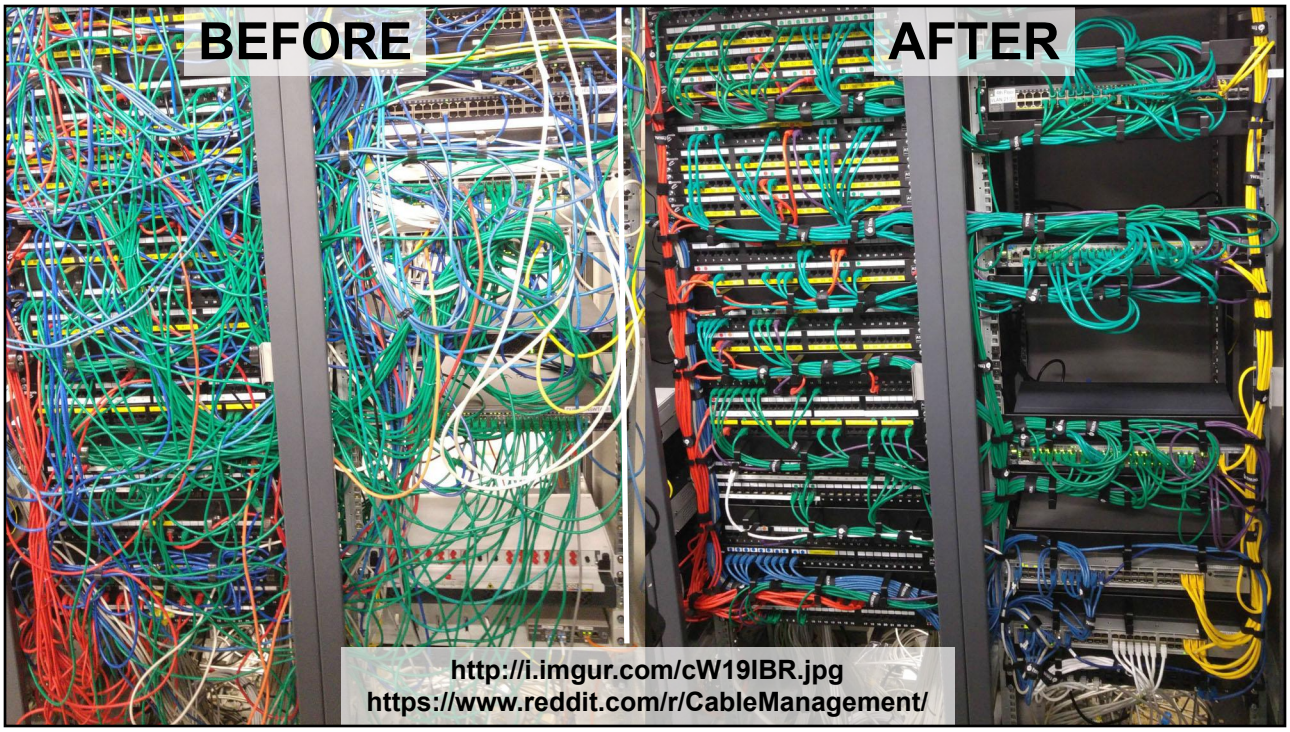
CSCI 4550/6550 Interactive Visualization

https://www.cs.rpi.edu/~cutler/classes/visualization/S24/

Lecture 4: Graph Visualization II: Clarity, Simplification, & Interaction



Today

- Worksheet & HW 3: Graphing RPI CSCI Course & Instructor Data
- Reading: Clustering/Hierarchical Edge Bundling
- Reading: Lombardi Graphs
- Emergency Management Graph Visualization
  - Sean Kim's masters project
- Preview of "High-Dimensional" Data Visualization
- Readings for Tuesday

RPI's Standard for Degree Requirement Templates

- Advantages?
  - You know what you're doing every term, you have a plan
  - Simpler to look at than degreeworks
  - Quick & easy, & symmetric (pretty)
  - Understand average CSCI student
  - Structure is comforting, don't need to ask for help to understand basic requirements
  - Template is helpful, see how its possible to complete requirements
  - Organized by term, chronologic
  - Familiar, standardized across RPI

- What important information is missing?
  - No information/guidelines on free electives
  - Lots of footnotes
  - Doesn't have important stuff like comm intensive requirement
  - If you miss or fail something, then everything later may be impacted
  - Fall vs spring offerings not visualized
  - Prereqs aren't visualized, don't know what will be impacted if a course is delayed
  - If have AP credits, don't know how to adapt
  - Doesn't have special case
  - Not shown that you can take > 16 per term (graduate a term early)
  - It is too rigid, if you get off schedule, you lose structure, have to talk to an advisor, that can be scary

First Year

Fall 2022		Spring 2023	
CSCI 1100 Computer Science I[1]	4	CSCI 1200 Data Structures	4
MATH 1010 Calculus I	4	MATH 1020 Calculus II	4
PHYS 1100 Physics I[2]	4	BIOL 1010 Intro. to Biology[2]	3
HASS Elective	4	BIOL 1015 Intro. to Biology Lab[2]	1
		HASS Elective	4

Second Year

Fall 2023		Spring 2024	
CSCI 2200 Foundations of CS	4	CSCI 2300 Intro. to Algorithms	4
CSCI 2500 Computer Organization[3]	4	CSCI 2600 Principles of Software[4]	4
Mathematics Option I	4	Mathematics Option II	4
HASS Elective	4	HASS Elective	4

Third Year

Arch Summer 2024		Fall 2024 -or- Spring 2025	
CSCI 4210 Operating Systems[8]	4	CSCI 4430 Programming Languages[6] -or- CS Option/Capstone	4
CS Option/Capstone -or- Free Elective[5]	4	Science Option	4
HASS Elective	4	HASS Elective	4
Free Elective	4	Free Elective	4

Fourth Year

Fall 2025		Spring 2026	
CS Option/Capstone -or- CSCI 4430 Programming Languages[6]	4	CS Option/Capstone	4
CS Option/Capstone	4	Free Elective -or- CS Option/Capstone	4
Free Elective	4	Free Elective	4
Free Elective	4	Free Elective	4

Today's Worksheet: Graph Drawing Design

- Pair up with someone new!
- Design a graph for the CSCI major template
- Visualize "Who teaches the CSCI courses"

First Year

Fall 2022		Spring 2023	
CSCI 1100 Computer Science I[1]	4	CSCI 1200 Data Structures	4
MATH 1010 Calculus I	4	MATH 1020 Calculus II	4
PHYS 1100 Physics I[2]	4	BIOL 1010 Intro. to Biology[2]	3
HASS Elective	4	BIOL 1015 Intro. to Biology Lab[2]	1
		HASS Elective	4

Second Year

Fall 2023		Spring 2024	
CSCI 2200 Foundations of CS	4	CSCI 2300 Intro. to Algorithms	4
CSCI 2500 Computer Organization[3]	4	CSCI 2600 Principles of Software[4]	4
Mathematics Option I	4	Mathematics Option II	4
HASS Elective	4	HASS Elective	4

Third Year

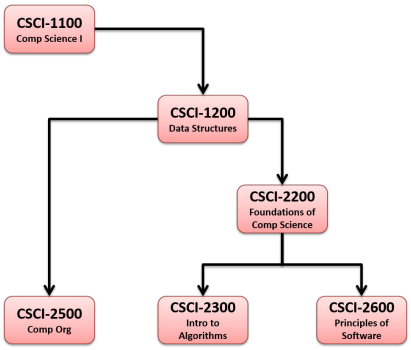
Arch Summer 2024		Fall 2024 -or- Spring 2025	
CSCI 4210 Operating Systems[6]	4	CSCI 4430 Programming Languages[6] -or- CS Option/Capstone	4
CS Option/Capstone -or- Free Elective[5]	4	Science Option	4
HASS Elective	4	HASS Elective	4
Free Elective	4	Free Elective	4

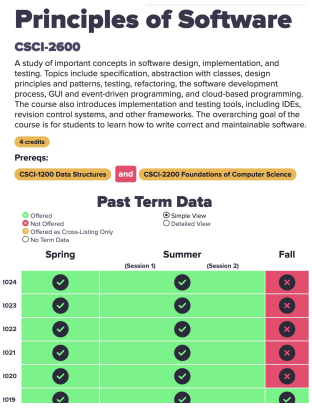
Fourth Year

Fall 2025		Spring 2026	
CS Option/Capstone -or- CSCI 4430 Programming Languages[6]	4	CS Option/Capstone	4
CS Option/Capstone	4	Free Elective -or- CS Option/Capstone	4
Free Elective	4	Free Elective	4
Free Elective	4	Free Elective	4

