Problem-Driven Interactive Visualization for Imperfect Models

Tamara Munzner
Department of Computer Science
University of British Columbia
@tamaramunzner

Huawei Vancouver
Jan 19 2022, virtual

http://www.cs.ubc.ca/~tmm/talks.html#huawei22

Outline

- methodology of problem-driven visualization research
- two case studies of visualizing imperfect models
  - NLP for temporal data
  - ML with graph neural networks
- brief overview of other problem-driven projects

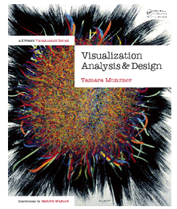
Visualization (vis) defined & motivated

Computer-based visualization systems provide visual representations of datasets designed to help people carry out tasks more effectively.

Visualization is suitable when there is a need to augment human capabilities rather than replace people with computational decision-making methods.

- human in the loop needs details about data
  - *entry point:* exploratory data analysis
    - don't know exactly what questions to ask in advance
  - *entry point:* presentation of known results
  - *entry point:* interplay of human judgement & computation/ML
    - refining model, trustbuilding/monitoring, mixed-initiative
- external representation: perception vs cognition
- intended task, measurable definitions of effectiveness

more at:
Visualization Analysis and Design, Chapter 1.
Munzner. AK Peters Visualization Series, CRC Press, 2014.



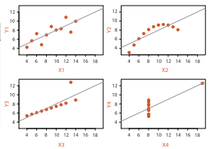
Unpacking data visually: From rollup to drilldown

Computer-based visualization systems provide visual representations of datasets designed to help people carry out tasks more effectively.

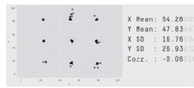
- summaries lose information, people can see a lot in the details
  - confirm expected and find unexpected patterns
  - assess validity of statistical model
  - sensitivity analysis for parameters

Anscombe's Quartet

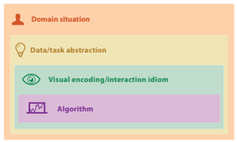
Identical statistics	
x mean	9
x variance	10
y mean	7.5
y variance	3.75
x/y correlation	0.816



Datasaurus Dozen



Same Stats, Different Graphs: Generating Datasets with Varied Appearance and Identical Statistics through Simulated Annealing. CHI 2017.
Matejka & Fitzmaurice



Tamara Munzner
@tamaramunzner

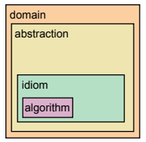
A Nested Model

for Visualization Design and Validation

http://www.cs.ubc.ca/labs/imager/tr/2009/NestedModel

A Nested Model for Visualization Design and Validation.
Munzner. IEEE Trans. Visualization and Computer Graphics (Proc. InfoVis 09), 15(6):921-928, 2009.

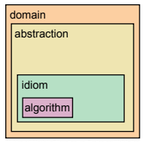
Nested model: Four levels of visualization concerns



[A Nested Model of Visualization Design and Validation. Munzner. IEEE TVCG 15(6):921-928, 2009 (Proc. InfoVis 2009).]

Nested model: Four levels of visualization concerns

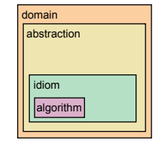
- *domain* situation
  - **who** are the target users?



[A Nested Model of Visualization Design and Validation. Munzner. IEEE TVCG 15(6):921-928, 2009 (Proc. InfoVis 2009).]

Nested model: Four levels of visualization concerns

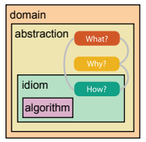
- *domain* situation
  - **who** are the target users?
- *abstraction*
  - translate from specifics of domain to vocabulary of vis



[A Nested Model of Visualization Design and Validation. Munzner. IEEE TVCG 15(6):921-928, 2009 (Proc. InfoVis 2009).]

Nested model: Four levels of visualization concerns

- *domain* situation
  - **who** are the target users?
- *abstraction*
  - translate from specifics of domain to vocabulary of vis
  - **what** is shown? **data abstraction**

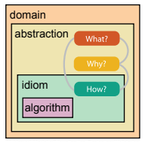


[A Nested Model of Visualization Design and Validation. Munzner. IEEE TVCG 15(6):921-928, 2009 (Proc. InfoVis 2009).]

[A Multi-Level Typology of Abstract Visualization Tasks Brehmer and Munzner. IEEE TVCG 19(12):2376-2385, 2013 (Proc. InfoVis 2013).]

