Overview

1. Designing effective visualizations

2. Visualization design basics
   2.1 Requirement analysis
   2.2 Visual perception and cognitive processing
   2.3 Visual mappings (slide 51)
   2.4 Interaction design

3. Visualization of
   3.1 Small networks
      - Hands-on visualization design
   3.2 Medium size networks
      - Hands-on visualization design
   3.3 Large networks

4. Challenges and opportunities

Schedule
5/17, 15:30-16:30
5/18, 15:30-16:30
5/19, 17:00-20:00
Preview: Hands-on visualization design

Datasets & tasks
(I) Network of highly frequent and bursty words in PNAS
   http://iv.slis.indiana.edu/ln/data/pajek_newinput.txt
(II) Network of papers that follow up on Milgram’s 76 work.
    http://ella.slis.indiana.edu/~katy/outgoing/histcite/milgrams.txt

Programs
- **IVC** and **Pajek** for word occurrence network
- **HistCite** for paper citation network
- **Pajek** and **VxInsight** for research collaboration network

1. Designing effective visualizations

"The success of a visualization is based on deep knowledge and care about the substance, and the quality, relevance and integrity of the content."

(Tufte, 1983)

Principle of Graphical Excellence
- Well-designed presentation of interesting data: substance, statistics, design.
- Complex ideas communicated with clarity, precision, and efficiency.
- Conveying the most knowledge in the shortest time with the least ink in the smallest space.
- It requires telling the truth about the data.
- It is nearly always multivariate.

(Tufte, 1983)
Five Principles in the Theory of Graphic Display

- Above all else show the data.
- Maximize the data-ink ratio, within reason.
- Erase non-data ink, within reason.
- Erase redundant data-ink.
- Revise and edit.

Visualizations should strive towards the following goals

- Focus on content of data not the visualization technique.
- Strive for integrity.
- Utilize classic designs and concepts proven by time.
- Comparative rather than descriptive visualizations.
- High resolution.

(Tufte, 1983)

Aesthetics

- Properly choose format and design
- Use words, numbers, drawings in close proximity
- Use lines of different weights as an attractive and compact way to display data.
- Reflect a balance, a proportion, a sense of relevant scale.
- Display an accessible complexity of detail.
- Let the graphics tell a story about the data.
- Avoid content-free decoration.
- Make use of symmetry to add beauty (although someone once said that "all true beauty requires some degree of asymmetry").
- Draw graphics an a professional manner, with the technical details of production down with care.

(Tufte, 1983)
Labeling

- Words spelled out.
- Words run left to right.
- Little messages explain data.
- Labels on the graphic; no legend needed.
- Graphic provokes curiosity.
- Blue contrasted with others.
- Clear, precise, modest type.
- Type is mixed case, with serifs

(Tufte, 1983)
The Internet

Donna Cox & Robert Patterson, NCSA
Visualization Study of the NSFNET, 1992

Social Networks / Activity

Figure 1: Friendship Choices Among Fourth Graders from Monroe, 1934, p. 396

K. Börner et al.

A. S. Klovdahl

K. C. Claffy, CAIDA

Young Hyun, CAIDA
General Philosophy for Increasing Data Comprehension

- High density is good: the human eye/brain can select, filter, edit, group, structure, highlight, focus, blend, outline, cluster, itemize, winnow, sort, abstract, smooth, isolate, idealize, summarize, etc. Give people the data so they can exercise their full powers – don’t limit them.
- Clutter/confusion are failures of design and not complexity.
- Information consists of differences that make a difference: "hide" that data which does not make a difference in what you are trying to depict.
- In showing parallels, only the relevant differences should appear.
- Value and power of parallelism: once you have seen one element all the others are accessible.
- Separate figure and background (for example, a blurry background often brings the foreground into sharper focus), layering & separation, use of white space.
- Graphics should emphasize the horizontal direction.

(Tufte, 1983)
Compare and Contrast

Yook, Jeong, & Barabasi, 2001

Geographic distribution of Internet routers (top) vs. global distribution of population (bottom).

George Chaplin

Map of human skin colors based on global ultraviolet radiation intensity and precipitation levels.

Map Attribute Overlays

http://worldprocessor.com/index_vis.htm
How to achieve excellence?