HW 3: Graphviz & RPI CSCI Course Data

- Learn GraphViz: Open-source software for automated graph drawing
- Parse JSON data from QuACS





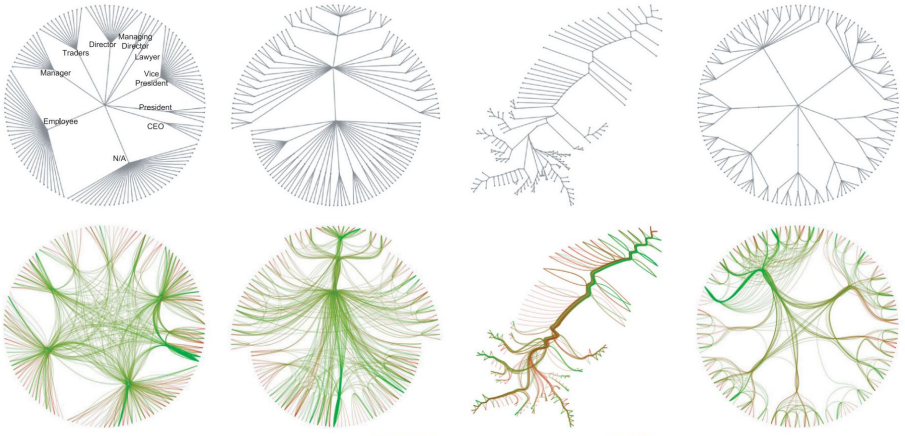

```
{
  "crse": 4550,
  "id": "CSCI-4550",
  "sections": [
    {
      "act": 39,
      "attribute": "Communication Intensive",
      "cap": 45,
      "credMax": 4.0,
      "credMin": 4.0,
      "crn": 93902,
      "crse": 4550,
      "rem": 6,
      "sec": "01",
      "subj": "CSCI",
      "timeslots": [
        {
          "dateEnd": "04/24",
          "dateStart": "01/08",
          "days": [
            "T",
            "F"
          ],
          "instructor": "Barbara Cutler",
          "location": "Troy Building 2012",
          "timeEnd": 1550,
```

Today

- Worksheet & HW 3: Graphing RPI CSCI Course & Instructor Data
- Reading: Clustering/Hierarchical Edge Bundling
- Reading: Lombardi Graphs
- Emergency Management Graph Visualization
  - Sean Kim's masters project
- Preview of "High-Dimensional" Data Visualization
- Readings for Tuesday

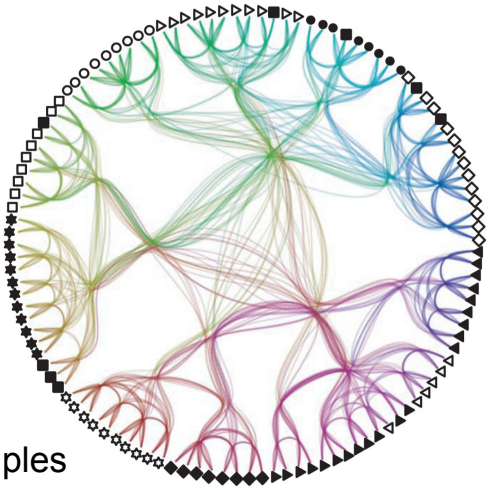
Readings for Friday *(pick one)*

"Social Network Clustering and Visualization using Hierarchical Edge Bundles", Jia, Garland, & Hart, Computer Graphics Forum, 2011.

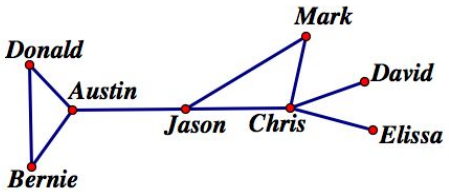


Manual (ideal) Strength clustering [ACJM03] BC communities [GN02] Our method

- Color is very helpful – except concerns about colorblindness
- Relation to '6 degrees of separation'
- Concern that small (but important) features or communities may be lost
- Concern about high cost of computing betweenness centrality (BC) metric
- Final graphs can still be confusing to interpret
- Diagrams very helpful in explaining steps of algorithm
- What are some anecdotal real-world examples of a "low BC edge" and a "high BC edge"?
  - Why does it make sense to preserve or simplify away these edges?
  - Why is a tree the best simplified representation of a big complex graph?



Betweenness Centrality



- For each pair of nodes "A" and "C" in the graph
- Compute the shortest path between A & C
- Is B on that path? A → B → C? How many of the paths?
- "Betweenness Centrality of node B" = # of times B appears in these shortest paths
- (Assuming?) Similar definition for "Betweenness Centrality of edge $B_1 \rightarrow B_2$"

	Jason	Austin	Donald	Bernie	Chris	Mark	David	Elissa
Jason		*	A	A	*	*	C	C
Austin			*	*	J	J	J,C	J,C
Donald				*	A,J	A,J	C,J,A	C,J,A
Bernie					A,J	A,J	C,J,A	C,J,A
Chris						*	*	*
Mark							C	C
David								C
Elissa								

Jason	12
Austin	10
Donald	0
Bernie	0
Chris	11
Mark	0
David	0
Elissa	0

http://mathforum.org/mathimages/index.php/Social_Networks#Betweenness_Centrality

Today

- Worksheet & HW 3: Graphing RPI CSCI Course & Instructor Data
- Reading: Clustering/Hierarchical Edge Bundling
- Reading: Lombardi Graphs
- Emergency Management Graph Visualization
  - Sean Kim's masters project
- Preview of "High-Dimensional" Data Visualization
- Readings for Tuesday

Readings for Friday *(pick one)*

"Force-directed Lombardi-style graph drawing", Chernobelskiy et al., Graph Drawing 2011.

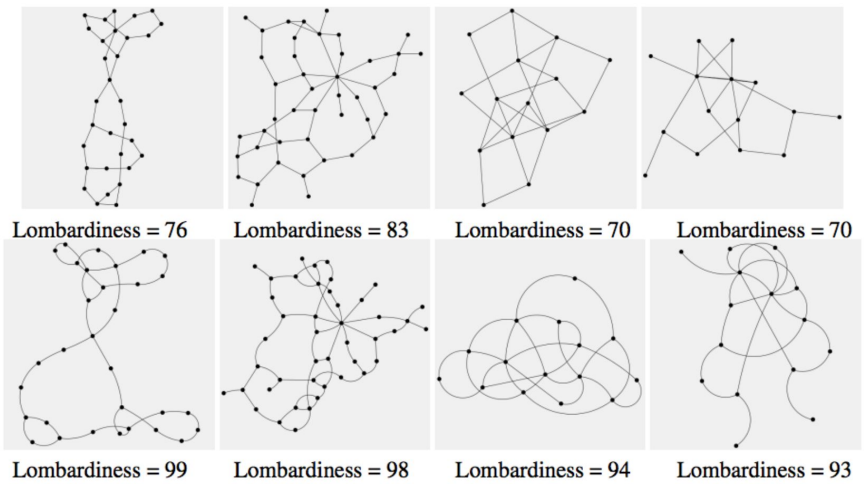
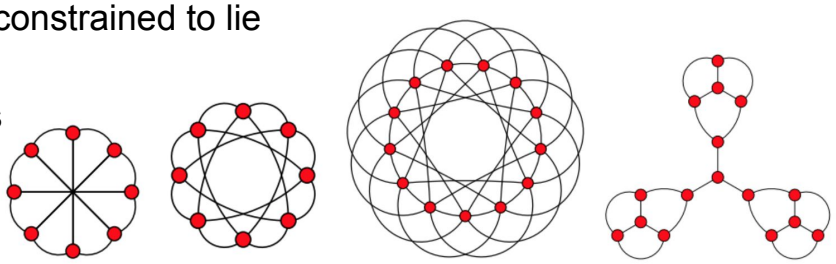


Fig. 5. Standard force-directed drawings (above) and near-Lombardi drawings (below).

