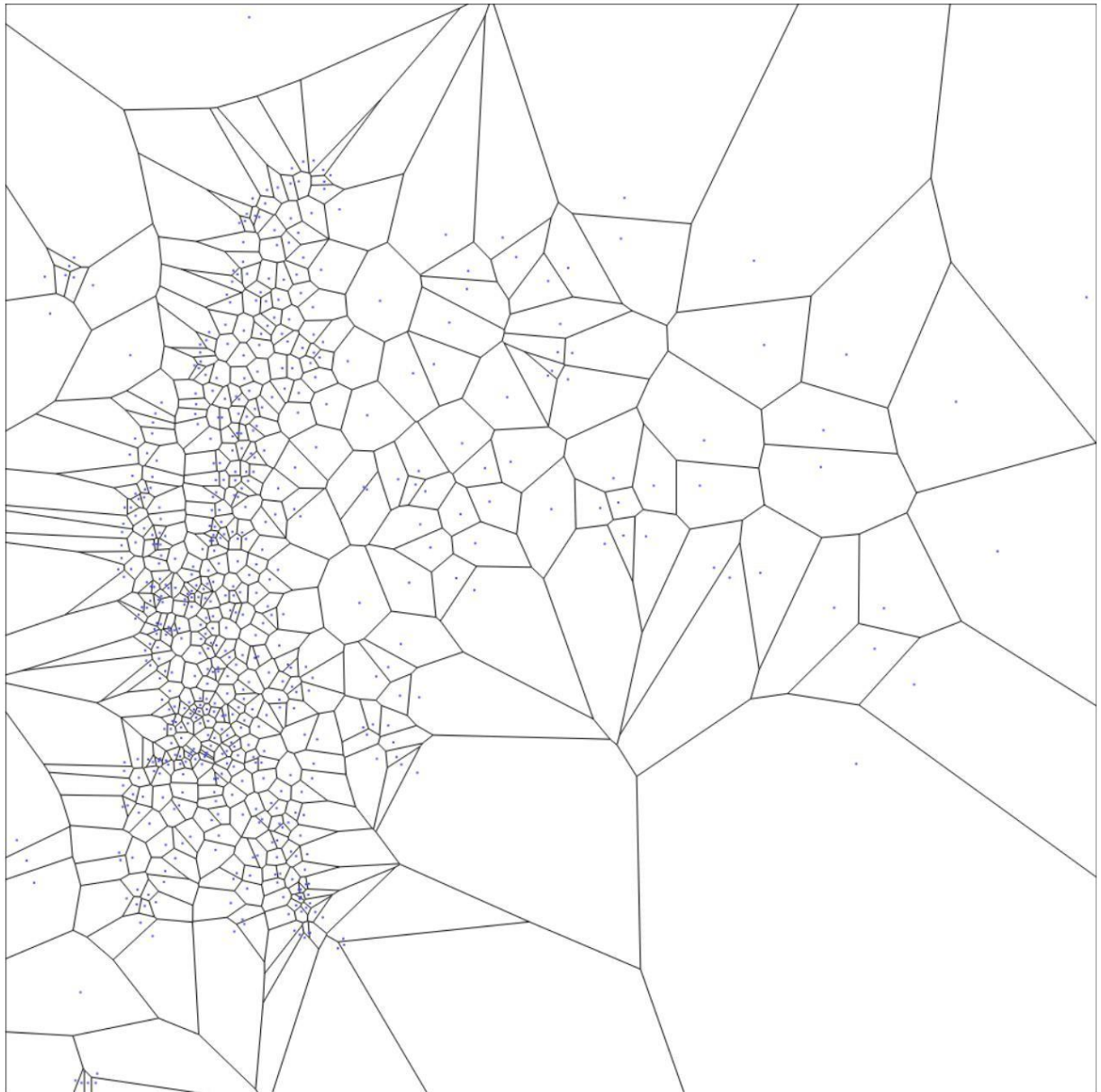
Voronoi Tessellation

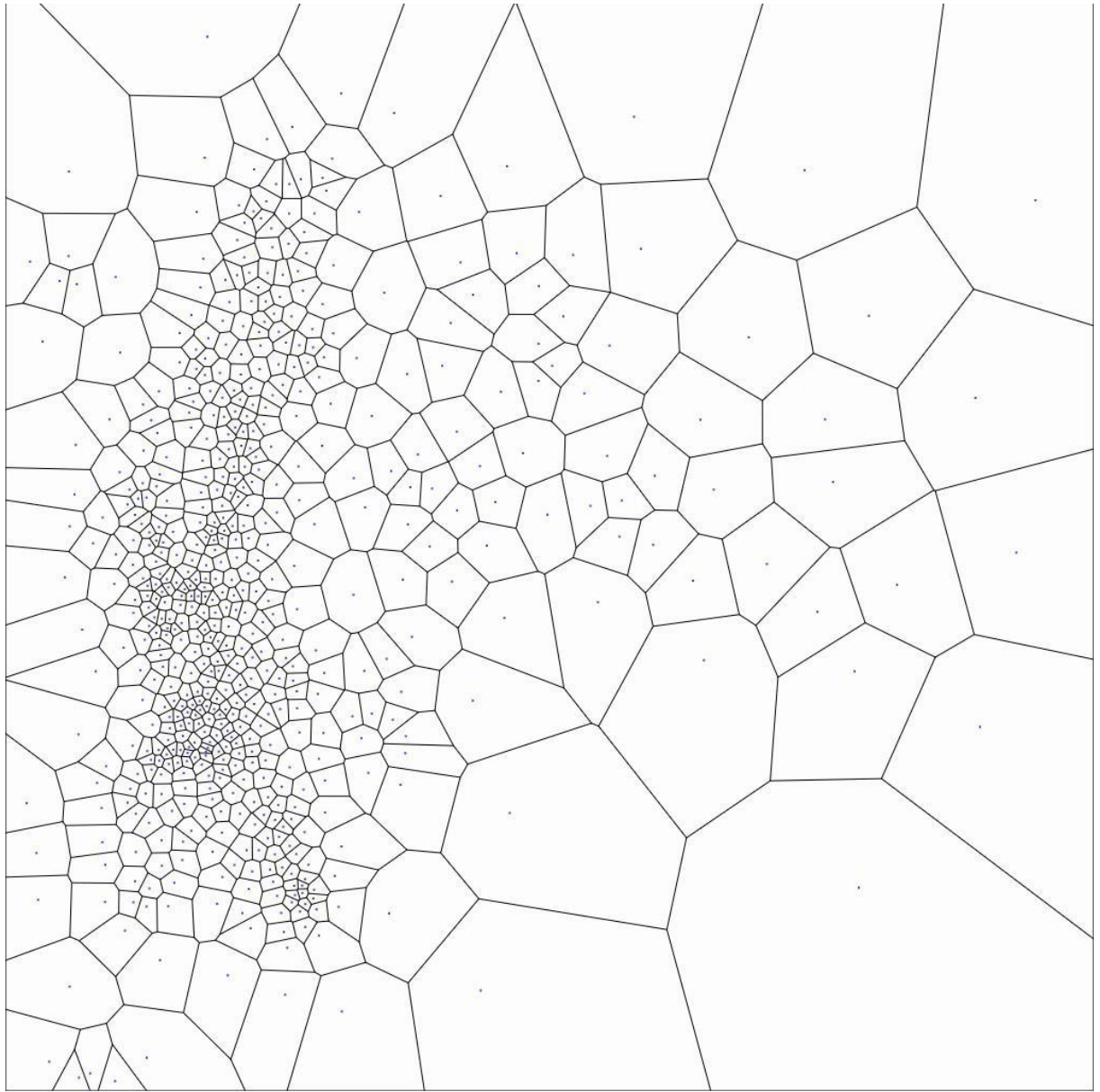
Western Europe



Centroidal Voronoi Tessellation – Animated!