Designing good visualizations requires a deep understanding of

- Users, their data, and their tasks. HCI
- Available hardware (displays, interaction technology) & software (applicability, scalability, etc.) CG, IV, UID
- Human perception and cognition. Psychology
- (Artistic) Visual representation criteria. Arts
- Usability studies and evaluation. HCI
- Work domain analysis. HCI
2. Visualization design basics

2.1 Requirement analysis – What users really want
2.2 Visual perception and cognitive processing – The only constant in this job
2.3 Visual mappings – How to visually encode what information
2.4 Interaction design – How do show all data using a limited number of pixels

2.1 Requirement analysis

First, you need to define your users and their tasks. Then you can select appropriate datasets, algorithms, visualization metaphors.

Acquire info on users and their tasks using
- Interviews
- Questionnaires
- Observation
- Document analysis
Defining your users and their tasks

- Who are your users (profession, location, gender, age, or lifestyle preferences)? There might be different user groups.
- What is their level of technical & subject expertise? Visual language used has to match the user's understanding of its function and/or content.
- Do users have information preferences? Which pieces of information do users want first, second, third, and so on?
- What are the user's information needs/tasks? Describe scenarios of use, or those situations or circumstances under which the IV may be used.
- What is their conceptualization of their tasks?
- What do your users need to understand, discover and communicate, etc. and at what point in time?
- When do they get promoted/fired?

- What is the visualization context? Describe your users' physical and social environments. Note any environmental challenges such as poor lighting or noise, and any technical challenges such as screen size and number of colors. Determine what hardware and browser software, monitors and screen resolutions your audience uses. In which work context will the IV be used?

Task analysis

Existing System

- What problems exist with the current interfaces (tasks, information objects, actions)?
- Prepare scenarios for each problem (use user terms not machine terms).
- Which problems are significant? Rank problems.
- How do users judge the result of their work?
- How is the system embedded in the work context?

New System

- Research how the potential users currently do their work.
- Determine a set of goals belonging to the target user through observation and interviews.
- Determine set of tasks that support these goals. Draw a diagram of the workflow.
- Prioritize tasks based on criteria such as the importance of the goal to the organization and the frequency of task performance.
General tasks

Should visualization help to identify
- Trends in the data.
- Outliers.
- Jumps in the data (gaps).
- Maxima and minima like largest, smallest, most recent, oldest, etc.
- Boundaries (not the same as maxima or jumps).
- Clusters in the data.
- Structure in heterogeneous information.
- A particular item of interest within the context of an enormous amount of contextual data.

Each of these tasks requires a different visualization design!

In the context of network visualizations

- What data entities should be represented as nodes?
- What nodes are important?
- What relationships are important and should be represented as edges?
- What node/edge attributes are important and need to be encoded?
- What subset of nodes, edges, subgraphs need to be labeled and how?
- Is there a temporal, geospatial, or semantic substrate that should augment the layout of nodes?
- Are there any existing metaphors that can guide the visual encoding of nodes, edges, and their attributes?
- How large is the network? What data can be omitted to provide users with a meaningful overview of the dataset?
2.2 Visual perception and cognitive processing

Our visual perception and cognitive processing (and body) are optimized for:
- Finding food and prey in natural habitats.
- Mating.
- Avoiding and/or escaping predators.

It is not optimized for extended keyboard, mouse, and screen usage.

---

Human visual information processing

**Stage 1:**
Parallel processing to extract low level properties of the visual scene.

- Rapid parallel processing.
- Extractions of features, orientation, color, texture, and movement patterns.
- Transitory nature of information, which is briefly held in an iconic store.
- Bottom-up, data-driven model of processing.
Stage 2:
Sequential goal-directed processing.

- Slow serial processing.
- Involvement of both working memory and long-term memory.
- More emphasis on arbitrary aspects of symbols.
- Top-down processing.
- Different pathways for object recognition and visually guided motion.
Receptors

- 100 million rods - extremely sensitive to low light levels.
- 6 million cones - sensitive under normal working levels.

- Nerve cells situated early in the visual pathway do not respond to absolute signals but space and time differences.
- Later stages take into account position of light, shadows, orientation of an object.

Figure 1.3.9 Distribution of rods (solid curve) and cones (shaded region) in the human retina. Notice that the fovea is populated almost exclusively by cones and that rods are much more plentiful than cones in the periphery.

Electromagnetic Spectrum

http://www2.ncsu.edu/unity/lockers/project/wiebeproj/webdes/color_models
5-10% of people are color blind to some degree (red-green is the most common type followed by blue-yellow, which usually includes blue-green). Mostly men (10%) are color blind. Female 1%.