Nested model: Four levels of visualization concerns

- *domain* situation
  - **who** are the target users?
- *abstraction*
  - translate from specifics of domain to vocabulary of vis
  - **what** is shown? **data abstraction**
    - often don't just draw what you're given: transform to new form

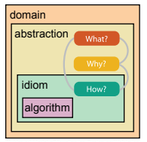


[A Nested Model of Visualization Design and Validation. Munzer. IEEE TVCG 15(6):921-928, 2009 (Proc. InfoVis 2009).]

[A Multi-Level Typology of Abstract Visualization Tasks Brehmer and Munzner. IEEE TVCG 19(12):2376-2385, 2013 (Proc. InfoVis 2013).]

Nested model: Four levels of visualization concerns

- *domain* situation
  - **who** are the target users?
- *abstraction*
  - translate from specifics of domain to vocabulary of vis
  - **what** is shown? **data abstraction**
    - often don't just draw what you're given: transform to new form
  - **why** is the user looking at it? **task abstraction**

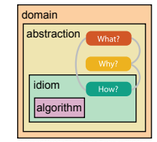


[A Nested Model of Visualization Design and Validation. Munzner. IEEE TVCG 15(6):921-928, 2009 (Proc. InfoVis 2009).]

[A Multi-Level Typology of Abstract Visualization Tasks Brehmer and Munzner. IEEE TVCG 19(12):2376-2385, 2013 (Proc. InfoVis 2013).]

Nested model: Four levels of visualization concerns

- *domain* situation
  - **who** are the target users?
- *abstraction*
  - translate from specifics of domain to vocabulary of vis
  - **what** is shown? **data abstraction**
    - often don't just draw what you're given: transform to new form
  - **why** is the user looking at it? **task abstraction**
- *idiom*
  - **how** is it shown?

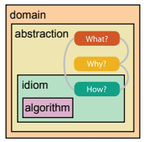


[A Nested Model of Visualization Design and Validation. Munzner. IEEE TVCG 15(6):921-928, 2009 (Proc. InfoVis 2009).]

[A Multi-Level Typology of Abstract Visualization Tasks Brehmer and Munzner. IEEE TVCG 19(12):2376-2385, 2013 (Proc. InfoVis 2013).]

Nested model: Four levels of visualization concerns

- *domain* situation
  - **who** are the target users?
- *abstraction*
  - translate from specifics of domain to vocabulary of vis
  - **what** is shown? **data abstraction**
    - often don't just draw what you're given: transform to new form
  - **why** is the user looking at it? **task abstraction**
- *idiom*
  - **how** is it shown?
    - **visual encoding idiom**: how to draw

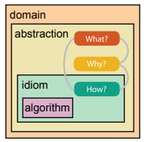


[A Nested Model of Visualization Design and Validation. Munzner. IEEE TVCG 15(6):921-928, 2009 (Proc. InfoVis 2009).]

[A Multi-Level Typology of Abstract Visualization Tasks Brehmer and Munzner. IEEE TVCG 19(12):2376-2385, 2013 (Proc. InfoVis 2013).]

Nested model: Four levels of visualization concerns

- *domain* situation
  - **who** are the target users?
- *abstraction*
  - translate from specifics of domain to vocabulary of vis
  - **what** is shown? **data abstraction**
    - often don't just draw what you're given: transform to new form
  - **why** is the user looking at it? **task abstraction**
- *idiom*
  - **how** is it shown?
    - **visual encoding idiom**: how to draw
    - **interaction idiom**: how to manipulate

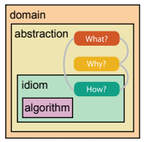


[A Nested Model of Visualization Design and Validation. Munzner. IEEE TVCG 15(6):921-928, 2009 (Proc. InfoVis 2009).]

[A Multi-Level Typology of Abstract Visualization Tasks Brehmer and Munzner. IEEE TVCG 19(12):2376-2385, 2013 (Proc. InfoVis 2013).]

Nested model: Four levels of visualization concerns

- *domain* situation
  - **who** are the target users?
- *abstraction*
  - translate from specifics of domain to vocabulary of vis
  - **what** is shown? **data abstraction**
    - often don't just draw what you're given: transform to new form
  - **why** is the user looking at it? **task abstraction**
- *idiom*
  - **how** is it shown?
    - **visual encoding idiom**: how to draw
    - **interaction idiom**: how to manipulate
- *algorithm*
  - efficient computation



[A Nested Model of Visualization Design and Validation. Munzner. IEEE TVCG 15(6):921-928, 2009 (Proc. InfoVis 2009).]