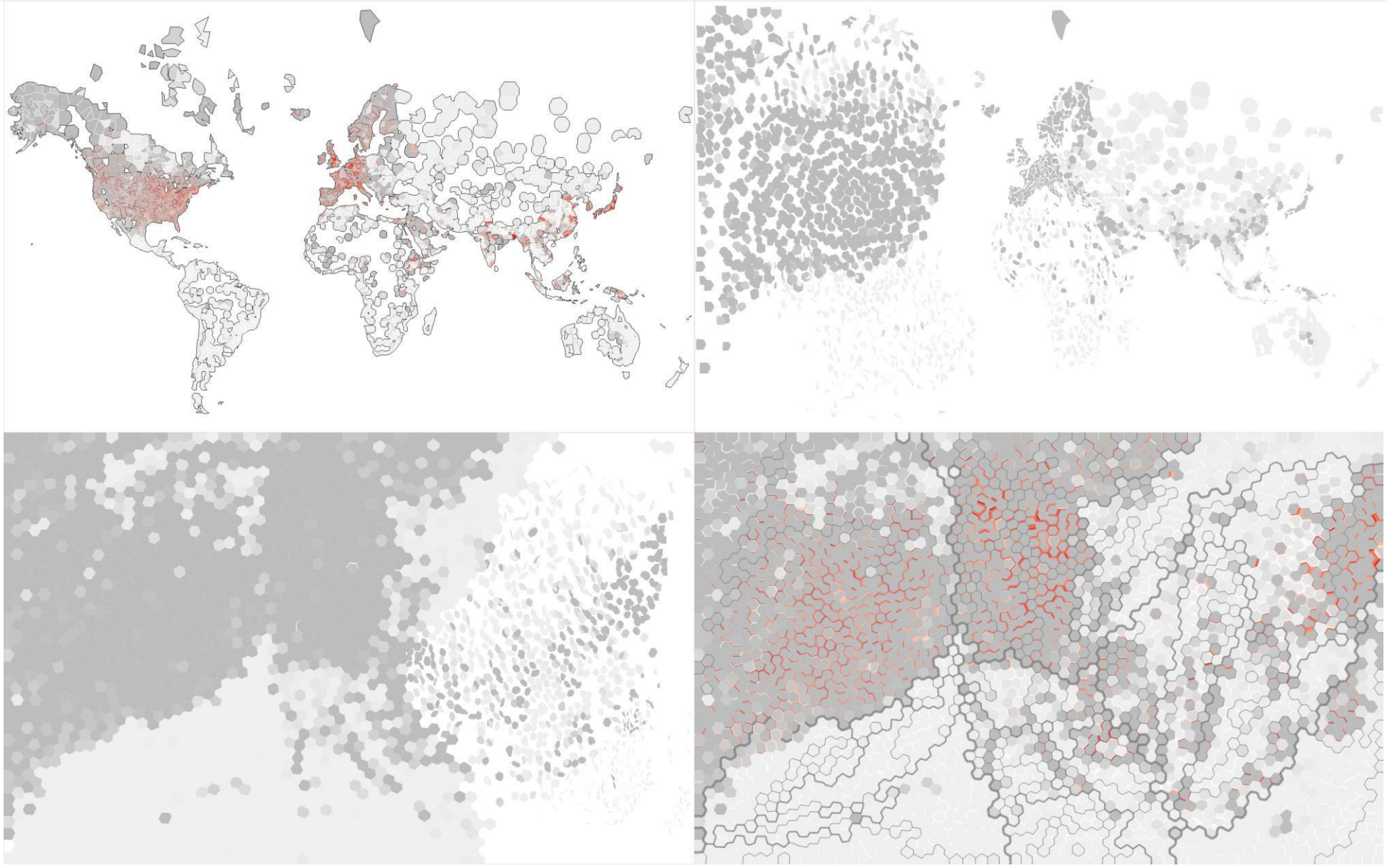
Western Europe





VoroGraph

Animated Transitions

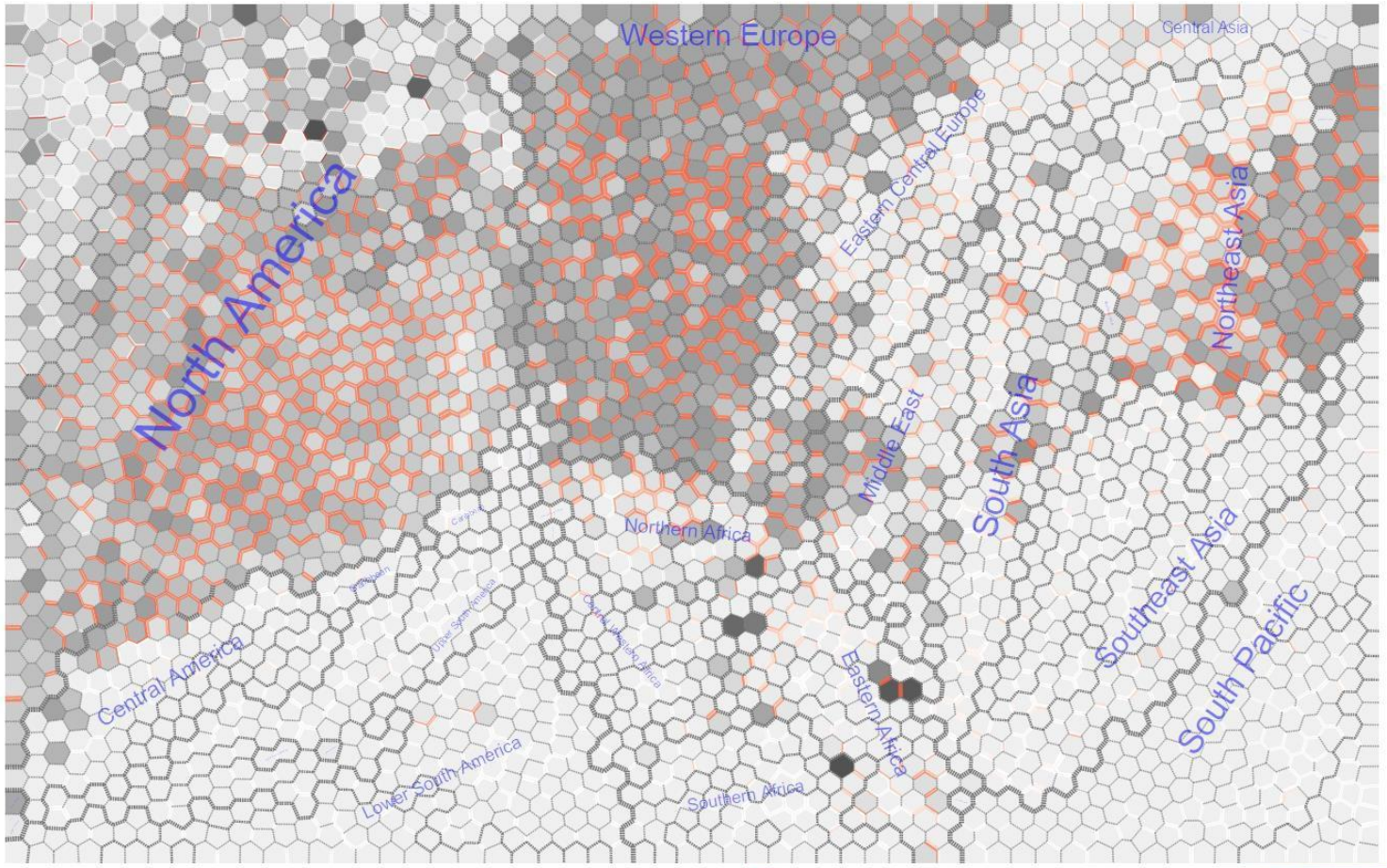


VoroGraph

Interface Demo

New York Influenza Scenario: Consider a new strain of influenza starting in New York City in mid February. In a dense population the disease could quickly reach pandemic proportions. Indeed, the millions of commuters and visitors could carry the virus home with them. As with H1N1, the authorities would have to face both the local and global spreading of the disease.