Many people never find out that they are color blind.

**Color**

- Color does not help to determine the layout of objects in space, how they are moving, or what their shapes are.

**Color Tests**

The individual with normal color vision will see a 5 revealed in the dot pattern. An individual with Red/Green (the most common) color blindness will see a 2 revealed in the dots.

http://www.visibone.com/colorblind/
Color does help

- Break camouflage.
- It tells us about material properties (condition of food, tools).
- Layering with colors is often effective.
- Color grids are a form of layer which provides context but which should be unobtrusive and muted.
- Pure bright colors should be reserved for small highlight areas and almost never used as backgrounds.
- Colors can be used as labels, as measures, and to imitate reality (e.g., blue lakes in maps).
- Note that surrounding colors can make two different colors look alike, and two similar colors look very different.
- Subtle shades of color or gray scale are best if they are delimited with fine contour lines.
Hue, Saturation, and Brightness

Hue (tint) is most closely related to the wavelength of the stimulus.

Saturation is related to how much white content is in the stimulus. Monochromatic hues are very highly saturated. The least saturated color is white.

Brightness (luminance) relates to the amount of light coming from a source or being reflected from an object.

(Source: http://www.yorku.ca/eye/glossary.htm)

Cross cultural color naming

Colors like red, green, yellow, and blue are most valuable in coding data.

People are excellent in generating pure yellow (2nm accuracy). Most people see unique green at 514 nm, a third sees it at 525 nm.

Unique hues do not change with different overall luminance levels. This supports that chromatic perception and luminance perception are independent.
Chromatic Abberation

- Different wavelengths of light are focused at different distances within the eye. Short-wavelength blue is refracted more than long-wavelength red light.
- Blue text on black background can be almost unreadable if there is white or red text nearby to attract focusing mechanism.

Demo

- Chromostereopsis - for most people red advances, blue recedes. 30% see reverse, 10% see both colors in same plane.
**Edge enhancement**

**Cornsweet effect** (Cornsweet, 1970)

Two areas that have the same lightness can be made to look different by having an edge between them that shades of gradually the two sides.

Technique can be used to adjust the background of a sophisticated data visualization to enhance key parts of the image.

---

*Georges Seurat. Bathing at Asnieres, 1883-84*
Color opponent theory

- Cone signals are transformed by neural mechanisms into three color opponent channels: a luminance black-white and two color red-green, and yellow-blue channels.

- Luminance channel dominates our perception of shape, space, and motion. Good luminance contrast in image is essential to perceive shape, space, and motion, especially when interpreting detailed information, small text, etc.

- Chromatic channels have much lower spatial resolution - tell us about surfaces of objects.

Think of color as attribute not as primary characteristic.

- Color is excellent for labeling and categorization. (However, only small number of colors can be used effectively.

- Simultaneous contrast with background colors can dramatically alter color appearance, making color look like another.

- Never use just chromatic differences. Always have a significant luminance difference in addition to color difference.

- Poor for displaying shape, detail or space.

- High saturation colors should be used for labeling small objects.

- Low saturation colors should be used for labeling large areas.
Iconic buffer

Is a short lived visual buffer that holds images for a second or two while we read symbols into short term memory.

Capacity of 7 items because of image decay, reading rate from iconic buffer, capacity of short-term working memory.

**Experiment:** How many of the objects can you remember after a glimpse of one-tenth of a second long?
Pre-attentive processing

Tasks can be performed on large multi-element displays in less than 200-250 msec. (Eye movements take at least 200 msec to initiate)

Count the 4’s in the table.

| 12342564738290298376534267389 |
| 48732652431452637849409837362 |
| 52435747584737236254647844894 |
| 12342564738290298376534267389 |
| 48732652431452637849409837362 |
| 52435747584737236254647844894 |

Lightness is pre-attentively processed.

As well as …

<table>
<thead>
<tr>
<th>Form</th>
<th>Color</th>
<th>Spatial position</th>
</tr>
</thead>
<tbody>
<tr>
<td>Line orientation</td>
<td>Hue</td>
<td>2D position</td>
</tr>
<tr>
<td>Line length</td>
<td>Intensity</td>
<td>Stereoscopic depth</td>
</tr>
<tr>
<td>Line width</td>
<td></td>
<td>Convex/concave shape from shading</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Curvature</td>
<td>- Flicker/blinking</td>
<td></td>
</tr>
<tr>
<td>Spatial grouping</td>
<td>- Detection of motion</td>
<td></td>
</tr>
<tr>
<td>Size</td>
<td>Added marks</td>
<td></td>
</tr>
<tr>
<td>Line collinearity</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Numerosity</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Network Visualization, Katy Borner, Indiana University

43
Most differences are pre-attentively processed.

Only juncture and parallelism are not.