"Lombardi Drawings of Graphs", Duncan, Eppstein, Goodrich, Kobourov, Nollenberg, Graph Drawing 2010

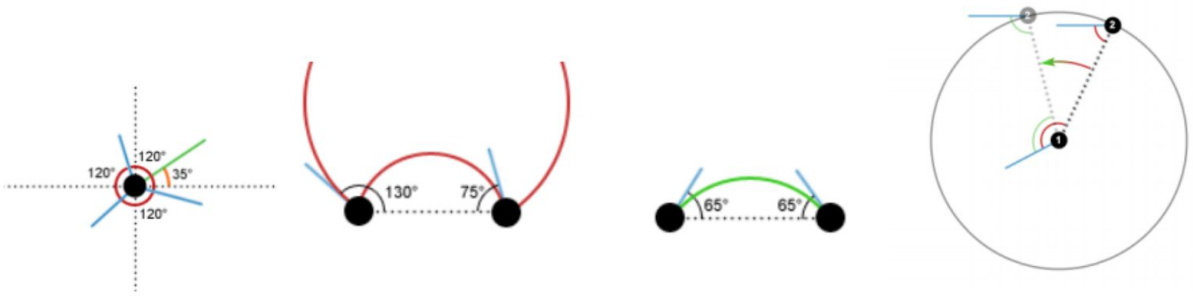
BONUS PAPER

- Circular arcs
- Perfect angular resolution (edges for equal angles at vertices)
- Arcs only intersect 2 vertices (at endpoints)
- *(not required to be crossing free)*
- Vertices may be constrained to lie on circle or on concentric circles



(a) (b) (c) (d)

- Relaxation of the Lombardi Graph requirements
- “straight-line segments rarely occur in nature ... it is not clear that humans prefer straight-line segments for the sake of graph readability”
- Forces on tangent angles as well as on vertex positions



(a) (b) (c) (d)

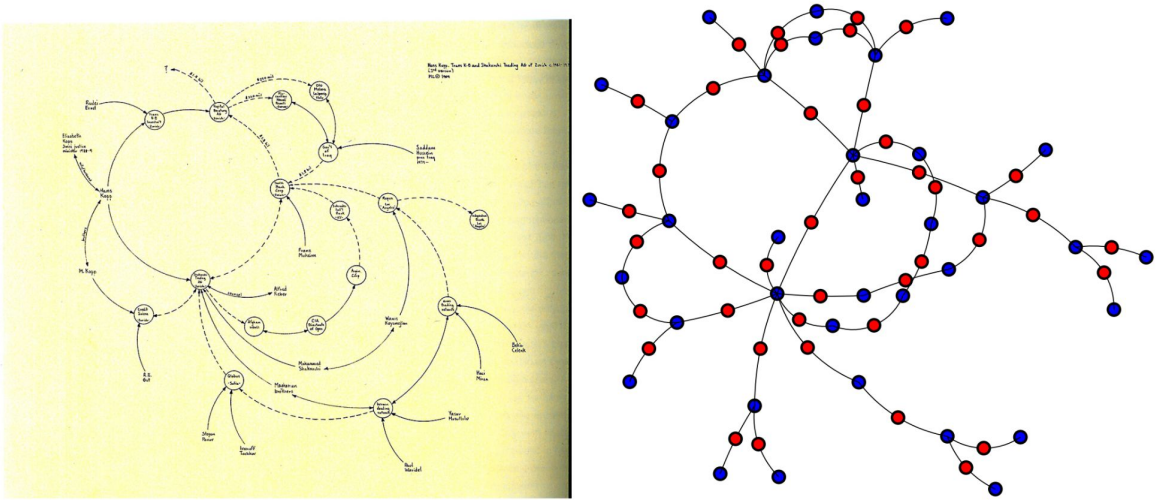
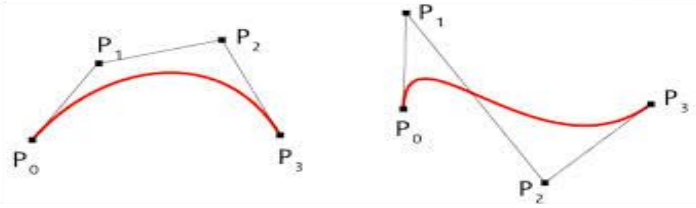


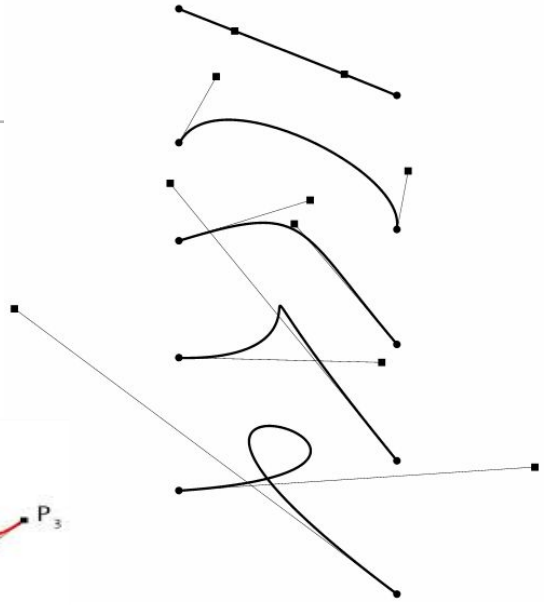
Fig. 8. The graph for Lombardi's Hans Kopp, Trans K-B and Shakarchi Trading [26], shown as rendered by Lombardi and as rendered by our force-directed method based on the use of dummy vertices.

- People are more patient with aesthetically pleasing graphs (will spend longer studying to learn/draw conclusions)
- What about relaxing the circular arc requirement and allowing Bezier arcs?
- How does it scale to larger data?
- Long curved arcs can be much harder to follow
- Circular layout of nodes is often very good!
- Would like more pseudocode
- *How well would graphs drawn by Mark Lombardi score on the "Lombardiness" metric?*
  - *Equal angles is not intrinsic to his drawings!*
  - *What alternate geometric properties are found in his drawnings? Is it possible to write code to mimic those properties?*

Cubic Bézier Curve

- 4 control points
- Curve passes through first & last control point
- Curve is tangent at $\mathbf{P_0}$ to ($\mathbf{P_1}$- $\mathbf{P_0}$) and at $\mathbf{P_3}$ to ($\mathbf{P_3}$- $\mathbf{P_2}$)



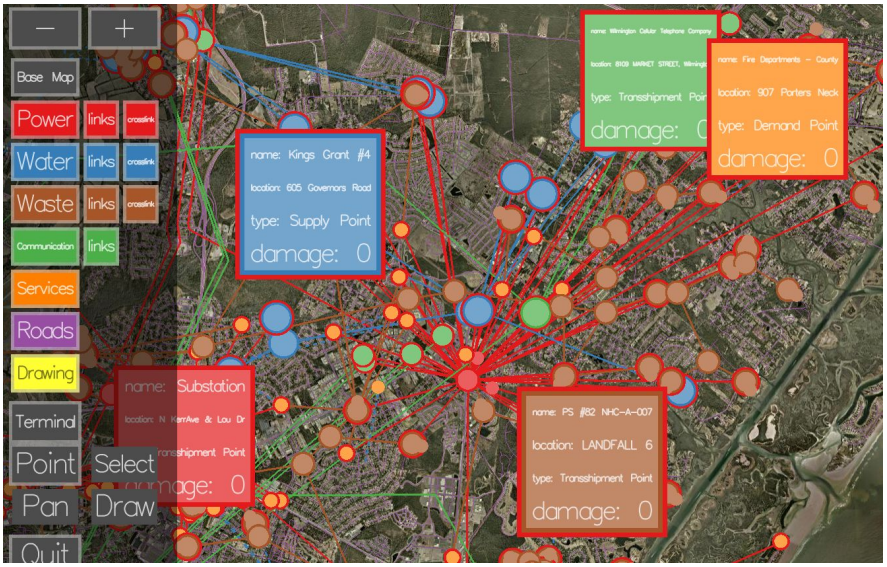


http://www.e-cartouche.ch/content_reg/cartouche/graphics/en/html/Curves_learningObject2.html
http://www.webreference.com/dlab/9902/bezier.html