Map Transform
Voronoi Transform
FD Transform
Open Controls

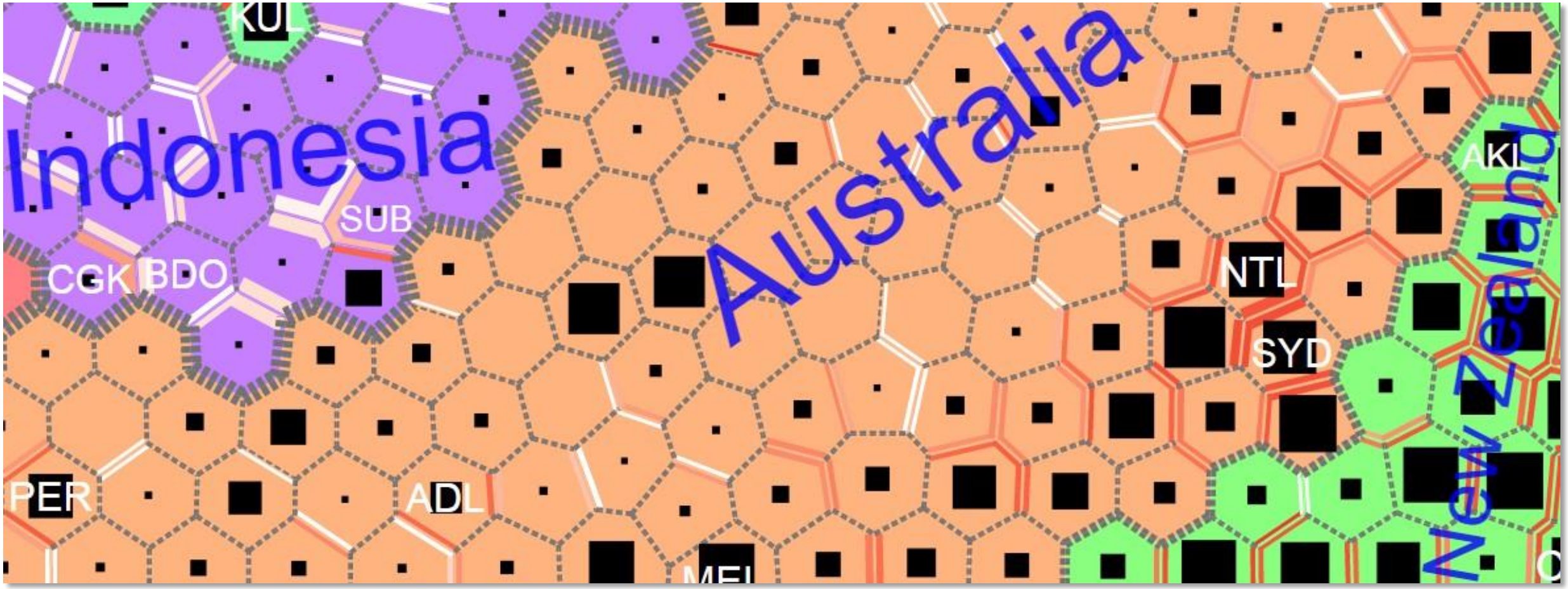


VoroGraph

Contiguous Edge Coding

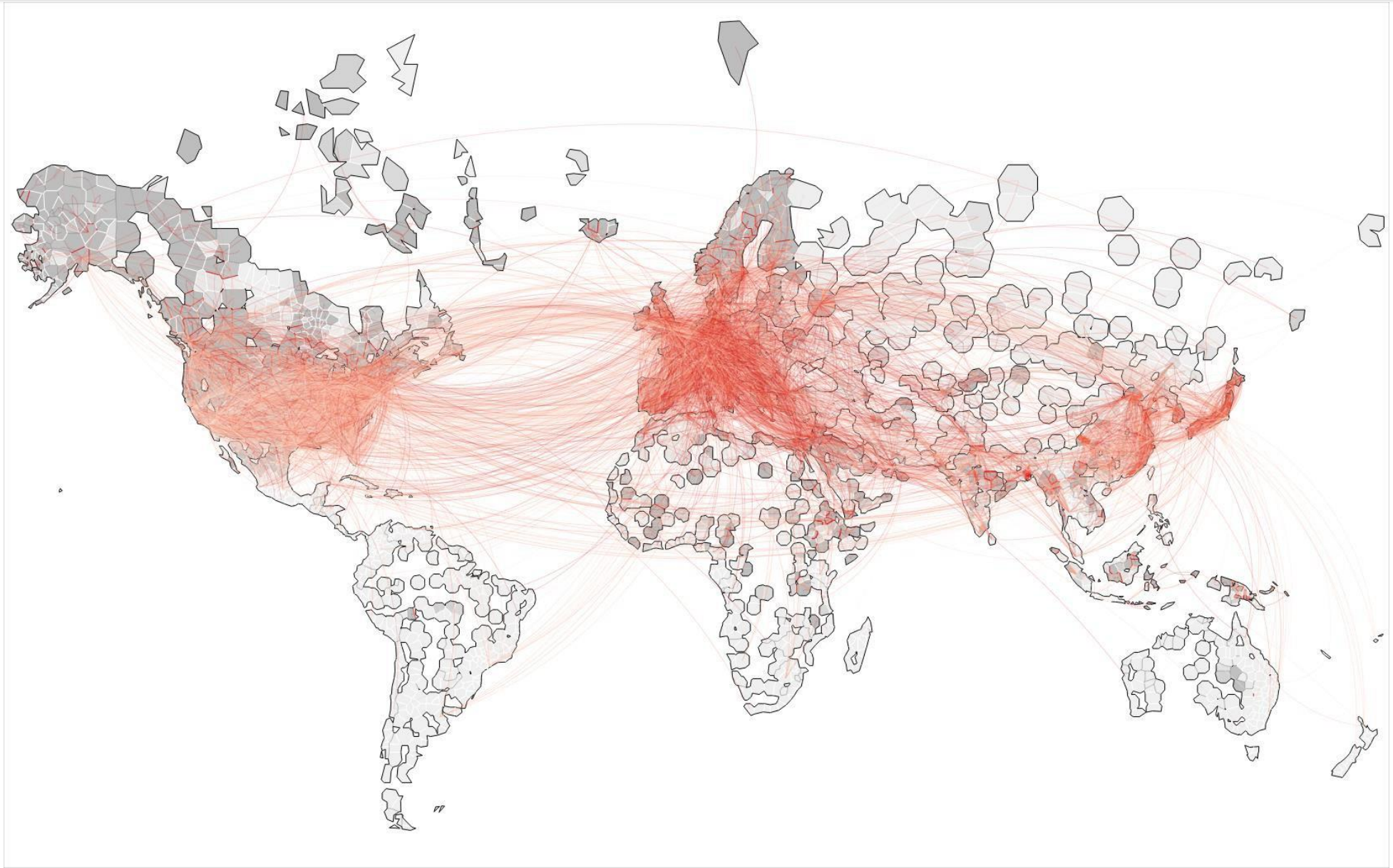
VoroGraph

Data Squares



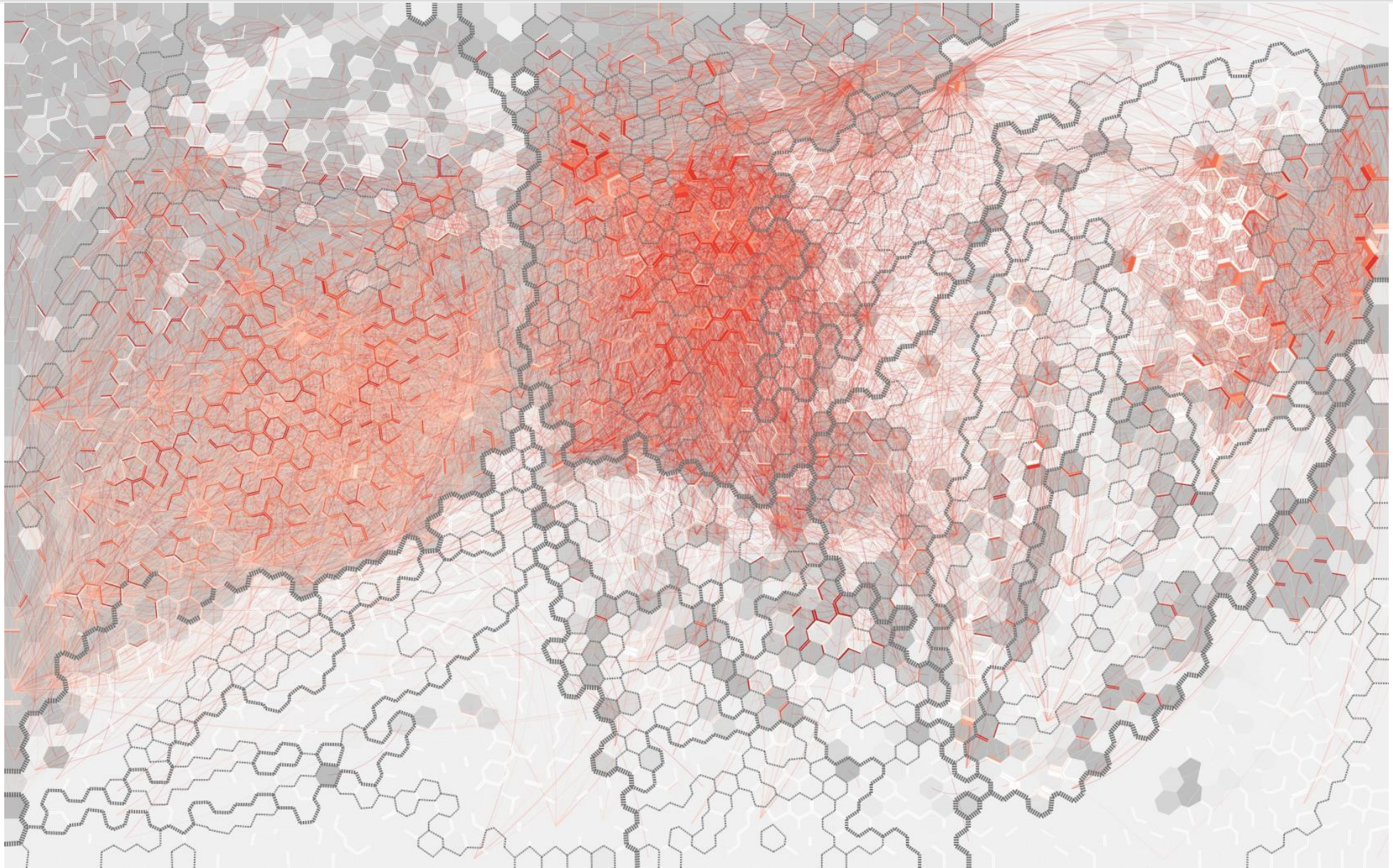
VoroGraph

Non-Contiguous Edge Coding



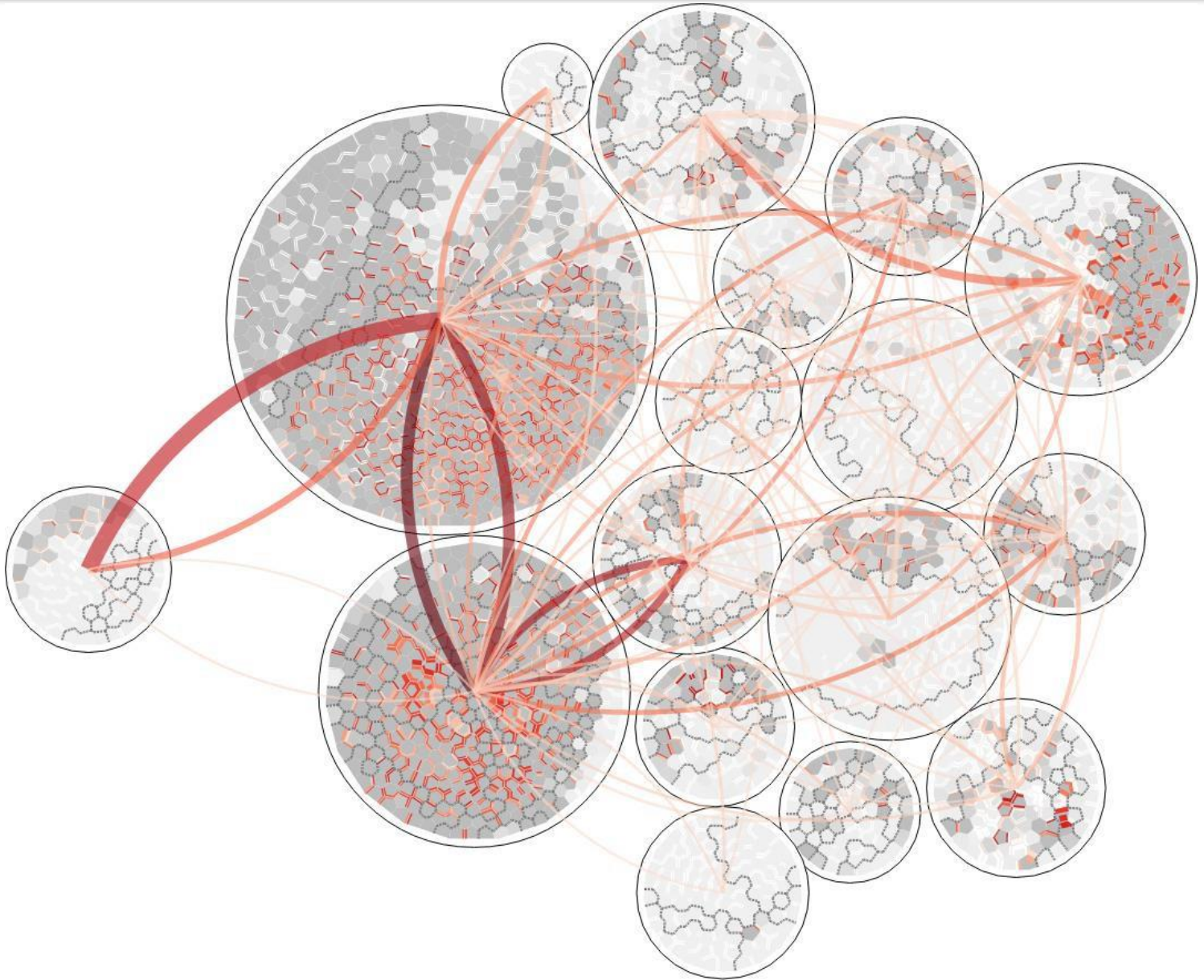
VoroGraph

Non-Contiguous Edge Coding



VoroGraph

Force-Directed Group-in-a-Box



VoroGraph

Discussion

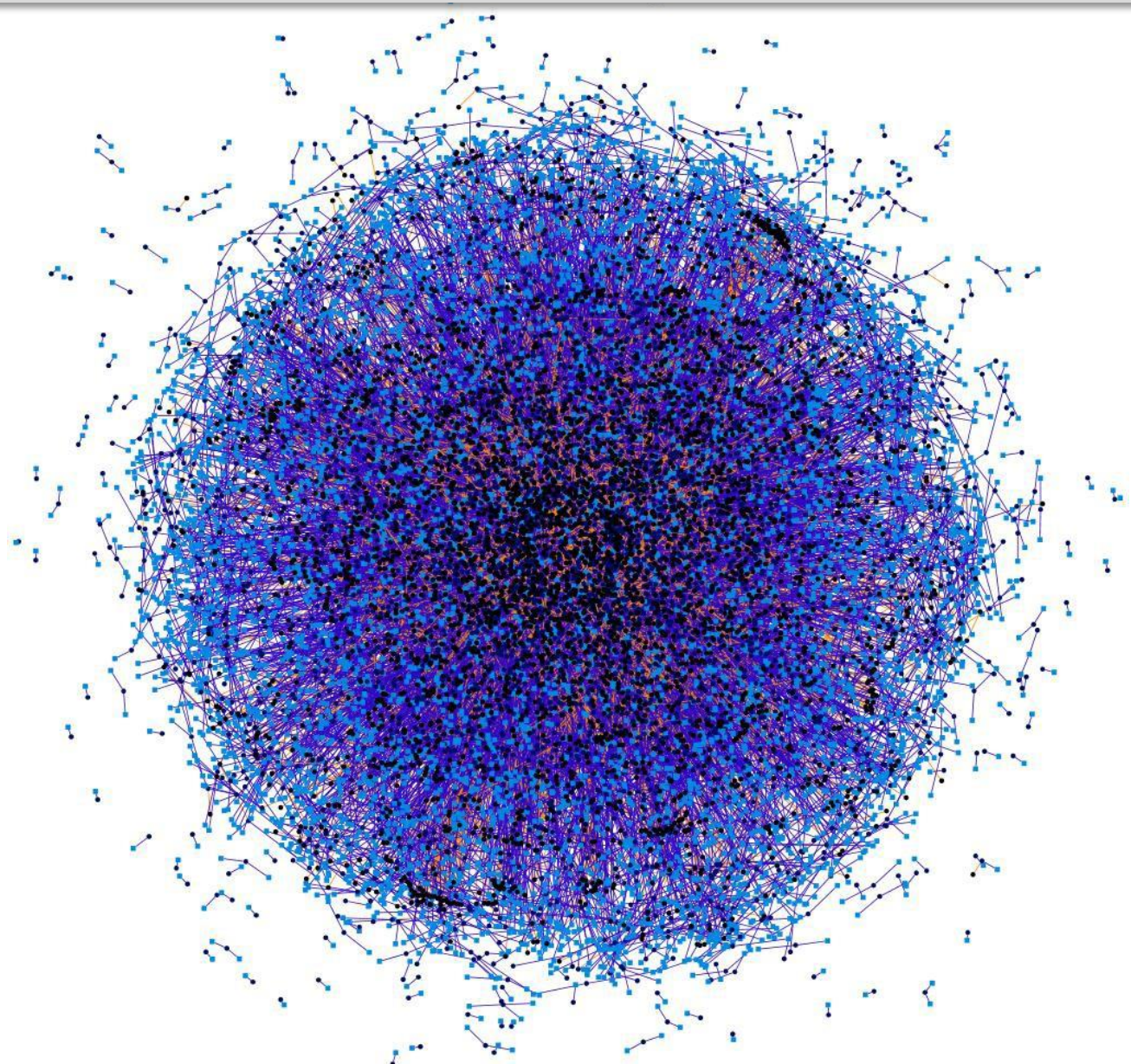
- Equal-population hexagons discretize the space for **countability**
- Easier **attribute comparison** with color/size coding
- Hexagons make clear it is an **artificial representation**
- Enforces a degree of **generalization**
- **Contiguous relationship** display

Dunne C, Muller M, Perra N, and Martino M. (2015) "*VoroGraph: Visualization Tools for Epidemic Analysis",*
In CHI '15 Interactivity. DOI:10.1145/2702613.2725459

GraphTrail


Introduction
Node-Link
VoroGraph
GraphTrail
StoryFacets
Discussion

Node-Link Visualization is Hard!



Networks can be

- Large & complex
- Multivariate
- Heterogeneous

Analysis can take

- Many sessions
- Many users

GraphTrail

Overview

- Aggregation
- Drag-and-drop interactions
- Integrated exploration history

Papers

Chart type: Bar Chart

Expand Paper to (+J

<Description>

Unique visible Papers: 4073 Papers in chart: 4073 Paper groups: 19/ 19

Duplicate visible Papers: 5514 Papers in graph: 4073

Export

Group

Group nodes by: topic*

Edge type: id_es (Paper =>Paper)