Properties of sensory and arbitrary representations

**Sensory representations**
- Understanding without training (visual system is build to perceive shapes of 3D surfaces).
- Resistance to instructional bias (illusions).
- Sensory immediacy - hard wired and fast info processing.
- Cross cultural validity.

Example: Segmentation of the visual world.

*Figure 1.8:* Five regions of texture. Some are easier to visually distinguish than others. Adapted from Beck (1966).
Arbitrary representations

- Arbitrary conventional representations.
- Hard to learn (alphabet, to read and to write).
- Easy to forget.
- Embedded in culture and applications (green=dead, red=luck/good fortune in China).
- Formally powerful (mathematic notations).
- Capable of rapid change.

Eye movements

- Saccadic movements: Eye makes 2-5 jerky movements, called saccades, per second. Dwell period is generally between 200 and 600 msec. Saccade takes 20-100 msec. Peak velocity of a saccade can be 900 deg/sec (Hallett, 1986; Barfield et al., 1995).

- Smooth-pursuit movements: Track moving object in visual field.

- Convergent movements: Saccadic or smooth eye movements when objects move away or towards us.

Be aware of saccadic suppression: Certain kinds of events can be missed when they occur while eyes are moving.
Gestalt School of Psychology was founded in 1912 by Max Westheimer, Kert Koffka, and Wolfgang Köhler to investigate the way we perceive form.

Gestalt = pattern in German

Priming effects

… occur even if information is not consciously perceived.

(Example: smiling face proceeds ambiguous face)

Might be useful in helping people search for particular patterns in data.

Prior exposure increases retrieval accuracy -> Provide sample images of the kind of pattern being looked for. Repeat samples at frequent intervals during search process.

(Example: Collecting mushrooms)
Remember

- Brain is powerful pattern matching engine.
- Structures, groups, trends can be discovered among hundreds of data values.
- Brain is especially good in discovering linear features and distinct objects.
- Gestalt laws of pattern perception important for visual design.
- Patterns of moving points can be perceived easily and rapidly.
- Effective use of motion is suggested as fertile area for investigation.

2.3 Visual mappings

- Visual layers
- Visual grammar of diagram/map elements.
- Images and words.
- The importance of reference systems.
- Using appropriate metaphors.

What helps people to understand and navigate abstract, alien information spaces?
Visual layers

Visual grammar of diagram elements

<table>
<thead>
<tr>
<th>Graphical Code</th>
<th>Visual Instantiation</th>
<th>Semantics</th>
</tr>
</thead>
<tbody>
<tr>
<td>2. Shape of closed region.</td>
<td></td>
<td>Entity type.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Larger = more.</td>
</tr>
<tr>
<td>5. Partitioning lines within enclosed region.</td>
<td></td>
<td>Entity partitions are created, e.g., TreeMaps.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Part of relations.</td>
</tr>
<tr>
<td>7. Shapes enclosed by contour.</td>
<td></td>
<td>Contained entities.</td>
</tr>
</tbody>
</table>
7. Shapes enclosed by contour. Contained entities.
10. Linking-line quality. Type of relationship between entities.
12. Tab connector. A fit between components.

Visual grammar of map elements

<table>
<thead>
<tr>
<th>Graphical Code</th>
<th>Visual Instantiation</th>
<th>Semantics</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Closed contour.</td>
<td><img src="image" alt="Closed contour" /></td>
<td>Geographic region.</td>
</tr>
<tr>
<td>2. Colored region.</td>
<td><img src="image" alt="Colored region" /></td>
<td>Geographic region.</td>
</tr>
<tr>
<td>3. Textured region.</td>
<td><img src="image" alt="Textured region" /></td>
<td>Geographic region.</td>
</tr>
<tr>
<td>4. Line.</td>
<td><img src="image" alt="Line" /></td>
<td>Linear map features such as rivers, roads, etc. Depends on scale.</td>
</tr>
<tr>
<td>5. Dot.</td>
<td><img src="image" alt="Dot" /></td>
<td>Point features such as town, building. Depends on scale.</td>
</tr>
<tr>
<td>6. Dot on line.</td>
<td><img src="image" alt="Dot on line" /></td>
<td>Point feature such as town on linear feature such as road.</td>
</tr>
</tbody>
</table>
Images and words

- Words (mathematical symbols, natural language, music) are better for representing procedural information, logical conditions, abstract verbal concepts (freedom).

