Perspectives on Network Visualization

Agenda

Introduction  Node-Link  Composite  Alternate  Coordinated  Discussion
My research focuses on information visualization, visual analytics, and cognitive computing. My goal is to design and build interactive visualization systems for users to effectively help them reach their goals.

Specifically I want to make data discovery, cleaning, exploration, and sharing results as fluid and easy and accessible as possible. While at IBM I have been involved in many projects to this end collaborating with teams across the organization, with more coming!

I received a PhD and M.S. in Computer Science under Ben Shneiderman at the University of Maryland Human Computer Interaction Lab, and a B.A. in Computer Science and Mathematics from Cornell College. I contribute to NodeXL and have worked at Microsoft Research and NIST.
Network Visualization in IBM
To name only a few

TED

WDA-LS

WDA

WGA

Alchemy

Twitter

i2

Designers

IOC Epidemics
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
Node-Link Network Visualizations
Layout Algorithms Matter

Immense variation in layout readability and speed

Hachul & Jünger, 2006

Shneiderman B and Dunne C (2012), "Interactive network exploration to derive insights: Filtering, clustering, grouping, and simplification", In Graph Drawing ‘12. pp. 2-18. DOI:10.1007/978-3-642-36763-2_2

Node-Link Network Visualizations: Motif Simplification
Observations

1: There are repeating patterns in networks (motifs)

2: Motifs often dominate the visualization

3: Motifs members can be functionally equivalent
Motif Simplification

Motif Design

Fan Motif

2-Connector Motif
Lostpedia articles
Lostpedia articles
Motif Simplification
Fan Glyph Design
Motif Simplification
Connector Glyph Design
Motif Simplification
Clique Glyph Design
Motif Simplification
Interactivity Video
Motif Simplification
Readable and scalable network visualizations for understanding relationships

- WDA-LS Drug Repurposing
- Concept Insights Biomarker Analysis
- SharpC Medical record concepts
- i2 Heterogeneous Networks
WDA-LS Drug Repurposing

- Conditions explored in client's use case (Asthma, Colitis, and Systemic Lupus Erythematosus (SLE)) seen in topology
- With the motifs we were able to easily find all the diseases with any required connectivity grouping
- Easily see neighboring/related conditions
  - Quickly find all diseases with required connectivity grouping
  - Existing conditions this drug is aimed at, such as Encephalomyelitis and Autoimmune Diseases, and positioned in the same region of the network.
  - If Colitis proves to be a promising target, might also consider exploring the importance of neighbors Liver Diseases, Psoriasis, Hepatitis, ...
- Rather than reduce size, see lossless overview then drill down in focused way
- Optional complement gives users greater control and wider variety of tools to accelerate discovery
- Prototype technique can be extended and refined to address client needs

Concept Insights Biomarker Network

- Visual encoding of network statistics and similarity measures can help identify insights
- Drastic reduction in network complexity through filtering by similarity (esp. >60%)
- Motif membership appears meaningful
- Interactive tools can support mental map while filtering through animations and iterative layout
- Flow of concepts between motifs at progressively higher filtering can be show in Sankey diagram to show strong relationships
- Can speed up clique through pre-processing

CHI Paper

- Controlled experiment with 36 users showed that motif simplification improves user task performance
  - Reducing complexity
  - Understanding larger or hidden relationships
- Algorithms for detecting fans, connectors, and cliques
- Publicly available implementation in NodeXL: nodexl.codeplex.com


Shneiderman B and Dunne C (2012), "Interactive network exploration to derive insights: Filtering, clustering, grouping, and simplification", In Graph Drawing '12. pp. 2-18. DOI:10.1007/978-3-642-36763-2_2
Node-Link Network Visualizations: Showing Group Membership
Disjoint Set Visualization
Network grouping/partitions

- Attributes
- Topology
- Combinations
- Manual

Twitter ties at the 2012 Collective Intelligence Conference @ MIT
Squarified Treemap
(Rodrigues et al., 2011)
Group-in-a-Box Meta-Layouts

Variants

• Squarified Treemap
  (Rodrigues et al., 2011)

• Croissant-Doughnut

• Force-Directed
Force-Directed
Furnas's *Generalized Fisheye Views* (1986)
Card, Robertson, & Mackinlay's *Information Visualizer* (1991)
Ishii & Ullmer's *Tangible Bits* (1997)
Temporal Spatial Attribute