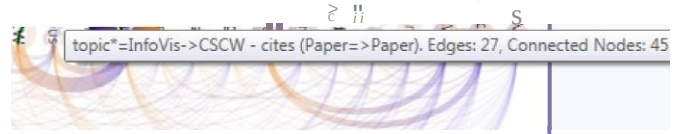
Size nodes by: node count

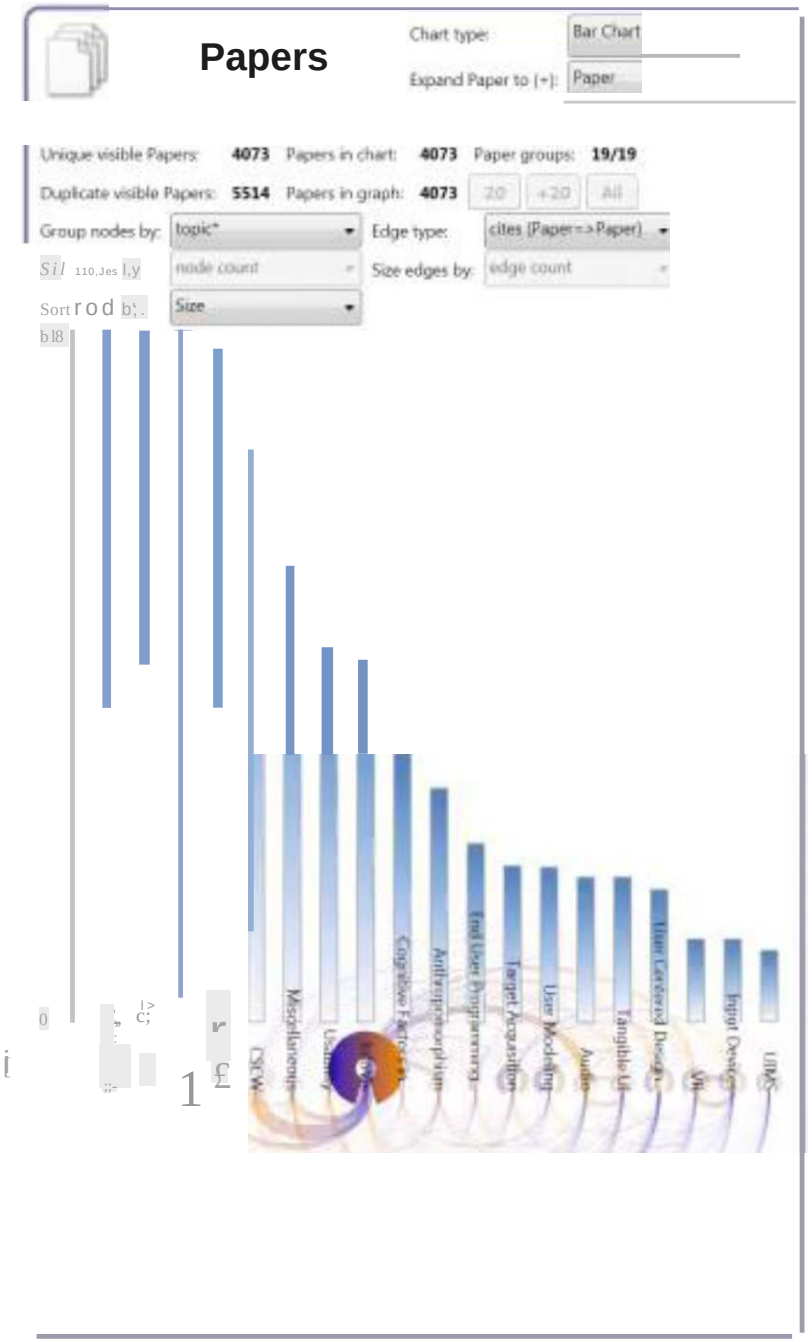
Size edges by: edge-count

Sort nodes by: Size

618

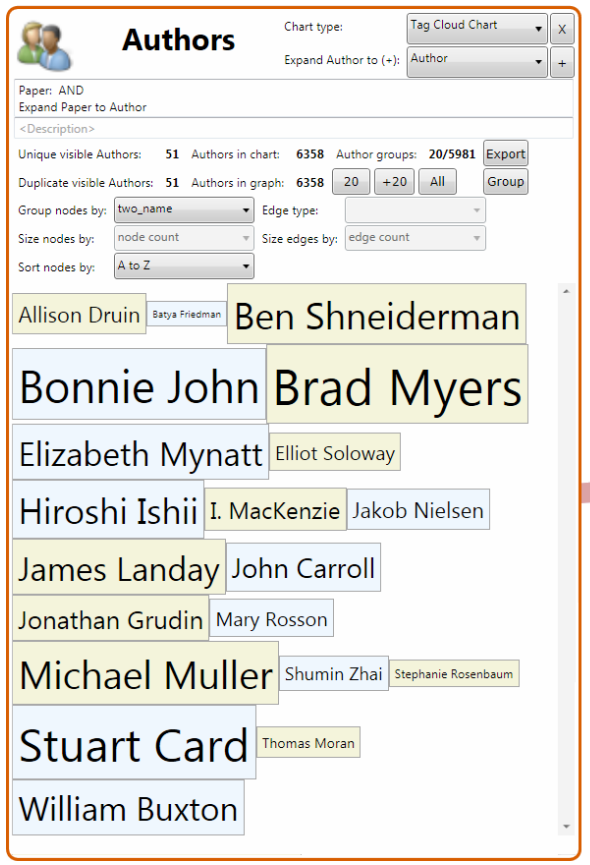
0

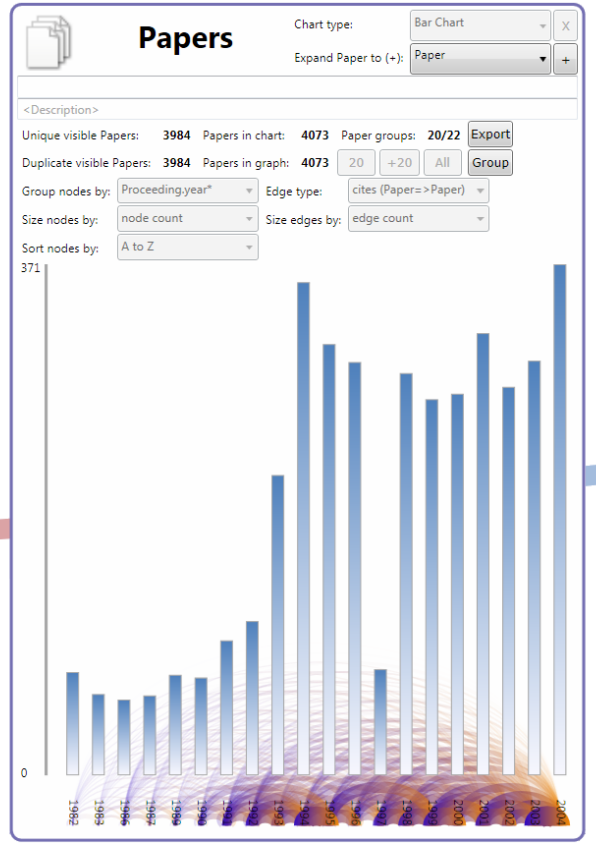


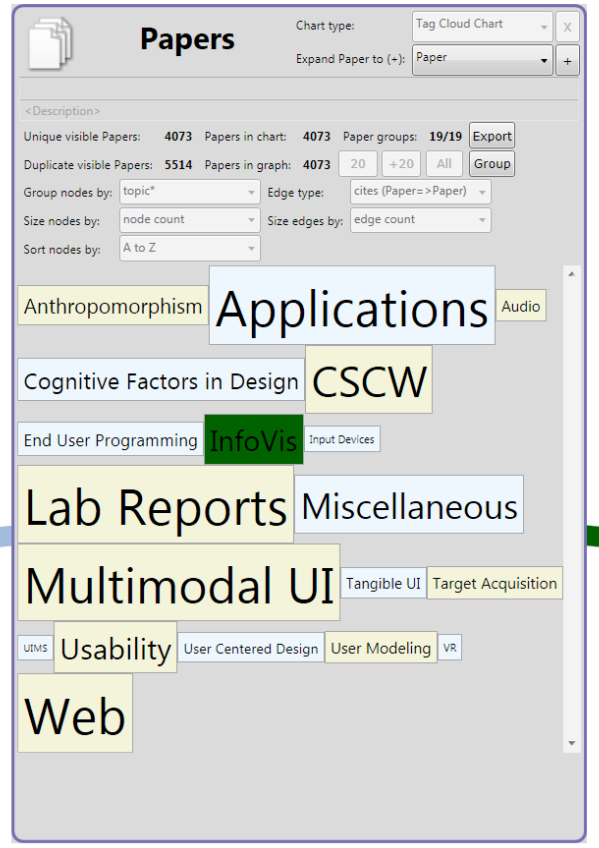

Papers
Chart type:
Bar Chart
Expand Paper to (+):
Paper
Unique visible Papers: 4073
Papers in chart: 4073
Paper groups: 19/19
Duplicate visible Papers: 5514
Papers in graph: 4073
20
+20
All
Group nodes by:
topic*
Edge type:
cites (Paper->Paper)
node count
Size edges by:
edge count
Size
0
Miscellaneous
Cognitive Factors
Target Acquisition
User Modeling
Tangible UI
Audio
Input Devices
UIMS