- Images (graphics, abstract & figurative imagery) are better for spatial structures, location, detail.

- Animation brings graphics closer to words in expressive capacity (pushing, causality, 'Sorting out Sorting', disassembly).
Images and words can be linked via
- Proximity
- Continuity/connectedness
- Common region
- Combinations thereof

Rules of thumb to integrate words and images:
- In written text - give text first then link to image.
- Highlight relevant part of info just before the start of relevant speech segment.
- Move viewpoint in visualization to draw attention to different features.

Cinematography: Static scenes 'go dead' visually after a few glances.

Cognitive map development or the importance of landmarks

<table>
<thead>
<tr>
<th>Lynch's Types</th>
<th>Examples</th>
<th>Functions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Paths</td>
<td>Street, canal, transit line</td>
<td>Channel for navigator movement</td>
</tr>
<tr>
<td>Edges</td>
<td>Fence, riverbank</td>
<td>Indicates district limits</td>
</tr>
<tr>
<td>Districts</td>
<td>Neighborhood</td>
<td>Reference region</td>
</tr>
<tr>
<td>Nodes</td>
<td>Town square, public building</td>
<td>Focal point for travel</td>
</tr>
<tr>
<td>Landmarks</td>
<td>Statue</td>
<td>Reference point into which we cannot enter</td>
</tr>
</tbody>
</table>

Figure 10.6 The functions of different kinds of landmarks in a virtual environment. Adapted from Vinson (1999).
Landmark knowledge is the basis of a cognitive map.

Navigational/wayfinding knowledge starts with highly salient visual landmarks. Locations in the environment are associated with reference to these landmarks.

Route knowledge uses landmark knowledge to make decisions when to take turns. It does not provide info on how to optimize a route.

Survey knowledge completes a cognitive map. Intensive use of maps increases survey knowledge in relatively short time (Lokuge et al., 1996).

Three major paradigms (Dourish and Chalmers, 1994)

- Spatial navigation: mimics our experiences in physical world
- Semantic navigation: driven by semantic relationships or underlying logic.
- Social navigation: takes advantage of the behavior of like-minded people.

Navigation/Wayfinding aids

- Provide overview maps if space is large
  - Track-up display (fewer errors) (Eley, 1988)
  - North-up display - provides common frame of reference for communication
- Indicate user location and direction of view in map.
- Provide imagery of key landmarks.
- Verbal or written set of procedural instructions enhanced with landmark imagery can be more useful than map.
- Sparsely populate with discrete but separately identifiable objects—there must be enough landmarks that several are always visible at any instant.
Implications for map design

- Provide overview maps if space is large
  - Track-up display (fewer errors) (Eley, 1988)
  - North-up display - provides common frame of reference for communication
- Indicate user location and direction of view in map.
- Provide imagery of key landmarks.
- Verbal or written set of procedural instructions enhanced with landmark imagery can be more useful than map.
- Sparsely populate with discrete but separately identifiable objects—there must be enough landmarks that several are always visible at any instant.
- Strong visual cues indicating paths and regions can help users to understand structure of a space (Darken & Sibert, 1996). Borders, boundaries and gridlines significantly improves navigation performance (Darken & Silbert, CHI 96).

- People like to point and talk when discussing maps. (Oviatt et al., 1997) – Provide opportunities to talk.

Reference systems & base maps

[Links to reference systems and base maps]

[Images of reference systems and base maps]
Base map distortion

![2004 Presidential Election Results by County](http://www-personal.umich.edu/~mejn/election)

Gastner, Shalizi & Newman, 2004

[http://www-personal.umich.edu/~mejn/election](http://www-personal.umich.edu/~mejn/election)

Metaphors that help ‘spatialize’ abstract spaces

[http://www.rhizome.org/starrynight/](http://www.rhizome.org/starrynight/)

[Xiong & Donath’s 1999, PeopleGarden](http://www.rhizome.org/starrynight/)

[http://www.aureka.com/0/aureka_online.html](http://www.aureka.com/0/aureka_online.html)
Composing the final visualization

1. (Distorted) Base map
2. Information overlay
3. Labels
4. Legend design
   - Title
   - Short explanation of unique features (if space permits)
   - All visual encodings (i.e., what do nodes, edges, colors, etc. represent?)
   - Credibility: Name of map maker, information on dataset, dataset preparation, date.