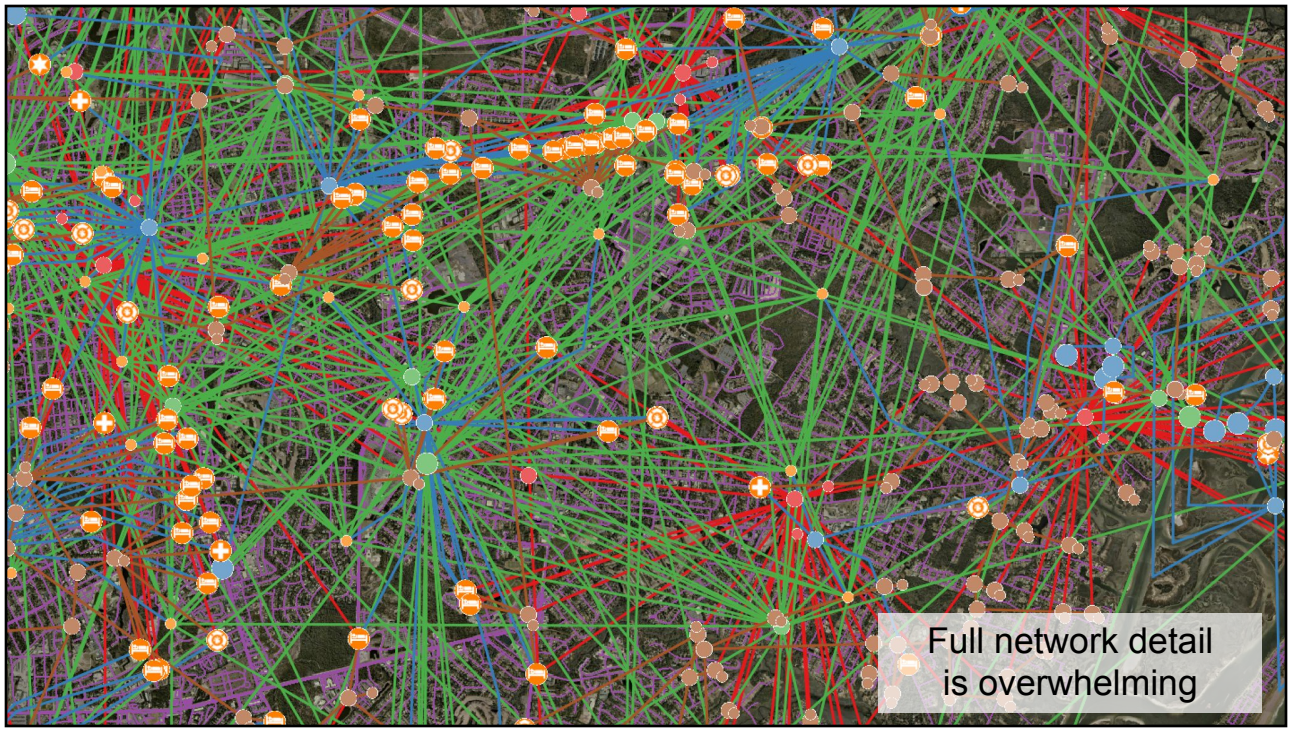
Today

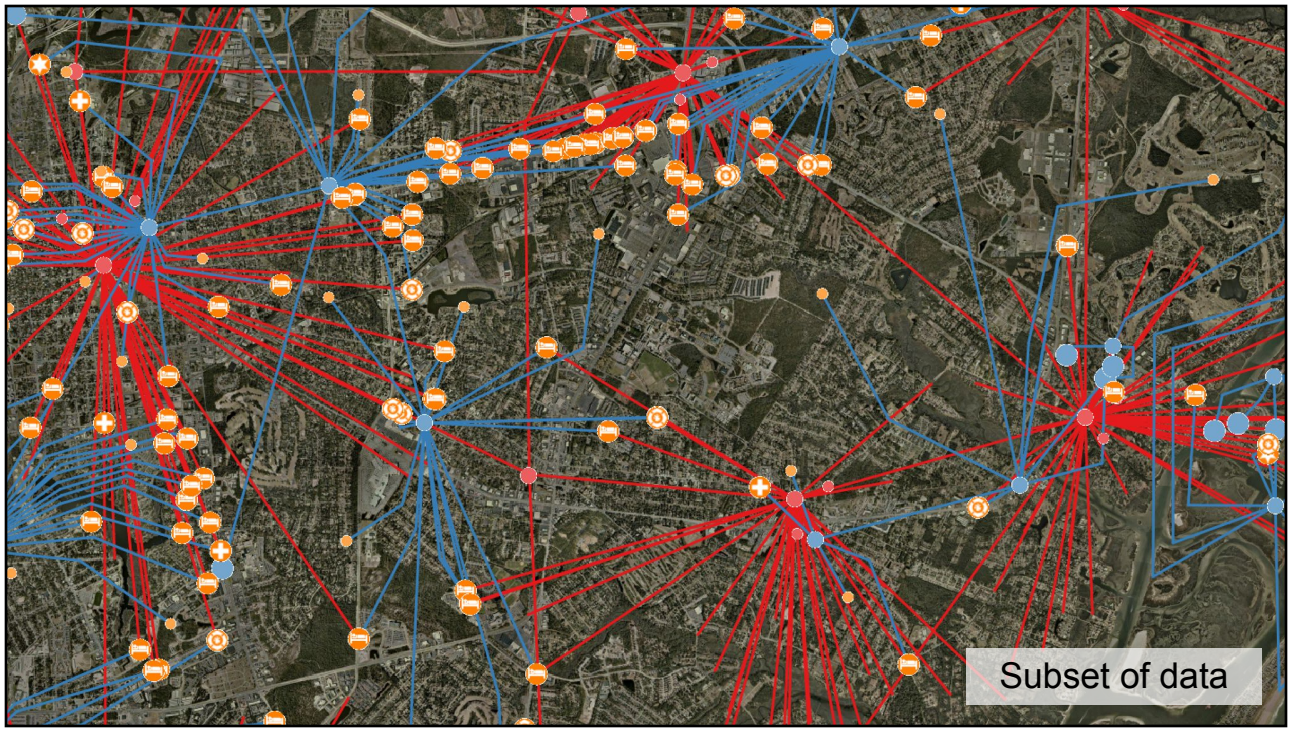
- Worksheet & HW 3: Graphing RPI CSCI Course & Instructor Data
- Reading: Clustering/Hierarchical Edge Bundling
- Reading: Lombardi Graphs
- Emergency Management Graph Visualization
  - Sean Kim's masters project
- Preview of "High-Dimensional" Data Visualization
- Readings for Tuesday