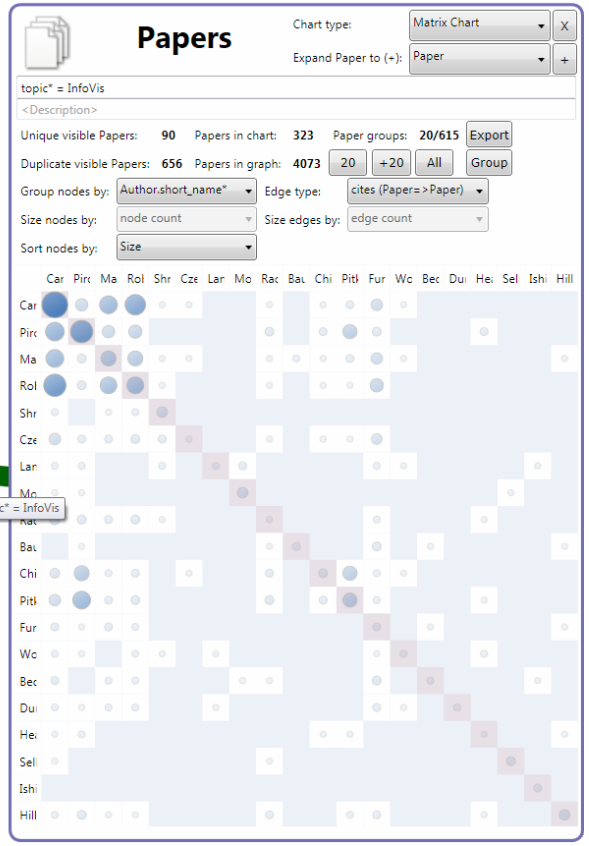
GraphTrail

Aggregating charts



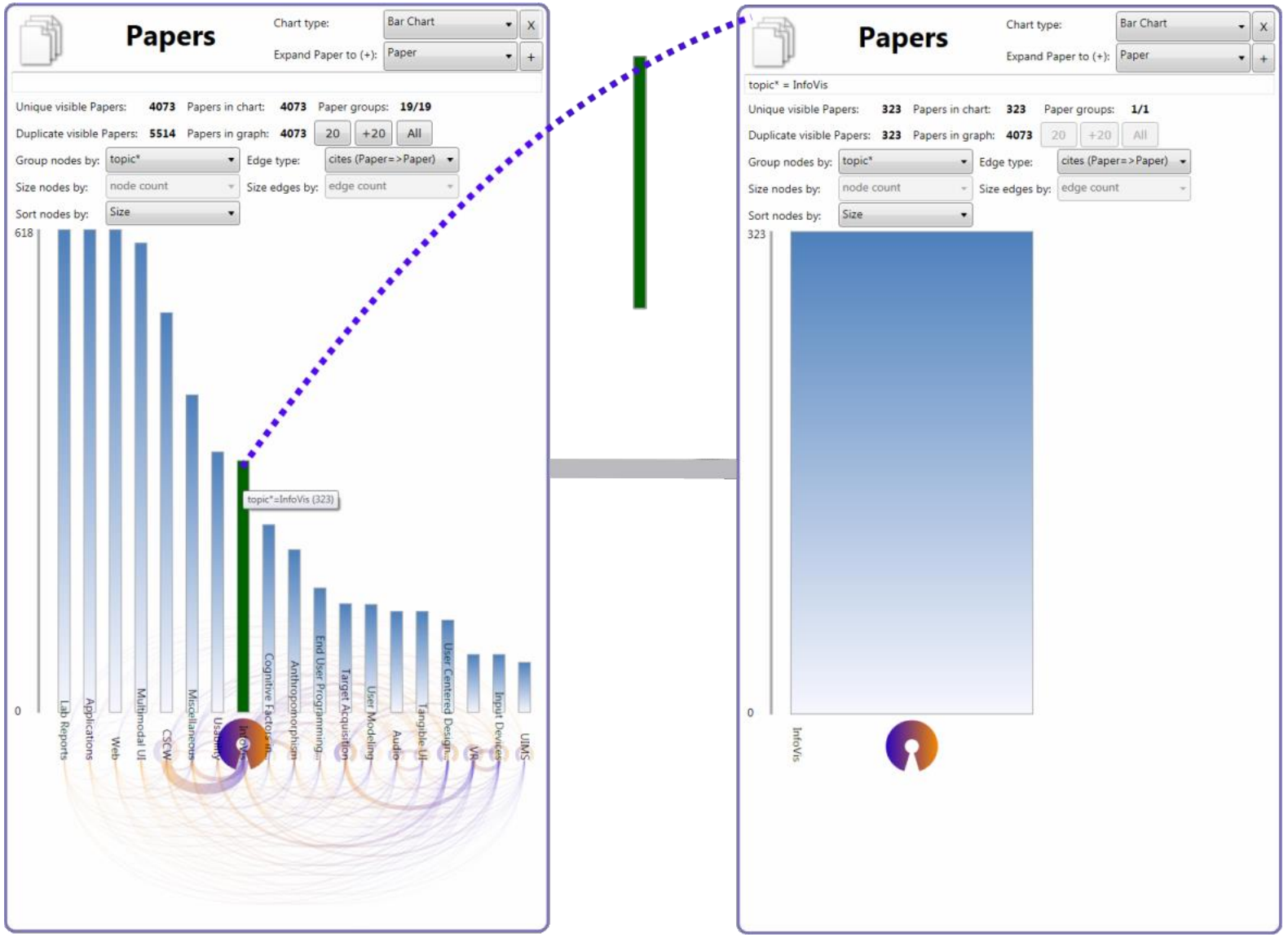






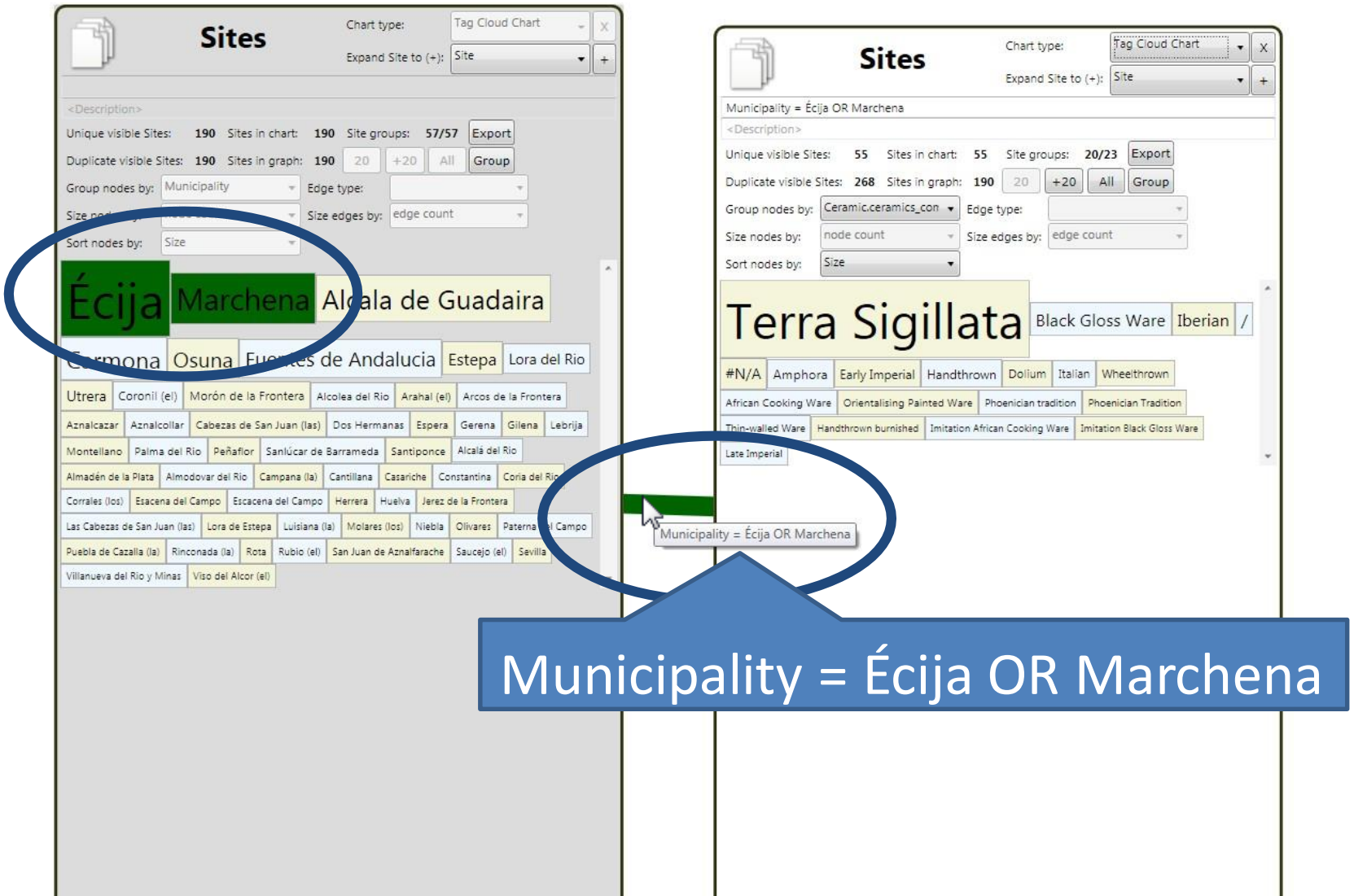
GraphTrail

Drag and drop interaction



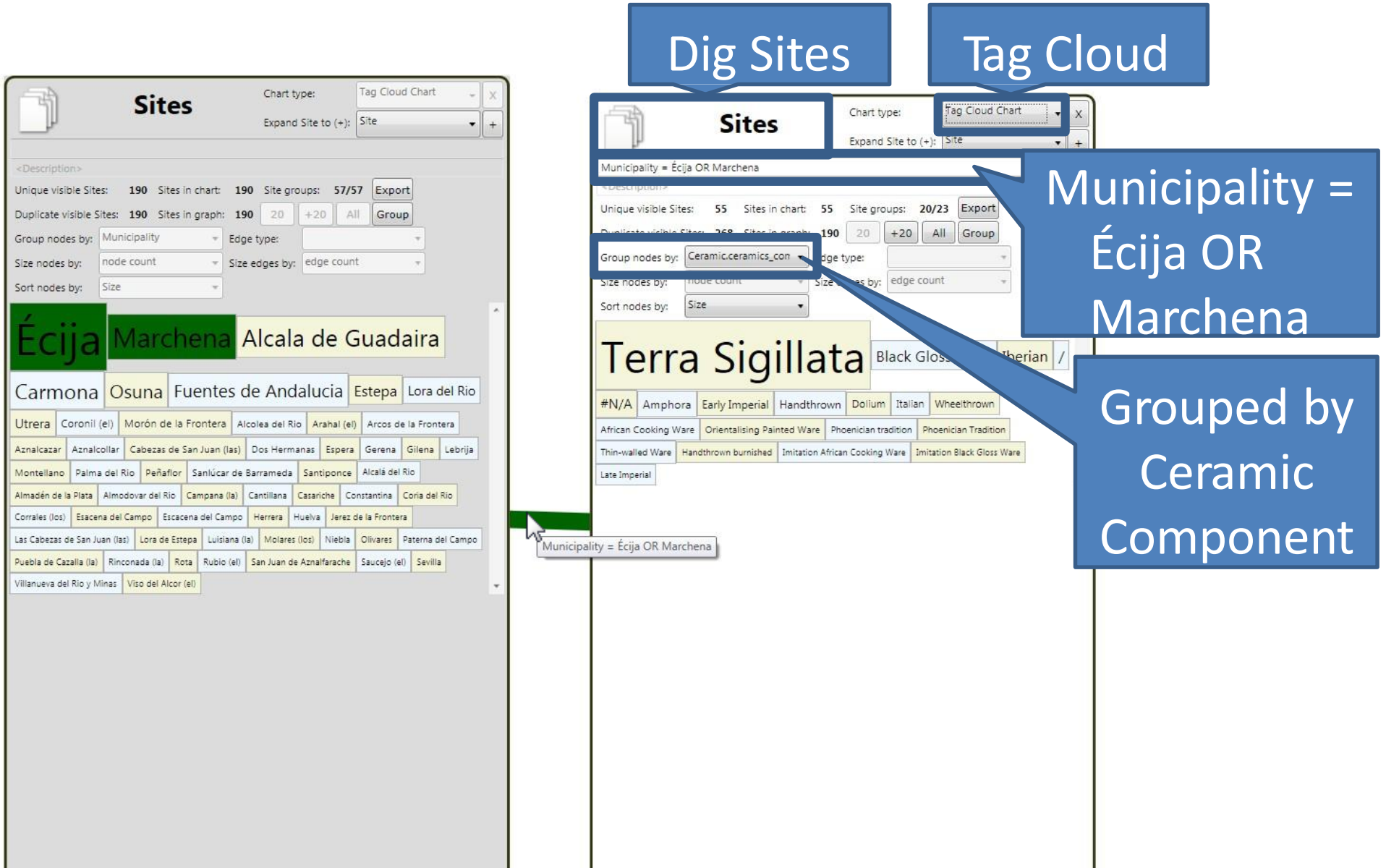
GraphTrail

Provenance/history visualization



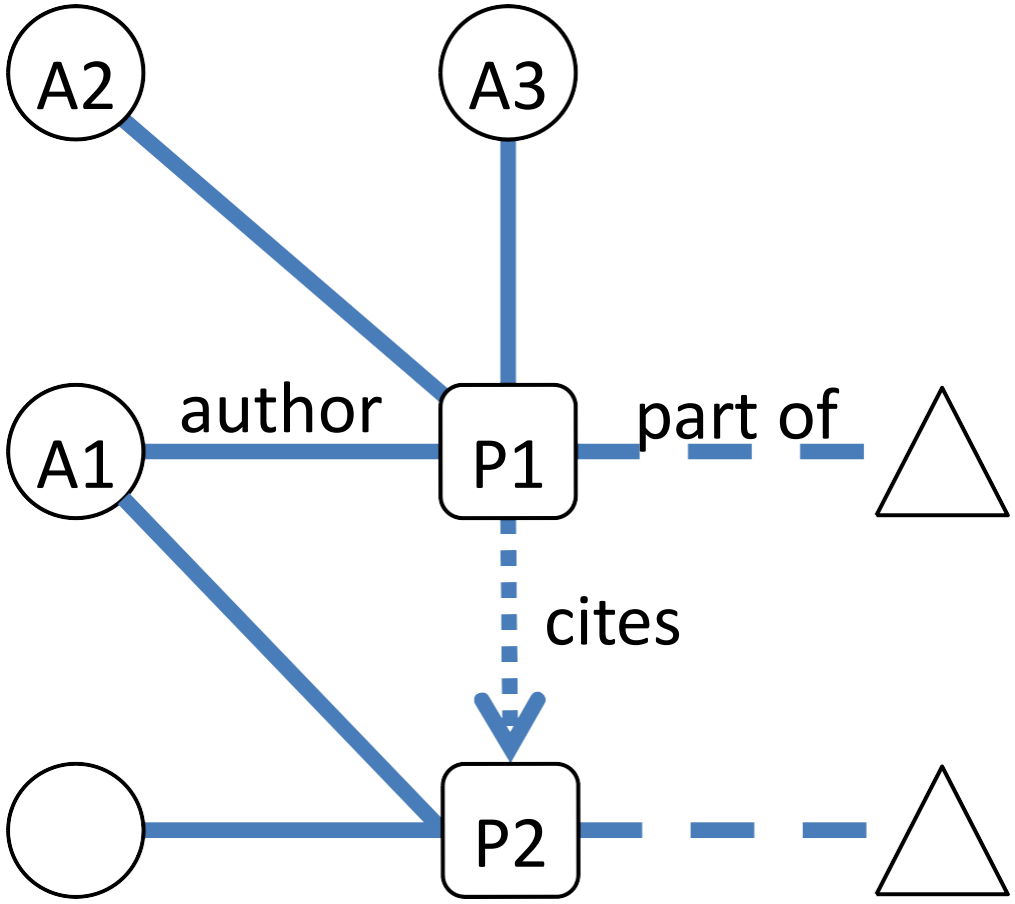
GraphTrail

Provenance & chart parameterization



GraphTrail

Pivoting & Derived Attributes



Author
- Name
- Affiliation
- Paper.Topic*

Paper
- Topic
- Year

Proceeding
- Country
- Date

GraphTrail

Lab study

- GraphTrail can make the same findings as other tools
  - And more!
- New users can make findings
- New users understand the exploration history
  - And usually motivation!