2.4 Interaction design

Principles of interaction design
- Mapping between data and their visual representation should be fluid and dynamic. -> Principle of transparency - 'the tools itself disappear' (Rutkowski, 1982).
- User obtains illusion of direct control.
- Provide visual feedback within 1/10 seconds (Shneiderman, 1987).
- Object constancy - use animation between displays instead of jumps.

Figure 1: Balancing challenges and skills is key to flow
Ways to decrease information density

Generalization of Line Features
Change of line features using elementary geometric operations


But: Don't Lie With Maps

Summary

- Bandwidth from computer to human is much higher than other way round.
- Design cognitive support systems that are semiautomatic, requiring occasional steering from users in desired direction.
- Use high bandwidth visualization channel to deliver results.

Attention is captured by motion!

Curiosity is an important aspect that can be supported in an interface that allows for discoverability of function. This is often suppressed in systems because of the fear of irrevocably screwing up. Systems should implement universal undo.

(Source: Robertson, Info Vis 1998)
3. Visualization of Networks

3.0 Networks and their representations

3.1 Small networks
3.2 Medium size networks
3.3 Large networks

General visualization objectives:
- Representing structural information & content information
- Efficient space utilization
- Easy comprehension
- Esthetics
- Support of interactive exploration
Aesthetic Criteria for Graph Drawing

- Symmetric.
- Evenly distributed nodes.
- Uniform edge lengths.
- Minimized edge crossings.
- Orthogonal drawings.
- Minimize area / bends / slopes / angles

Optimization criteria may be relaxed to speed up layout process.

(Source: Fruchterman & R. alg p. 76, see Table & discussion Hearst, p 88)

General network representations

Matrices

<table>
<thead>
<tr>
<th></th>
<th>0</th>
<th>0</th>
<th>0</th>
<th>0</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>250.5</td>
<td>0</td>
<td>-280</td>
<td>31.32</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>12</td>
</tr>
</tbody>
</table>

Structure Plots

Equivalenced representation of US power network

Lists of nodes & links

*Netzer 3
1 Dot'080000:0: Green be Brown
2 Dot'080000:0: Green be Brown
3 Dot'080000:0: Green be Brown
*Arcs
1  1 3 c Green
2  3 6 c Black
*Edges
1  3 4 c Green

Network layouts of nodes and links

When to use what kind of representation?
3.1 Small networks

- Up to 100 nodes
- All nodes and edges and most of their attributes can be shown.

**General mappings for nodes**
- # -> (area) size
- Intensity (secondary value) -> color
- Type -> shape

**General mappings for edges**
- # -> thickness
- Intensity, age, etc. -> color
- Type -> style
Network Visualization, Katy Börner, Indiana University

**Tree layouts**

![Tree layout images]

Network Visualization, Katy Börner, Indiana University

**Hands-on visualization design**

**Datasets & tasks**

(I) Network of highly frequent and bursty *words* in PNAS
   [http://iv.slis.indiana.edu/lm/data/pajek_newinput.txt](http://iv.slis.indiana.edu/lm/data/pajek_newinput.txt)

(II) Network of *papers* that follow up on Milgram’s 76 work.
   [http://ella.slis.indiana.edu/~katy/outgoing/histcite/milgrams.txt](http://ella.slis.indiana.edu/~katy/outgoing/histcite/milgrams.txt)

**Programs**

- *Pajek* and *IVC* for word co-occurrence network
- *HistCite* for paper citation network
- *Pajek* and *VxInsight* for research collaboration network
Network Visualization, Katy Börner, Indiana University

7 Colors in Pajek

Available at http://vlado.fmf.uni-lj.si/pub/networks/pajek

Pajek Reference Manual
Using the IVC

- Download from http://ivc.sourceforge.net/ivc.zip
- Double click 'IVC'
- Select Help>Update
- Select 'Search for new features to install', click 'Next'
- Click 'New remote site', enter Name: 'ive', URL: http://ivc.sourceforge.net/update, click 'ok'.
- Open directory
- Check sample data and plugins
Run the IVC Software Framework

Use menu system to
- Load or simulate a data set.
- Analyze, visualize, or interact with a data set.
- To start a tool.
- To access code reference pages, learning modules, javadoc, get updates, etc.