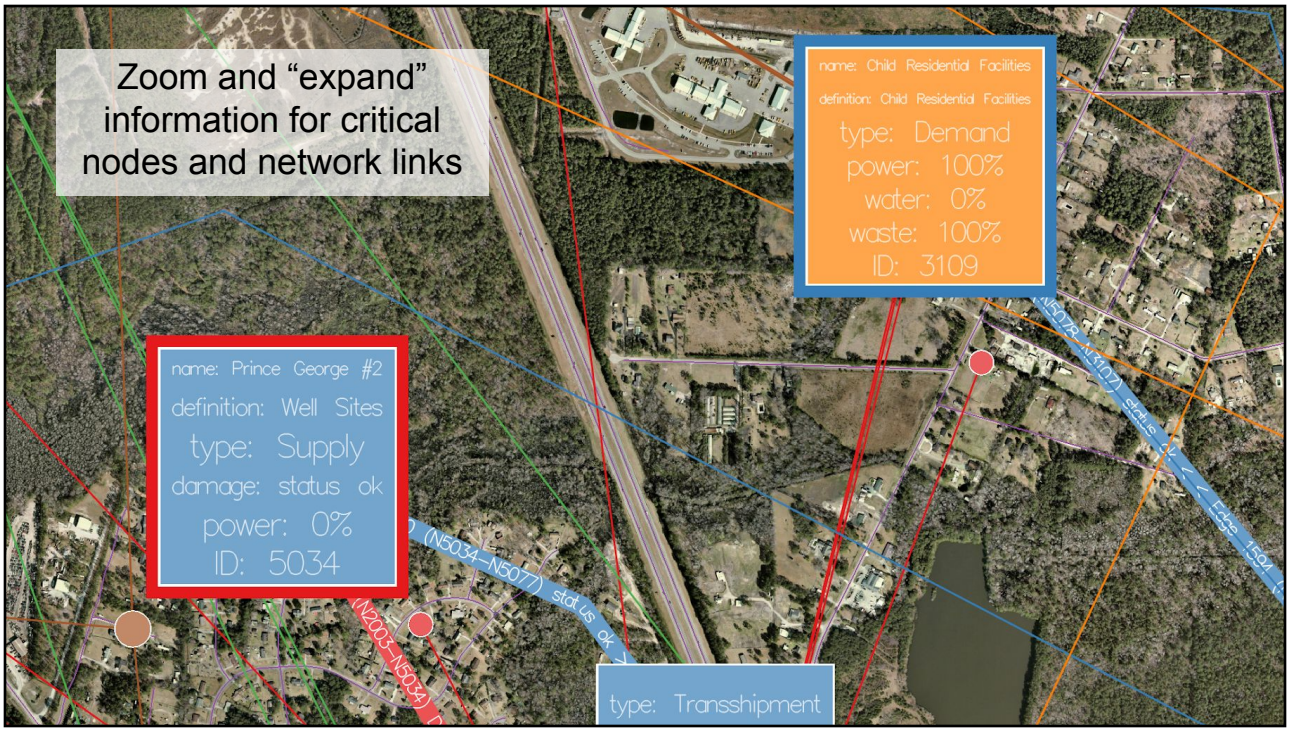
Emergency Response Decision Making

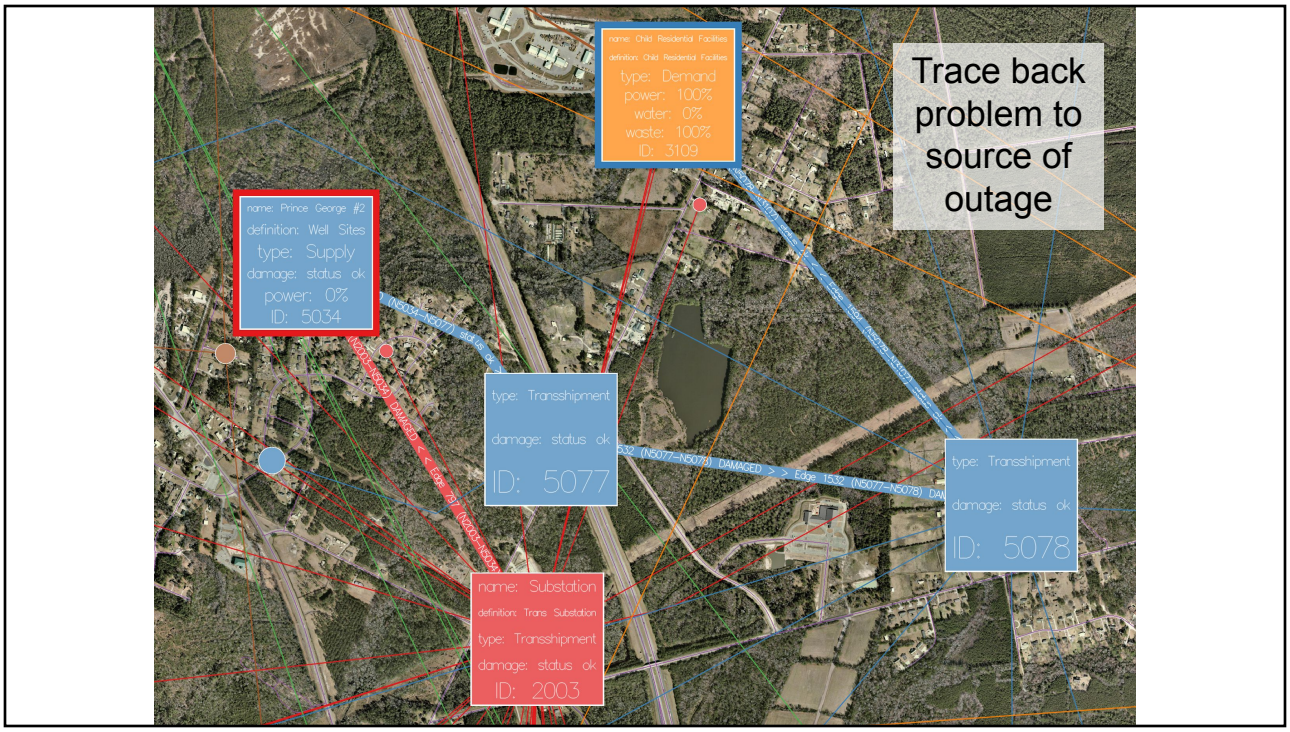


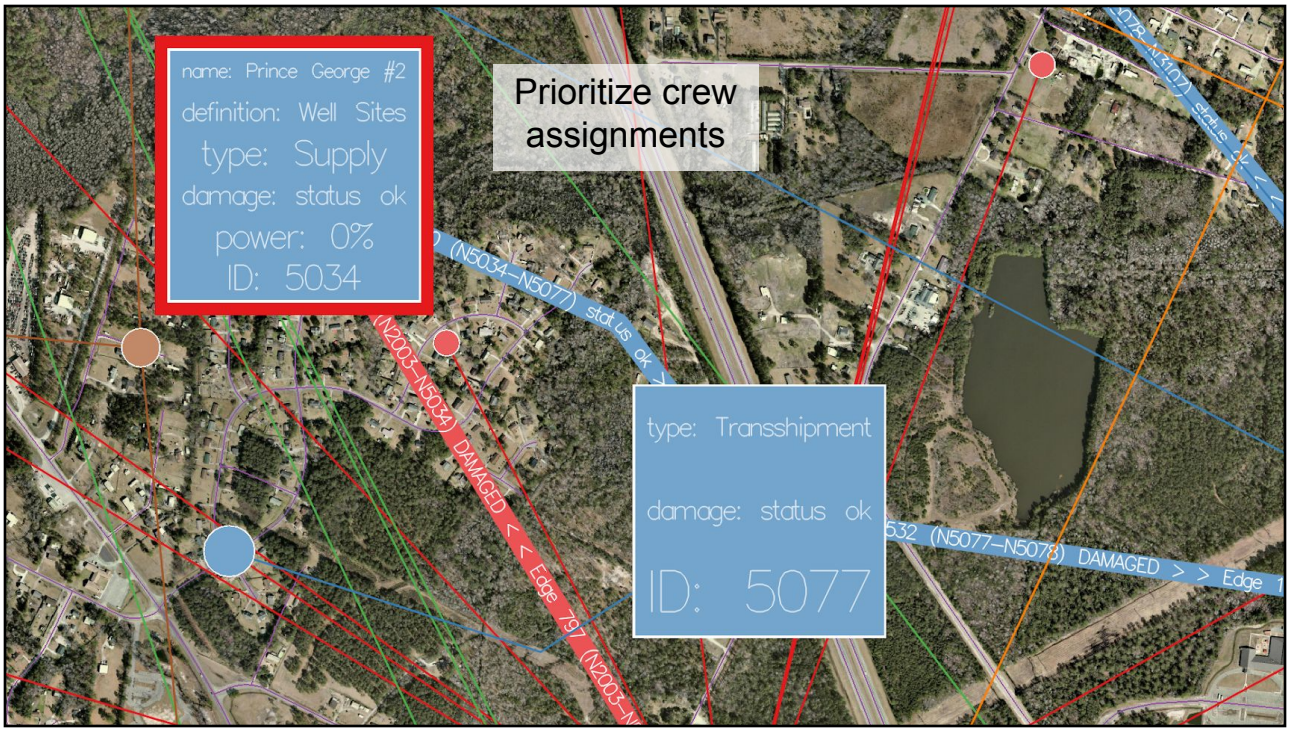
Emergency Response Decision Making

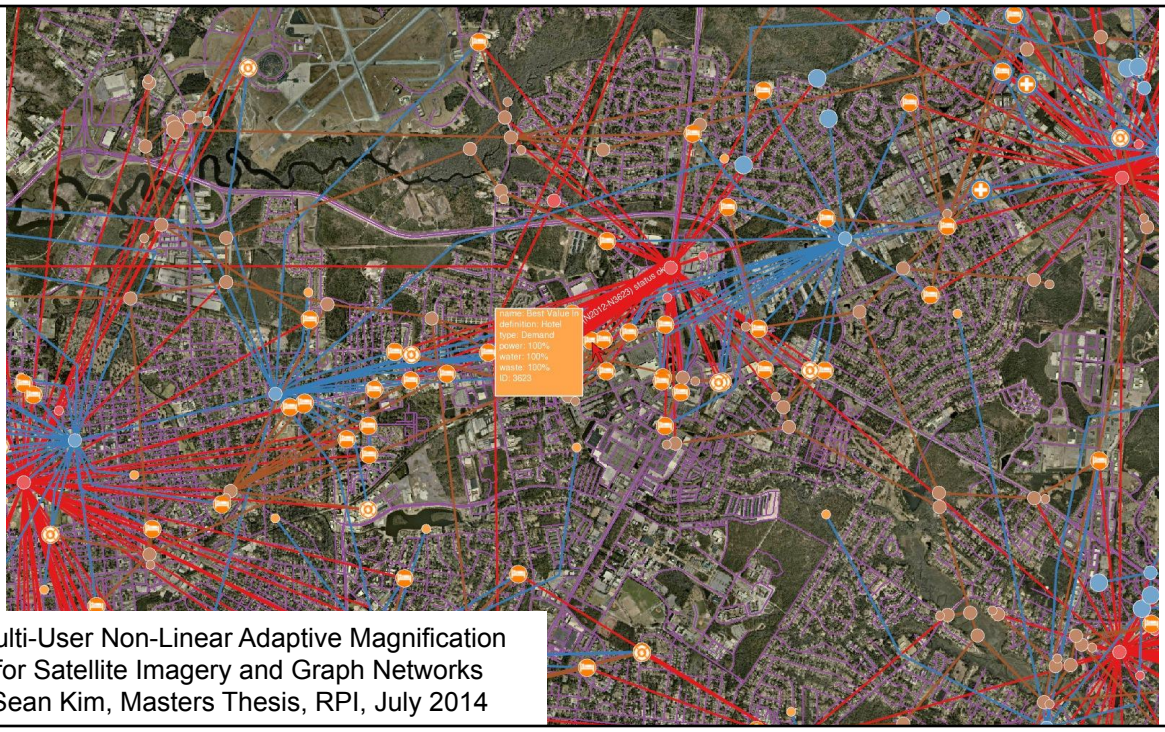

Full network detail
is overwhelming


Subset of data

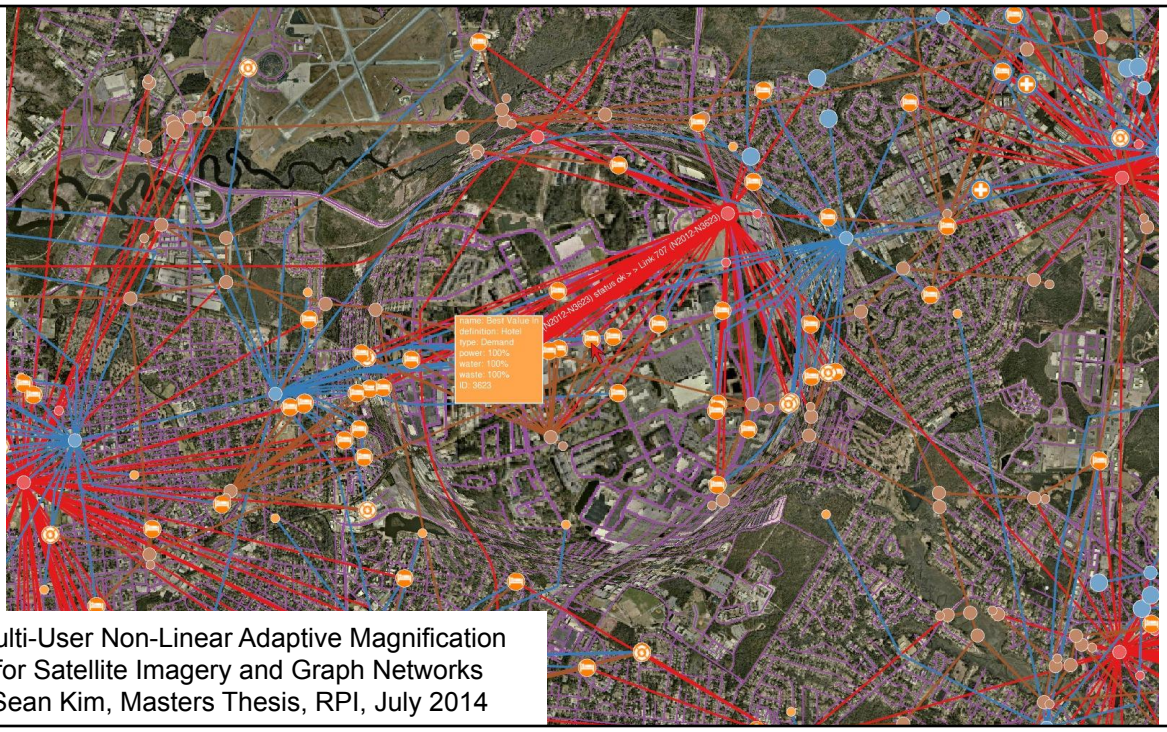

Zoom and “expand” information for critical nodes and network links
name: Child Residential Facilities
definition: Child Residential Facilities
type: Demand
power: 100%
water: 0%
waste: 100%
ID: 3109
name: Prince George #2
definition: Well Sites
type: Supply
damage: status ok
power: 0%
ID: 5034
type: Transshipment


Trace back problem to source of outage
name: Child Residential Facilities
definition: Child Residential Facilities
type: Demand
power: 100%
water: 0%
waste: 100%
ID: 3109
name: Prince George #2
definition: Well Sites
type: Supply
damage: status ok
power: 0%
ID: 5034
type: Transshipment
damage: status ok
ID: 5077
type: Transshipment
damage: status ok
ID: 5078
name: Substation
definition: Trans Substation
type: Transshipment
damage: status ok
ID: 2003