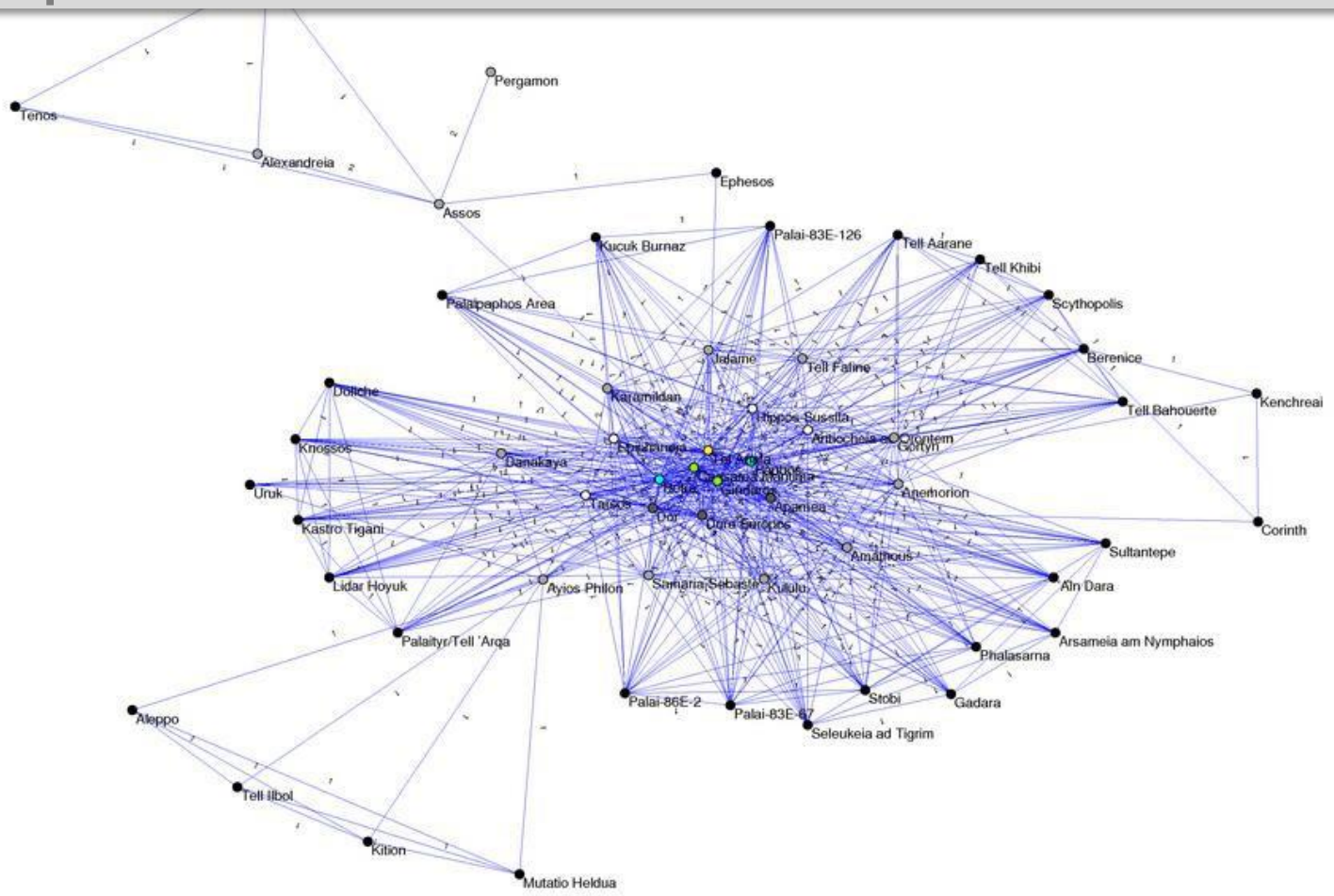
Field Study With Archaeologists

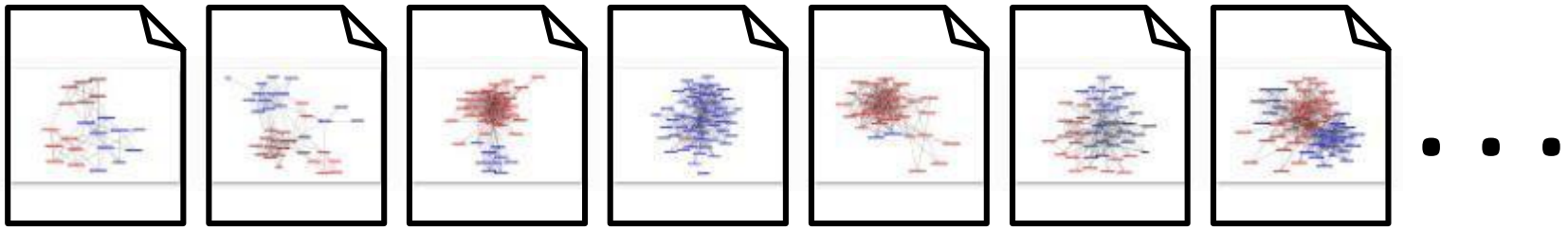
“How were Iron-Age communities integrated into the political and economic structure of the Roman Empire?”

“How were urban social hierarchies within the Roman provinces structured and articulated?”

GraphTrail

Field study – current practice





GraphTrail

Field study – analyses

1. Number of nodes, edges, types

	Nodes	Types	Edges	Types
CHI	10K	3	20K	3+
Archaeology	13K	24	20K	35

2. Number of charts

20 – 30 per session

GraphTrail

Discussion

- A system for exploring **large multivariate**, **heterogeneous networks** using **aggregation** by node and edge attributes,
- A method for capturing a user's **exploration history** and integrating it directly into the workspace, and
- A longitudinal **field study** and a qualitative **lab study** that prove the utility of these approaches.

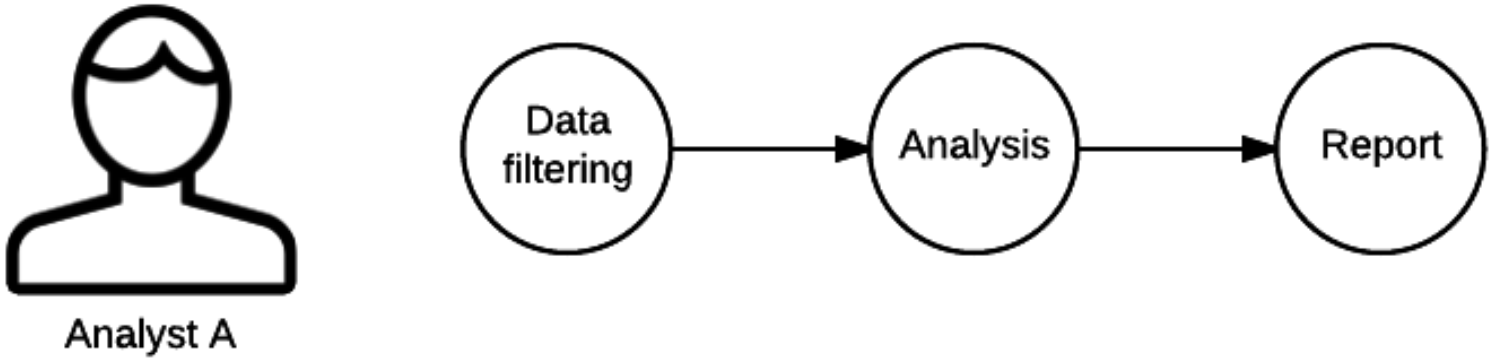
Dunne C, Riche NH, Lee B, Metoyer RA and Robertson GG (2012), *"GraphTrail: Analyzing large multivariate, heterogeneous networks while supporting exploration history"*, In CHI '12. pp. 1663-1672. DOI:10.1145/2207676.2208293

Riche N, Lee B and Dunne C (2011), *"Interactive visualization for exploring multi-modal, multi-relational, and multivariate graph data"*. U.S. Patent Application No. (13/041474).

StoryFacets

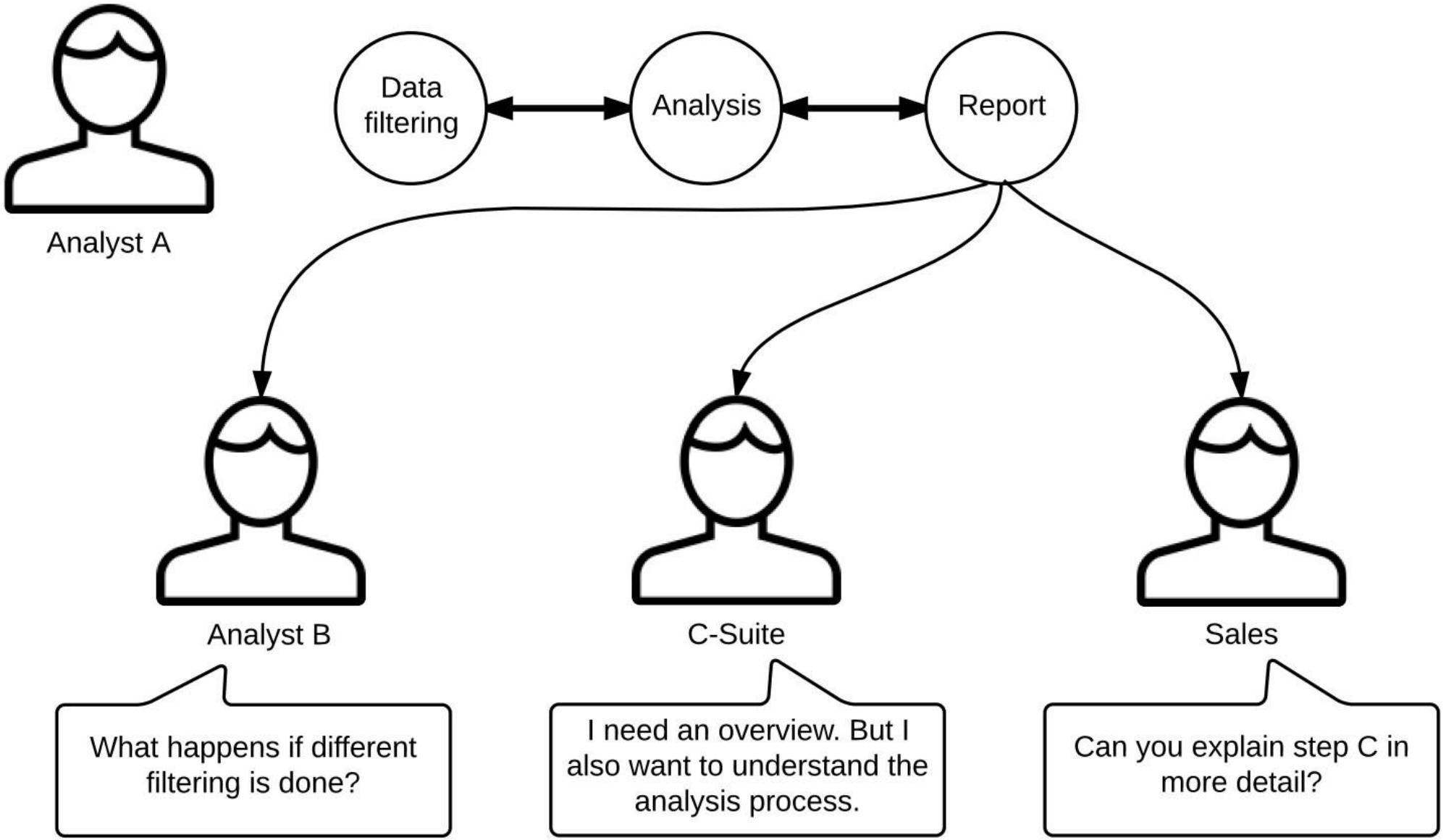
StoryFacets

Individual exploration



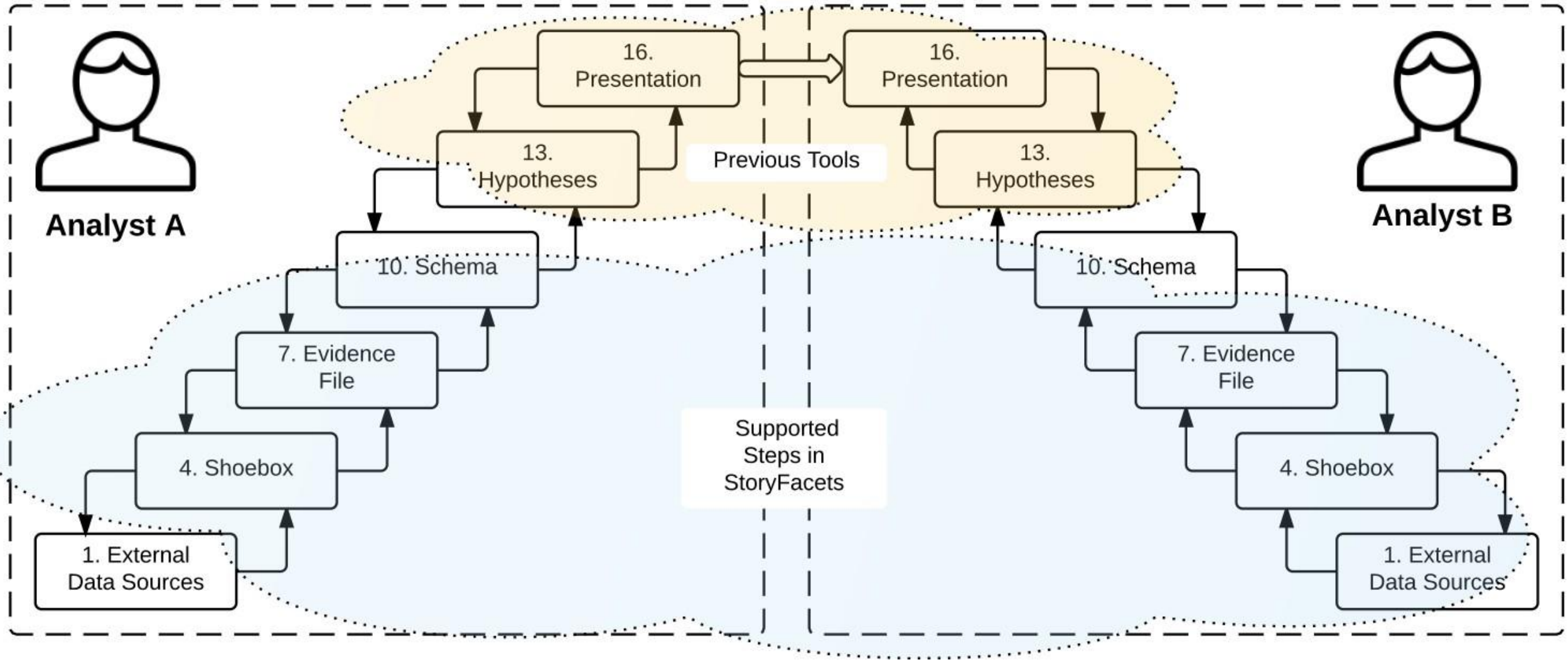
StoryFacets

Real-world exploratory analysis



StoryFacets

Pirolli and Card sensemaking loop




I am your database.