Whenever an algorithm is selected, the user receives feedback on what algorithm was run, what parameters were used, any textual results, who developed this algorithm, etc.

The amount of feedback can be customized.
In addition, a log file is generated as a permanent track of all user actions.

Each loaded/simulated data set is internally stored as a data model. All data models of a session are listed on the right hand side. Right click a model to rename it.
Only algorithms that can be used with a selected data model are selectable (all others are grayed out).*

* The ‘Visualization’ plugins shown in black are prefuse and other demo’s that do not (yet) use the IVC data modules but require their own data format.

---

Currently Available Algorithms

**Data Modeling**
- General network models (Barabasi, Watts, Random)
- P2P network modeling and search algorithms
  - Structured P2P Systems (Content-Addressable Network Model (CAN), Chord Model)
  - Unstructured P2P System (PRU Model, Hypergrid Model)
- TARL model for the co-evolution of author-paper networks

**Data Analysis**
- Search Algorithms (Breadth First Search, k-Random Walk Search, CAN, Chord)
- Timeseries analysis (Burst)
- Network analysis
- Clustering

**Visualization**
- JUNG network layout algorithms
- Prefuse demos

**Interaction**
- Prefuse demos

**Toolkits**
- Active Worlds Toolkit
- Network Analysis Toolkit

**Converters**
Graph <-> Matrix
Dataset

Aims:
➢ Map the co-word space of the top 50 highly frequent and bursty words used in the top 10% most highly cited PNAS publications in 1982-2001.

Our goal here:
Application of burst detection algorithm.
Declutter the visual display of the co-word space.

Name of dataset:
/data/pnas-words.net
With node size coding \sim \text{word burst strength}

Reducing clutter by deleting nodes with low degree: many nodes are missing.

- 82 - 85 = SpringGreen
- 86 - 89 = Goldenrod
- 90 - 93 = Orange
- 94 - 97 = Mahogany
- 98 - 01 = Black
Reducing the number of edges by deleting edges with low weight: many nodes are unconnected.

Reducing the number of edges by applying pathfinder network scaling.
Reducing the number of edges via pathfinder network scaling.

Co-word space of the top 50 highly frequent and bursty words used in the top 10% most highly cited PNAS publications in 1982-2001.

(Mate & Börner, 2004)

3.2 Medium size networks

- Up to 1000 nodes
- Most nodes can be shown but not all their labels.
- Frequently, the number of edges and attributes need to be reduced.

Major design strategies:
Show only important nodes, edges, labels, attributes
Order nodes spatially
Reduce number of displayed nodes

See examples on subsequent slides.
**Hands-on visualization design**

(II) Network of papers that follow up on Milgram 76

**Dataset:**
Papers that cite "Small World Problem" by Milgram S., 1967, PSYCHOL TODAY, or contain "Small World" in the title.

**Aims:**
- Plot the citation network of major papers on the small world concept.
- Identify major (highly cited) papers.
- Understand the temporal dynamics of citation interlinkages.

**Our goal here:**
Plot historiograph of papers using Garfield’s HistCite software.
Interact with and interpret the visualization.

**Name of datasets:**
/data/milgrams.txt (ISI file)
/data/milgrams.mdb
This HistCite software is the property of Eugene Garfield, who also owns the copyright and controls all rights to its use and distribution. Contact Eugene Garfield (garfield@codex.cis.upenn.edu) to use this software.

HistCite Guide:
http://garfield.library.upenn.edu/histcomp/guide.html

Network Visualization, Katy Börner, Indiana University

Co-author network in Pajek
3.3 Large networks

- More than 1000 nodes.
- Neither all nodes nor all edges can be shown at once. Sometimes, there are more nodes than pixels.

Examples of large networks
- Communication networks:
  - Internet, telephone network, wireless network.
- Network applications:
  - The World Wide Web, Email interactions
- Transportation network/road maps
- Relationships between objects in a data base:
  - Function/module dependency graphs
  - Knowledge bases

Challenges
- Positioning nodes without overlap
- De-cluttering links
- Labeling
- Navigation/interaction

Major design strategies:
Tight coupling of data analysis and visualization.

Select most important nodes
- Based on node degree, BC value, connectivity (hubs & authorities), frequency of usage, depth in hierarchy, etc.

Select most important edges
- Based on their BC value, using pathfinder network scaling, flux values, etc. – *show strong and weak links!*
- Show ‘backbone’.