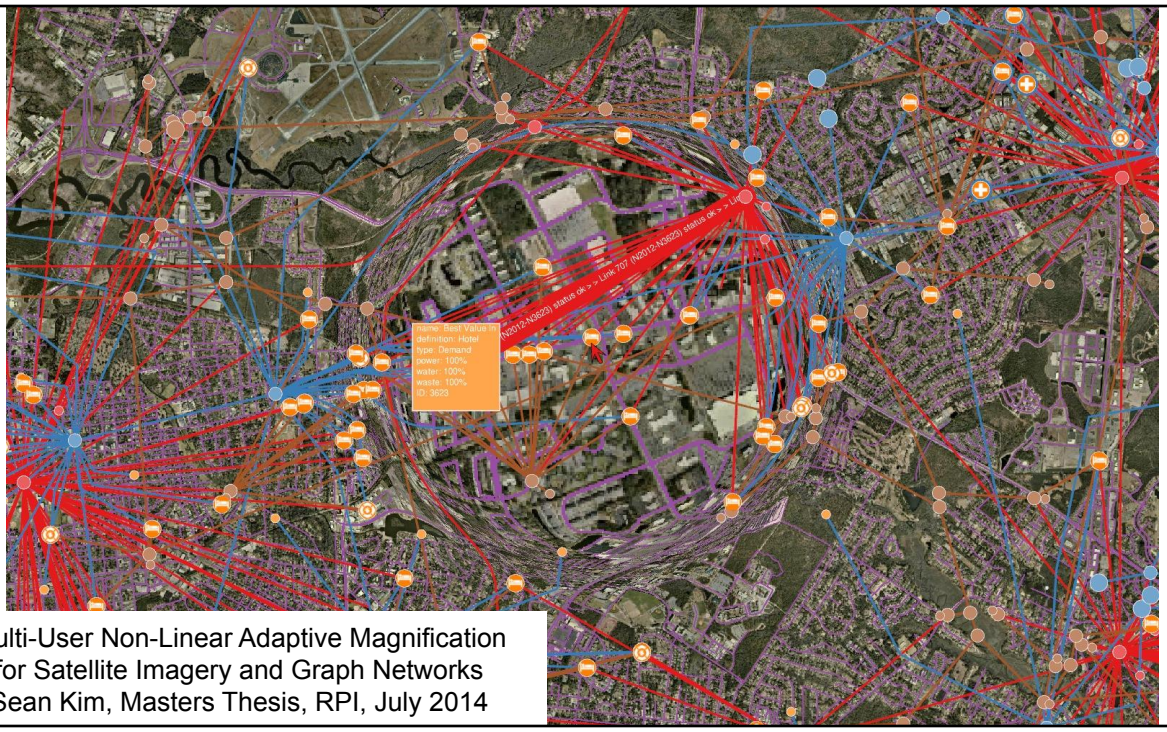




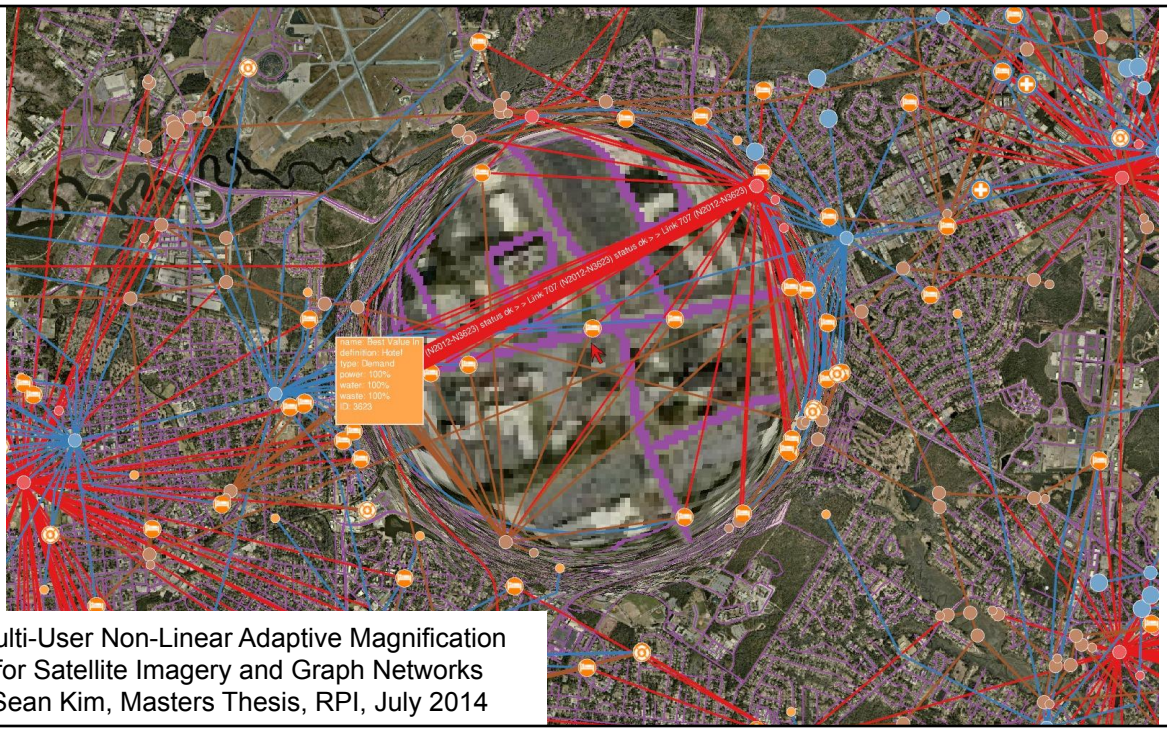
Multi-User Non-Linear Adaptive Magnification
for Satellite Imagery and Graph Networks
Sean Kim, Masters Thesis, RPI, July 2014



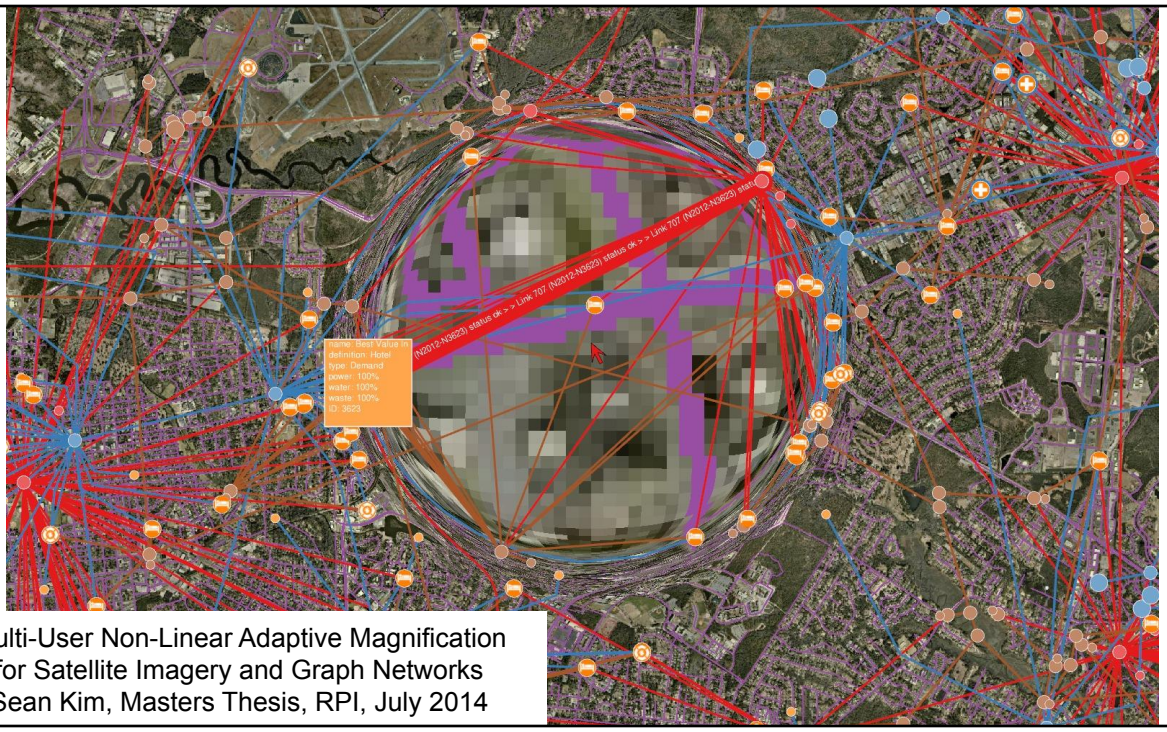
Multi-User Non-Linear Adaptive Magnification
for Satellite Imagery and Graph Networks
Sean Kim, Masters Thesis, RPI, July 2014



Multi-User Non-Linear Adaptive Magnification
for Satellite Imagery and Graph Networks
Sean Kim, Masters Thesis, RPI, July 2014



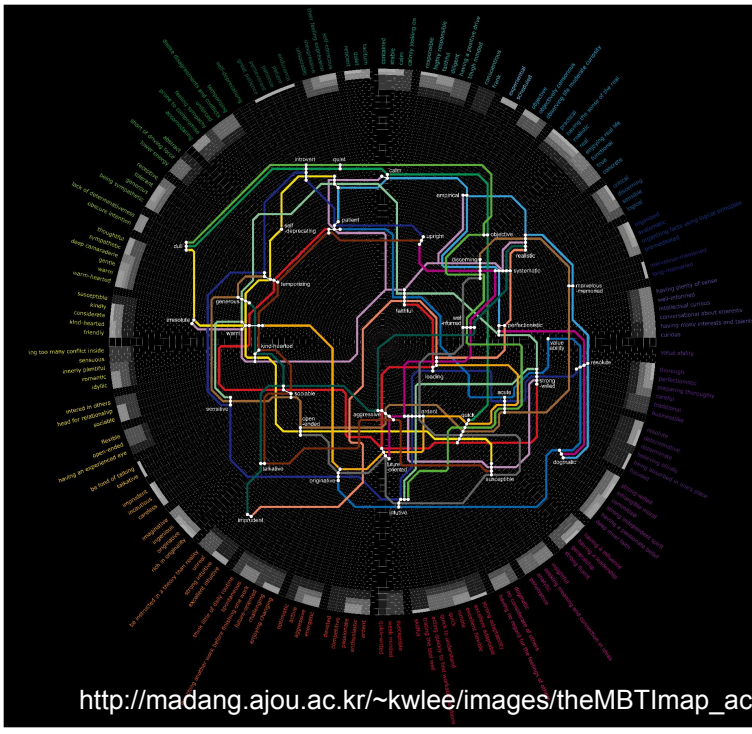
Multi-User Non-Linear Adaptive Magnification
for Satellite Imagery and Graph Networks
Sean Kim, Masters Thesis, RPI, July 2014



Multi-User Non-Linear Adaptive Magnification
for Satellite Imagery and Graph Networks
Sean Kim, Masters Thesis, RPI, July 2014

Today

- Worksheet & HW 3: Graphing RPI CSCI Course & Instructor Data
- Reading: Clustering/Hierarchical Edge Bundling
- Reading: Lombardi Graphs
- Emergency Management Graph Visualization
  - Sean Kim's masters project
- Preview of "High-Dimensional" Data Visualization
- Readings for Tuesday

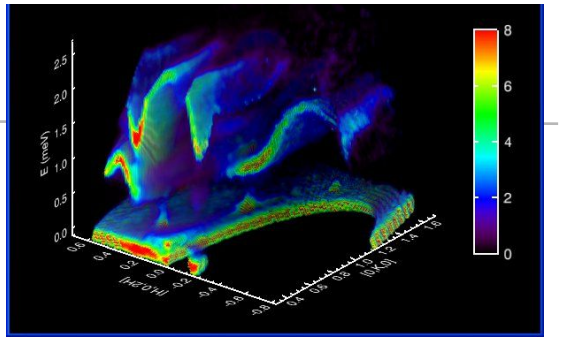


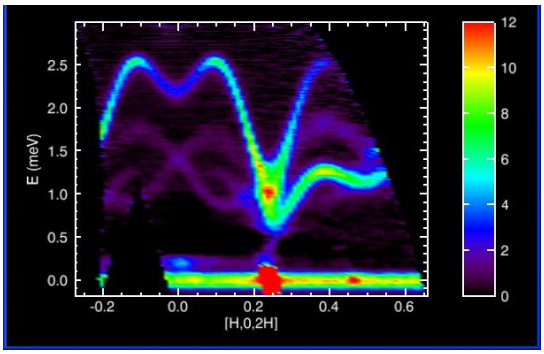
- Many datasets don't (initially) have coordinate systems!
- This graph shows the relationship among the words that describe people's personalities.

http://madang.ajou.ac.kr/~kwlee/images/theMBTImap_actual%20poster%20image.pdf

Scientific Data

- For many 3D(or 2D) spatial locations during an experiment or simulation…
- Collect time-varying temperature, velocity, pressure, humidity, etc.