StoryFacets

Star Wars data

Films

Films	
title	string
episode_id	integer
opening_crawl	string
director	string
producer	string
release_date	date
species	array
starships	array
vehicles	array
characters	array
planets	array
url	string
created	string
edited	string

Starships

Starships	
name	string
model	string
starship_class	string
manufacturer	string
cost_in_credits	string
length	string
crew	string
passengers	string
max_atmosphere_speed	string
hyperdrive_rating	string
MGLT	string
cargo_capacity	string
consumables	string
films	array
pilots	array
url	string
created	string
edited	string

Species

Species	
name	string
classfication	string
average_height	Type
average_lifespan	string
eye_colors	string
hair_colors	string
skin_colors	string
language	string
homeworld	string
people	array
films	array
url	string
created	string
edited	string

People

People	
name	string
birth_year	string
eye_color	string
gender	string
hair_color	string
height	string
mass	string
skin_color	string
homeworld	string
films	array
species	array
starships	array
vehicles	array
url	string
created	string
edited	string

Vehicles

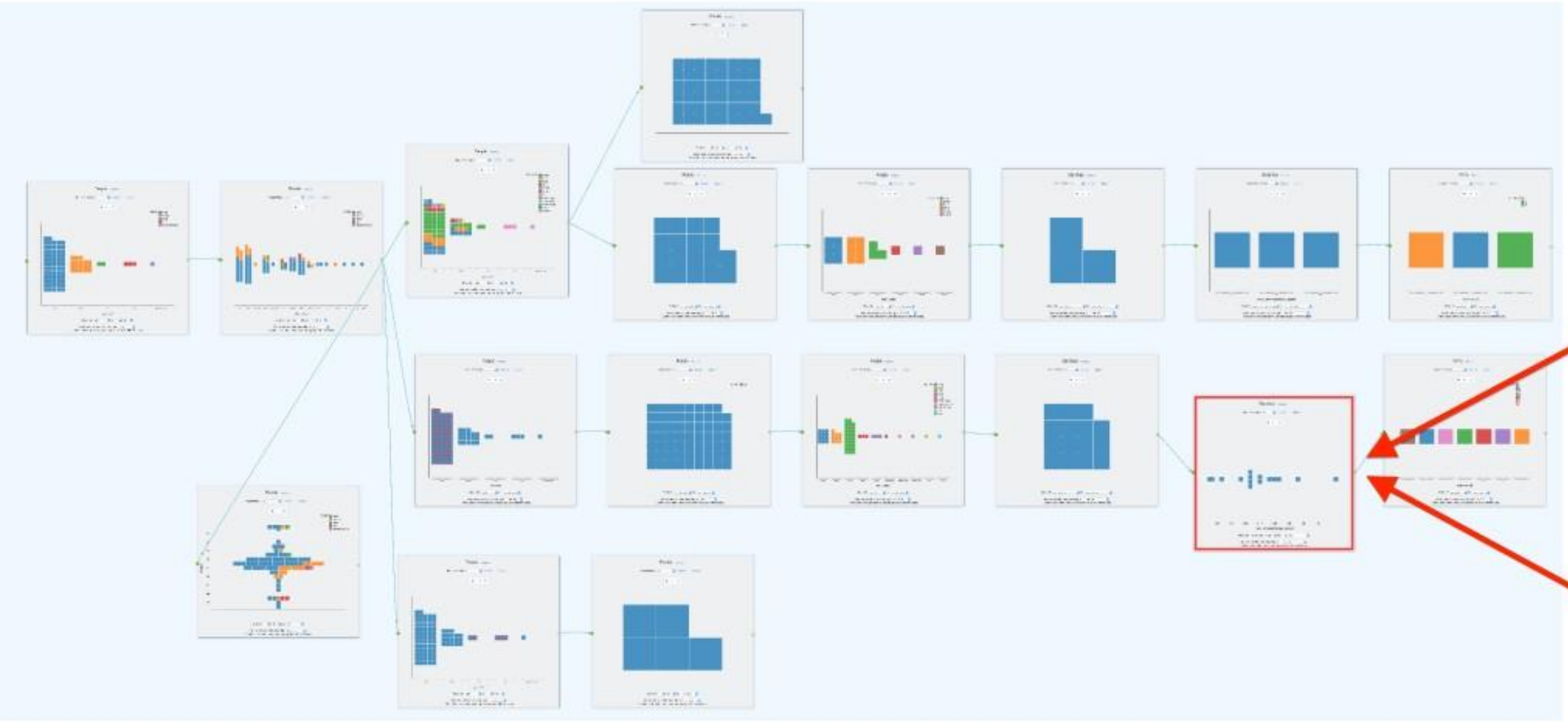
Vehicles	
name	string
model	string
vehicle_class	string
manufacturer	string
length	string
cost_in_credits	string
crew	string
passengers	string
max_atmosphering_speed	string
cargo_capacity	string
consumables	string
films	array
pilots	array
url	string
created	string
edited	string

Planets

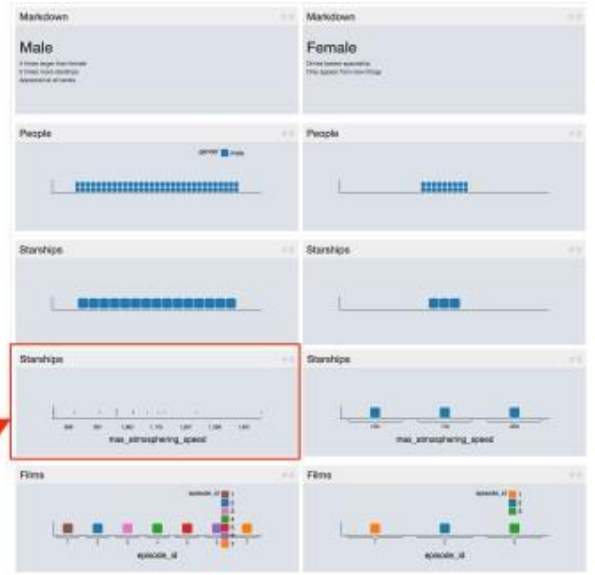
Planets	
name	string
diameter	string
rotation_speed	string
orbital_period	string
gravity	string
population	string
climate	string
terrain	string
surface_water	string
residents	array
films	array
url	string
created	string
edited	string

StoryFacets

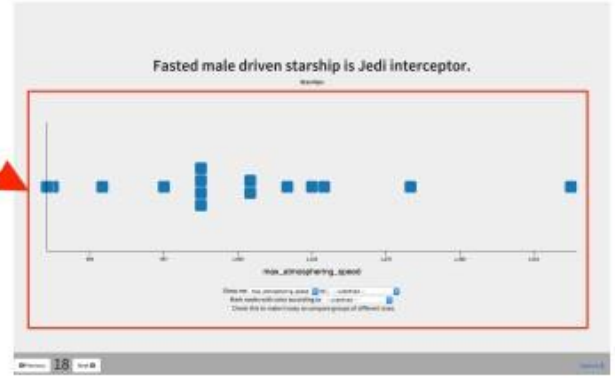
Overview



(a) Trail Facet



(b) Dashboards or Info Graphics Facet



(c) Story Facet

Demo Preview

STORYFACETS

Generating Multiple Representations of Exploratory Data Analysis for Communication

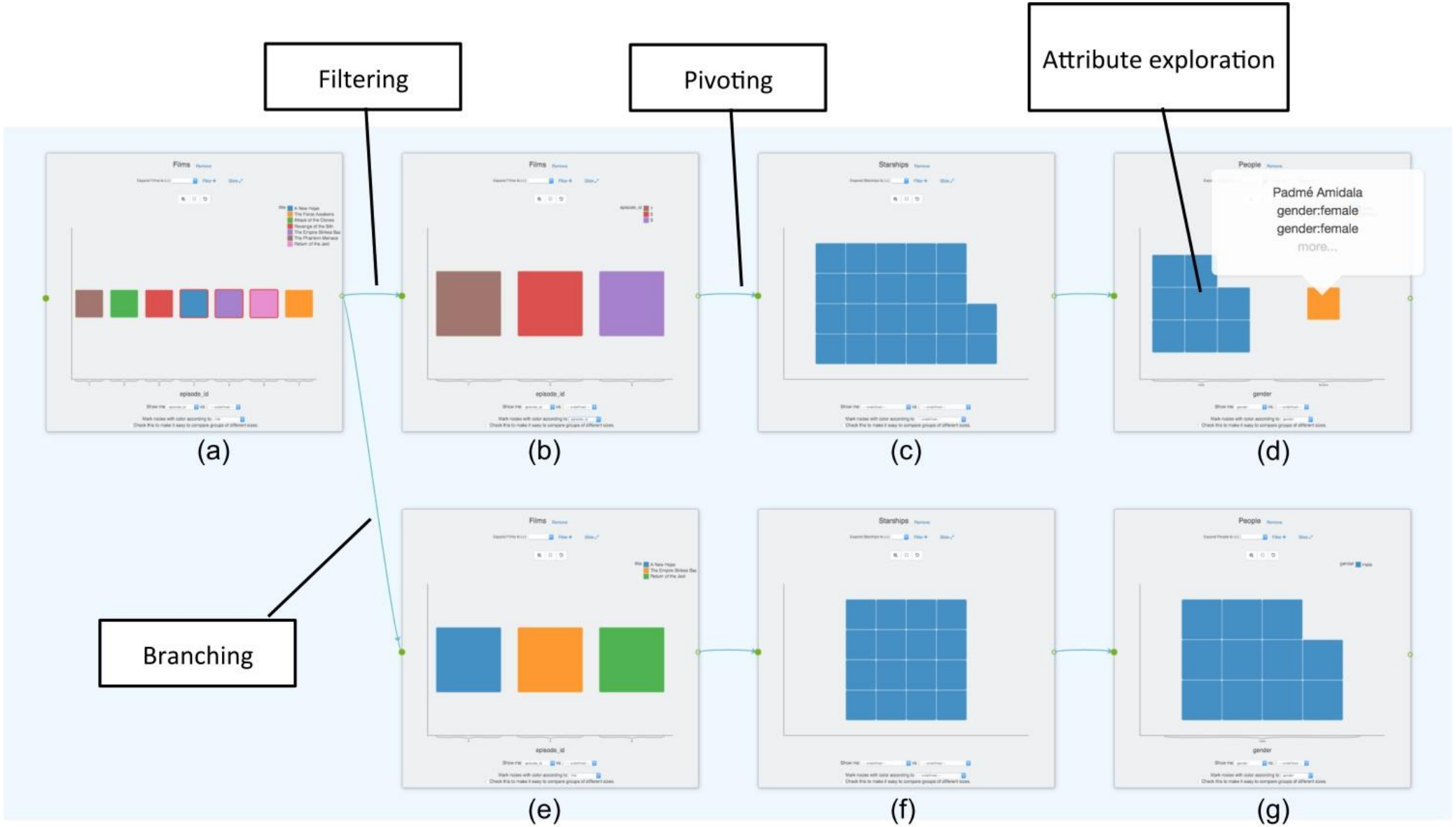
Previous 0 Next

subTitle:

```
# StoryFacets
#### Generating Multiple Representations of
Exploratory Data Analysis for Communication
```

StoryFacets

Trail facet – Star Wars trilogy ships & pilots



StoryFacets

Dashboard facet – Star Wars male vs. female

StoryFacets

Infographic facet – Star Wars Jabba the Hutt

Markdown

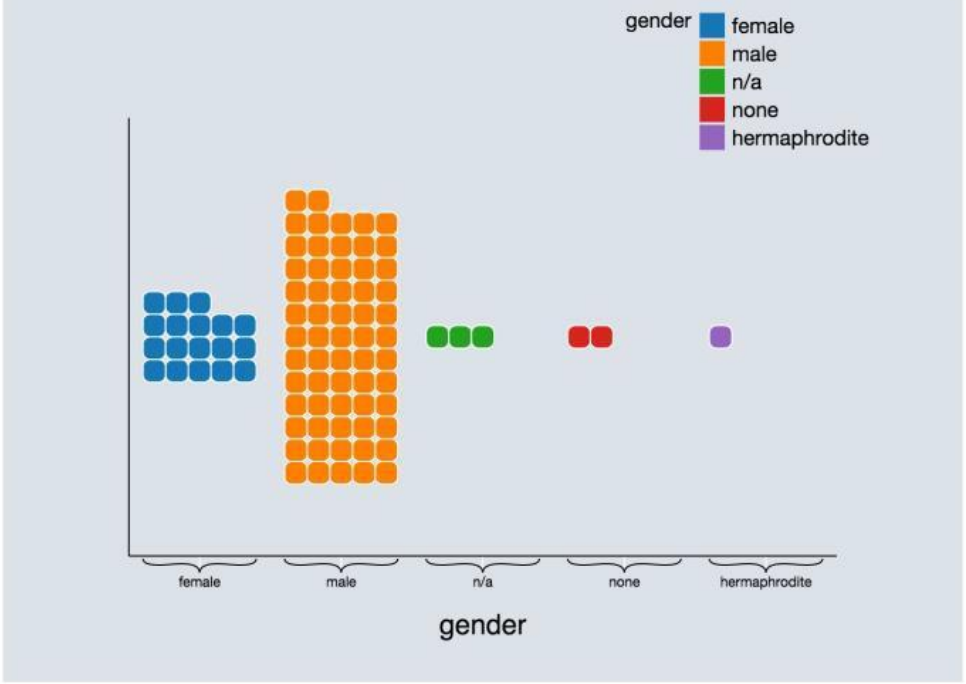
(a)