Grapes-in-a-Box
Group-in-a-Box Meta-Layouts
See local and global connections

- Concept Insights Biomarker Analysis
- SharpC Medical record concepts
- Similarity for Active Learning
- Innovation in Pennsylvania
- U.S. Senate Voting Patterns
Group-in-a-Box Meta-Layouts

Discussion

• Three Group-in-a-Box \textbf{layout algorithms} for dissecting networks
  • Improved group and overview visualization
• \textbf{Empirical evaluation} on 309 Twitter networks using readability metrics
• Publicly available \textbf{implementation in NodeXL}: nodexl.codeplex.com


Shneiderman B and Dunne C (2012), "Interactive network exploration to derive insights: Filtering, clustering, grouping, and simplification", In Graph Drawing ’12. pp. 2-18. DOI:10.1007/978-3-642-36763-2_2

Overlapping Set Visualization
ABACUS API – Fortune 500
Composite Network Visualizations

Introduction
Node-Link
Composite
Alternate
Coordinated
Discussion
Composite Network Visualization: VoroGraph
GLEAM Epidemic Model
Population basins, local commuting, and global flights

population layer -- census areas
short range mobility layer -- commuting
long range mobility layer -- air travel
Voronoi Tessellation
Western Europe
Centroidal Voronoi Tessellation – Animated!

Western Europe
**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.
VoroGraph
Non-Contiguous Edge Coding
VoroGraph
Non-Contiguous Edge Coding
VoroGraph
Force-Directed Group-in-a-Box
• 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

Composite Network Visualization: BrainVis
Brain From Back: Back-view of the brain with the network of activity and connections overlaid onto colored brain regions

Brain From Below: Bottom-view of the brain with the network of activity and connections overlaid onto colored brain regions
Brain From Above: Top-view of the brain with the network of activity and connections overlaid onto colored brain regions

Brain From Side: Side-view of the brain with color-coded brain regions and physical connections
Alternate Network Visualizations
Line Charts
Tree visualization citation network aggregation

Matrix/Heatmap Visualizations
Co-occurrence of organizations related to business intelligence

Business Intelligence 2000-2009:

Tech1
• Google
• Yahoo
• Stanford
• Apple

Tech2
• IBM, Cognos
• Microsoft
• Oracle

Finance
• NASDAQ
• NYSE
• SEC
• NCR
• MicroStrategy
Matrix/Heatmap Visualizations for Time
Co-occurrence of organizations related to business intelligence

Alternate Network Visualizations: GraphTrail
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!
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.


Alternate Network Visualizations: CACTI
How to control a directed complex network with minimum number of nodes?

What is the minimum number of driver nodes ($N_D$) of real-world networks?

How to locate them efficiently?

Which topological characteristics determine $N_D$?

How robust is network controllability?
Complex systems are difficult to understand and represent. An organism’s genome, an ecosystem’s food web, and even a social network are just a few examples of complex systems important to diverse scientific disciplines but also to art and media. Connectedness is the essence of complex systems. A change in any one member element can propagate and affect the entire system—even if parts of it are loosely connected. How can we draw many interdependent connected components? Our technique is based on research on network controllability, and allows us to identify the “skeleton” of a variety of networks. With these technique data scientist do not need to represent every relationship to have a full sense of what the system does. Moreover, meta-information such as the overall system structure becomes comprehensible and the class of network can be perceived.
A) A directed network $G(A)$. All nodes in the network are called state vertices.

B) The network is controlled by three input vertices (origins), which are marked in blue. The controlled network is described by a digraph $G(A, B)$.

C) The U-rooted factorial connection of the digraph $G(A,B)$ is composed of vertex-disjoint stems and cycles.

D) The cacti is built from the U-rooted factorial connection. The cactus in the left contains a stem and four buds.
Coordinated Network Visualizations

Introduction
Node-Link
Composite
Alternate
Coordinated
Discussion

Mohammad S, Dunne C and Dorr B (2009), "Generating high-coverage semantic orientation lexicons from overtly marked words and a thesaurus", In EMNLP '09. pp. 599-608.

Coordinated Network Visualizations
Watson News Explorer

http://news-explorer.mybluemix.net/


http://www.cs.umd.edu/hcil/ase/
NetGrok
Visualizing real-time computer network traffic


http://www.cs.umd.edu/projects/netgrok/
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