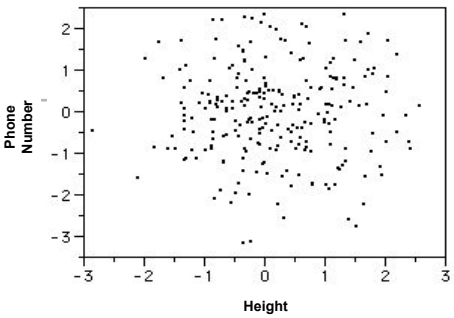


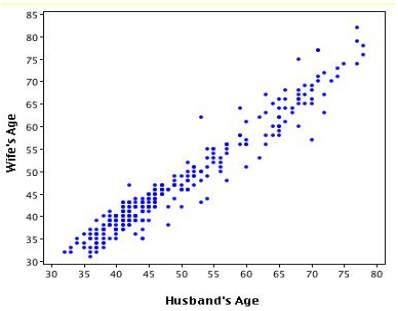


http://www.ncnr.nist.gov/dave/screenshots.html

Miscellaneous Personal Data

- Height, weight, eye color, phone number, address, IQ, age, cholesterol score, grade in Data Structures, etc.
- Example Hypotheses:
  - There is no correlation between height & phone number
  - There is a positive correlation between the ages of spouses
- Scatter plot: Look at 2 dimensions at a time
- *Scale/units on each axis are often very different!*

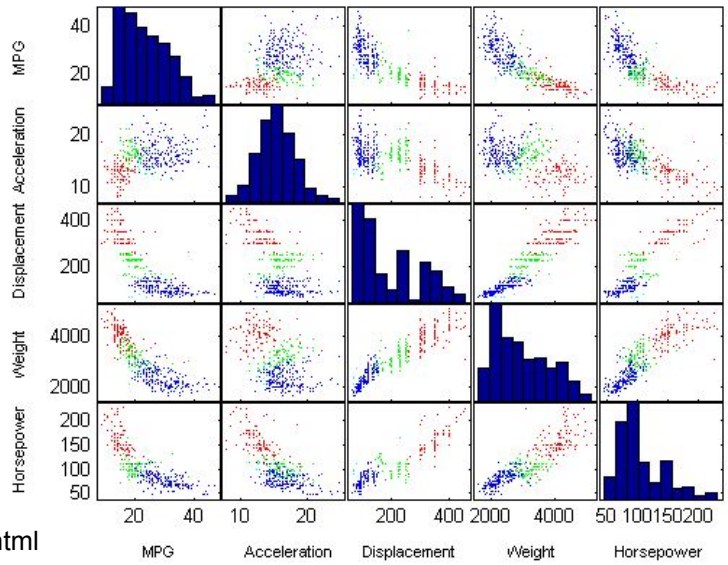




http://www.mzandee.net/~zandee/statistiek/stat-online/chapter4/pearson.html

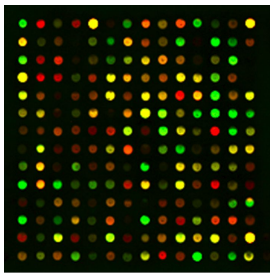
Scatterplot Matrix

Color coded by
# of cylinders

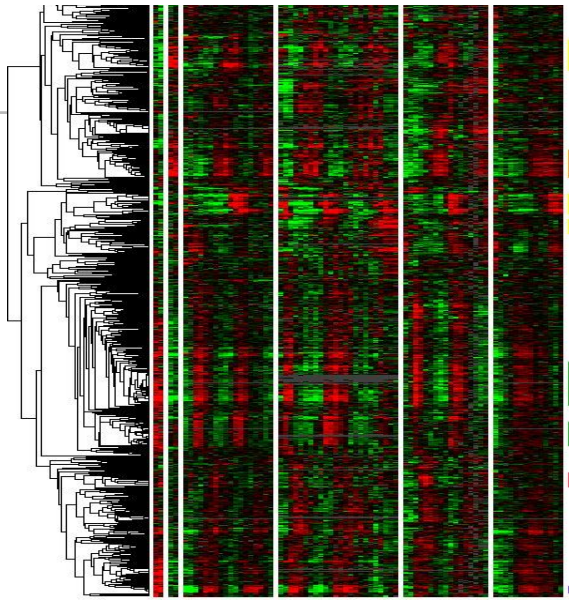


http://www.mathworks.com/help/stats/examples/visualizing-multivariate-data.html

Gene Expression



- Expression level for hundreds of genes
- For many trials, different individuals/conditions)
- Automatically discover correlations: genes that commonly work together or in opposition



http://www.imbb.forth.gr/people/poirazi/researchEP.html

http://www.bioss.ac.uk/~dirk/essays/GeneExpression/bayes_net.html

>3 Dimensions vs. “Really High Dimensions”

- Obvious/intuitive dimensions
  - Position, Orientation, Time, Temperature, Color, etc.

vs.

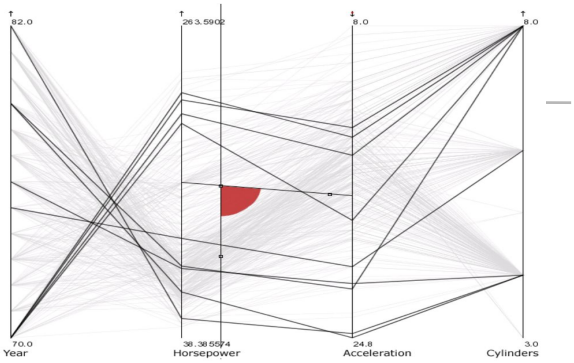
- Hundreds or thousands of attributes
  - May be floating point values or binary values or even *unordered categorical values*
  - Stored as a “feature vectors” for each data point
  - Nearest Neighbor calculations become very expensive & visualization seems *impossible*

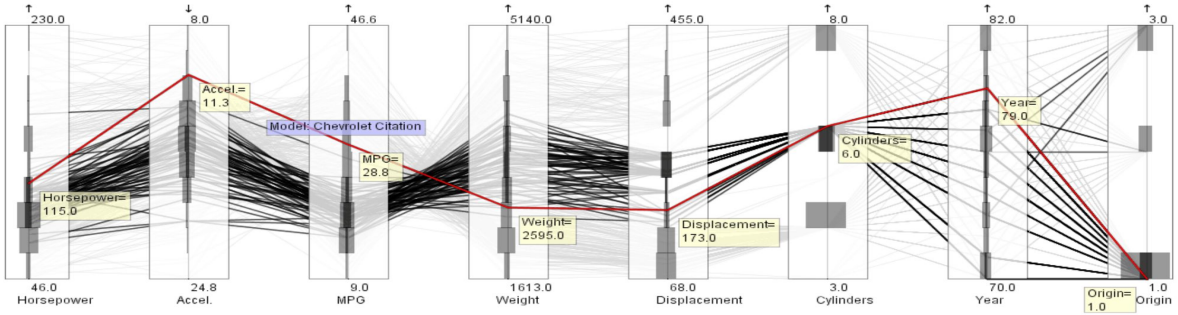
Today

- Worksheet & HW 3: Graphing RPI CSCI Course & Instructor Data
- Reading: Clustering/Hierarchical Edge Bundling
- Reading: Lombardi Graphs
- Emergency Management Graph Visualization
  - Sean Kim’s masters project
- Preview of “High-Dimensional” Data Visualization
- Readings for Tuesday

Reading for Tuesday

"Angular Brushing of Extended Parallel Coordinates", Hauser, Ledermann, and Doleisch, InfoVis 2002





Reading for Tuesday

“LineUp: Visual Analysis of Multi-Attribute Rankings”, Gratzl, Lex, Gehlenborg, Pfister and Streit, IEEE INFOVIS 2013