Markdown

The only hermaphrodite in the Starwars is Jabba Desilijic Tiure

(b)

People



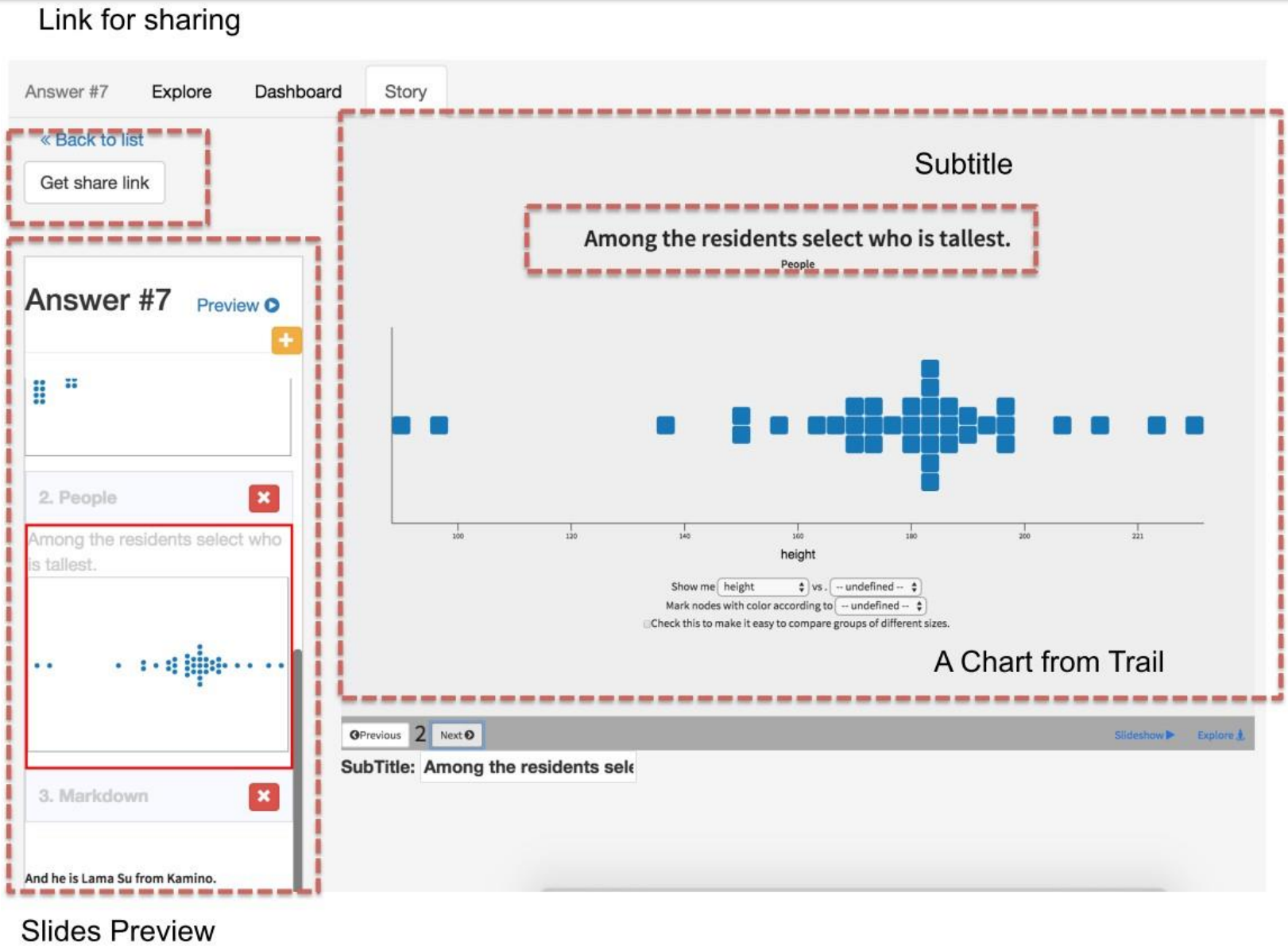
(c)

Markdown

(d)

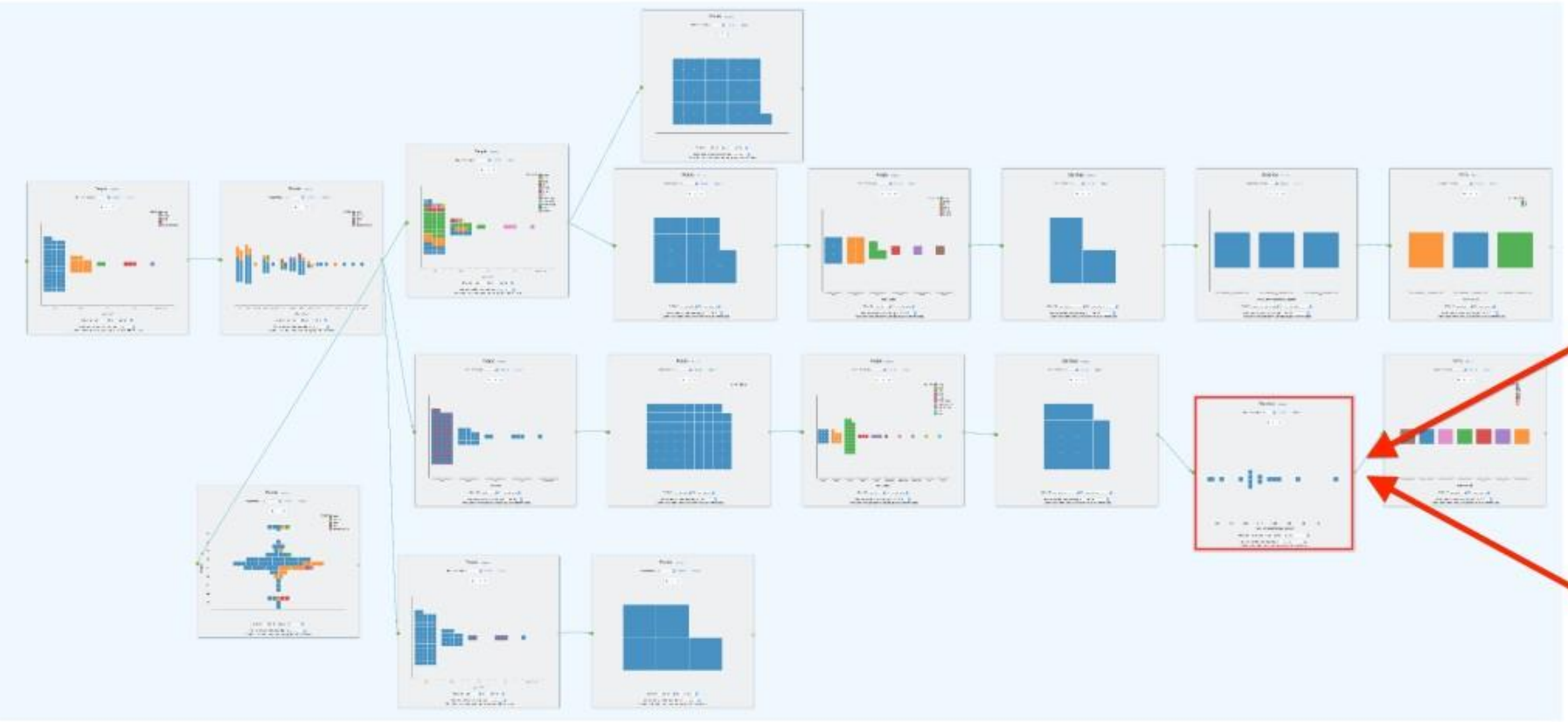
StoryFacets

Story (slideshow) facet – Star Wars character height

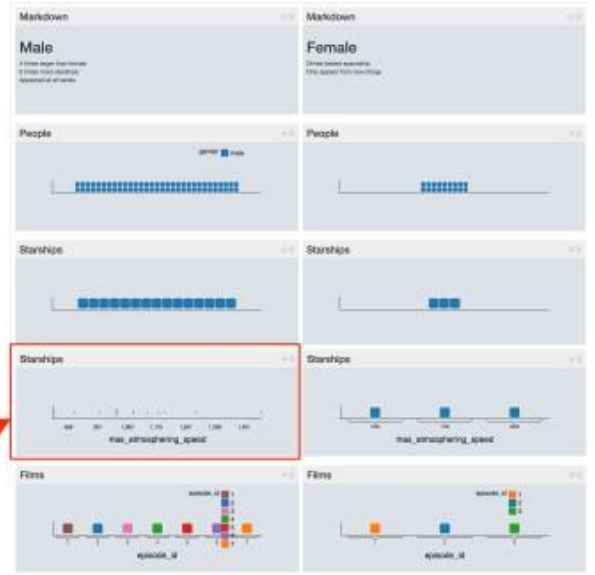


StoryFacets

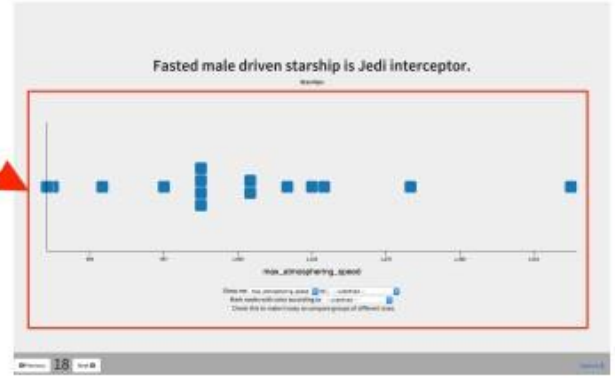
Linked back to trail facet



(a) Trail Facet



(b) Dashboards or Info Graphics Facet



(c) Story Facet

StoryFacets

Evaluation

- 19 casual participants in a public setting
  - Novice users were able to select different communication format based on the target audience and nature of the content
- Three visualization professional expert reviewers

StoryFacets

Discussion

- Exploratory data analysis is much more than the initial exploration session
- Unified platform to support exploration, collaboration, discussion, and presentation
- Study with casual participants and expert reviewers reveal future directions

Park DG, Dunne C, Ragan E, Elmqvist N (2015) *"StoryFacets: Generating Multiple Representations of Exploratory Data Analysis for Communication",* Under submission.

Discussion


Introduction
Node-Link
VoroGraph
GraphTrail
StoryFacets
Discussion

Graph Drawing 2017

Hosted in September by IBM

Cambridge, MA, USA

Cody Dunne, IBM Watson
cdunne@us.ibm.com

T. Alan Keahey, IBM Watson
alan.keahey@us.ibm.com

iCity: Urban Informatics for Sustainable Metropolitan Growth

Version 2, November 2015

A project funded by:

The Ontario Research Fund - Research Excellence Round 7

Principal Investigator:

Prof. Eric J. Miller, Research Director

University of Toronto Transporlation Research Institute

The research team

University of Toronto
Eric Miller
Baher Abdulhai
Steve Easterbrook
Judy Farvolden
Mark Fox
Khandker Nurul Habb
Paul Hess
Matt Roorda
Amer Shallaby
Matti Siemiatycki
Enid Slack

OCAO University
Sara Diamond
Greg Van Alstyne
Isabel Meirelles
University of Waterloo
Jeff cassello
University of Regina
Mohamed El-Darleby
IBM
Cody Dunne
Magda Mourad
Sambit Sahu

Our partners

Cellint
City of Toronto
IBM Canada
Maximum City
Region of Waterloo
Waterfront Toronto
and supporters
I-Canada
Metrolinx
Ontario Ministry of Transportation
Ontario Ministry of Municipal Affairs & Housing
VARDEC

UNIVERSITY OF TORONTO
FACULTY o, APPLIED SCIENCE " ENGINEERING
TrJ.nSportation RCSf"arch Institute

GraphTrail & StoryFacets

Blocks to build on

- Additional visualizations
- Context aware comments
- Exploration hints for new paths
- Streaming/temporal data, intelligent updates, and resurgent relevancy
- Linked chart parameterization for comparisons with auto chain layout and compression
- Advanced modeling and analytics (IBM Catalyst, IBM Watson)
- User management & security