Show major network components
- Represent each component as a ‘super node’ in the size of the # of its nodes.

Careful interactivity design
- Overview first, then zoom and filter, then details on demand.
Finding and evaluating community structure in networks

(a) Largest component of a network of coauthorships between physicists who have published on topics related to networks.
(b) Communities identified using the shortest-path betweenness version of the community structure algorithm.
(c) A coarse-graining of the network in which each community is represented by a single node, with edges representing collaborations between communities. The thickness of the edges is proportional to the number of pairs of collaborators between communities.

(Boyan & Girvan, 2004)
Direct manipulation

Modify focusing parameters while continuously provide visual feedback and update display (fast computer response).

- Conditioning: filtering, setting background variables and displaying foreground parameters.
- Identification: highlight, color, shape
- Linkmap parameter control: line thickness, length, color legend, time slider, animation
- Bird’s Eye view and zooming

Hands-on visualization design

**VxInsight** is a general purpose knowledge visualization software package developed at Sandia National Laboratories.

It enables researchers, analysts, and decision-makers to accelerate their understanding of large databases.
4. Challenges and opportunities

Challenges:
- Utilizing handheld and higher resolution displays
- Visualizing growing networks
- Visualizing network activity
- Visualizing very large scale networks
- Interconnecting network algorithms
  developers and their users.

http://loadrunner.uits.iu.edu/weathermaps/abilene/

Amsterdam RealTime project, WIRED Magazine, Issue
11.03 - March 2003
Logicaland

Logicaland is a project study for visualizing our world’s complex economical, political and social systems. Logicaland tries to engage people into strategies of raising human sensibility and responsibility within the global networked society. The challenge is to develop ideas, tools and visualizations that fit the requirements of complex correlating systems and our world’s complex participative environment.

http://www.logicaland.net
## Opportunities

- Information explosion (amount doubles every 18 months).
- Work is becoming more ‘knowledge-oriented’.
- Increasing computing power (doubles every 18 months - Moore’s Law).
- Decreasing cost of storage.
- Fast graphics processors.
- Larger hard disk sizes -> more information.
- High resolution color monitors.
- Alternative user interfaces Idesk, CAVE.
- Connectivity between systems is expanding.
- Increasing visual intelligence.
- There is a bad mismatch between computer displays and the human perceptual system and between computer controls and human motor functions.

## The power of mapping

- Visualizations are not objective, neutral artifacts.
- They are created. They include or leave out information.
- They communicate particular messages.
- Commonly, the messages are those of the powerful who pay for the visualizations.

**Deconstruct Visualizations!**

by questioning who the visualization was made for, by whom, why, and based on which data!
Further reading

Readings in Information Visualization: Using Vision to Think
Information Visualization: Perception for Design
by Colin Ware, Dec 1999
Information Visualisation and Virtual Environments
by Chaomei Chen, Nov 1999
Information Visualization
   http://www.ee.ic.ac.uk/research/information/www/Bobs.html
Mapping Cyberspace
by Martin Dodge and Rob Kitchin, Nov 2000
http://www.mappingcyberspace.com/
The Craft of Information Visualization: Readings and Reflections
by Benjamin B. Bederson, Ben Shneiderman

Software

See position papers and slides linked from
Workshop on Information Visualization Software Infrastructures
at IEEE InfoVis, Austin, Texas, Oct 9, 2004.
http://www.indiana.edu/ivsi2004/

Description
Information visualization systems and tools are becoming available for a large range of visualization and interaction techniques and are used in diverse application domains. This workshop is aimed at gathering experts interested in building multi-organizational to share their vision, understand the issues involved and trying to find ways to avoid fragmentation and enhance collaboration.
Acknowledgements

I would like to thank Weimao Ke for preparing the datasets used in this workshop material. Shashikant Penumarthy, Bruce Herr, and James Ellis designed and implemented the IVC software framework.

Several images, references, and text fragments presented in these slides were taken from Colin Ware’s book “Information Visualization: Perception for Design”, Laszlo Barabasi’s Image Gallery at http://www.nd.edu/~networks/gallery.htm and Martin Dodge’s Atlas of Cyberspaces at http://www.cybergeography.org/atlas.

Funding for the preparation & presentation of this workshop was provided by the Abdus Salam International Centre for Theoretical Physics (ICTP), NSF DUE-0333623 award and NSF IIS-0238261 Career award.