[A Multi-Level Typology of Abstract Visualization Tasks Brehmer and Munzner. IEEE TVCG 19(12):2376-2385, 2013 (Proc. InfoVis 2013).]

Why is validation difficult?

- different ways to get it wrong at each level

[A Nested Model of Visualization Design and Validation. Munzner. IEEE TVCG 15(6):921-928, 2009 (Proc. InfoVis 2009).]

Why is validation difficult?

- different ways to get it wrong at each level

Domain situation
You misunderstood their needs

[A Nested Model of Visualization Design and Validation. Munzner. IEEE TVCG 15(6):921-928, 2009 (Proc. InfoVis 2009).]

Why is validation difficult?

- different ways to get it wrong at each level

Domain situation
You misunderstood their needs

Data/task abstraction
You're showing them the wrong thing

[A Nested Model of Visualization Design and Validation. Munzner. IEEE TVCG 15(6):921-928, 2009 (Proc. InfoVis 2009).]

Why is validation difficult?

- different ways to get it wrong at each level

Domain situation
You misunderstood their needs

Data/task abstraction
You're showing them the wrong thing

Visual encoding/interaction idiom
The way you show it doesn't work

[A Nested Model of Visualization Design and Validation. Munzner. IEEE TVCG 15(6):921-928, 2009 (Proc. InfoVis 2009).]

Why is validation difficult?

- different ways to get it wrong at each level

Domain situation
You misunderstood their needs

Data/task abstraction
You're showing them the wrong thing

Visual encoding/interaction idiom
The way you show it doesn't work

Algorithm
Your code is too slow

[A Nested Model of Visualization Design and Validation. Munzner. IEEE TVCG 15(6):921-928, 2009 (Proc. InfoVis 2009).]

Validation solution: use methods from appropriate fields at each level

[A Nested Model of Visualization Design and Validation. Munzner. IEEE TVCG 15(6):921-928, 2009 (Proc. InfoVis 2009).]

Validation solution: use methods from appropriate fields at each level

computer science

Algorithm
Measure system time/memory
Analyze computational complexity

[A Nested Model of Visualization Design and Validation. Munzner. IEEE TVCG 15(6):921-928, 2009 (Proc. InfoVis 2009).]

Validation solution: use methods from appropriate fields at each level

computer science

Algorithm
Measure system time/memory
Analyze computational complexity

technique-driven work

[A Nested Model of Visualization Design and Validation. Munzner. IEEE TVCG 15(6):921-928, 2009 (Proc. InfoVis 2009).]

Validation solution: use methods from appropriate fields at each level

design

Visual encoding/interaction idiom
Justify design with respect to alternatives

computer science

Algorithm
Measure system time/memory
Analyze computational complexity

technique-driven work

cognitive psychology

Analyze results qualitatively
Measure human time with lab experiment (*lab study*)

[A Nested Model of Visualization Design and Validation. Munzner. IEEE TVCG 15(6):921-928, 2009 (Proc. InfoVis 2009).]

Validation solution: use methods from appropriate fields at each level

anthropology/ ethnography

Domain situation
Observe target users using existing tools

Data/task abstraction

design

Visual encoding/interaction idiom
Justify design with respect to alternatives

computer science

Algorithm
Measure system time/memory
Analyze computational complexity

technique-driven work

cognitive psychology

Analyze results qualitatively
Measure human time with lab experiment (*lab study*)

anthropology/ ethnography

Observe target users after deployment (*field study*)
Measure adoption

[A Nested Model of Visualization Design and Validation. Munzner. IEEE TVCG 15(6):921-928, 2009 (Proc. InfoVis 2009).]

Validation solution: use methods from appropriate fields at each level

anthropology/ ethnography

Domain situation
Observe target users using existing tools

problem-driven work

Data/task abstraction

design

Visual encoding/interaction idiom
Justify design with respect to alternatives

computer science

Algorithm
Measure system time/memory
Analyze computational complexity

technique-driven work

cognitive psychology

Analyze results qualitatively
Measure human time with lab experiment (*lab study*)

anthropology/ ethnography

Observe target users after deployment (*field study*)
Measure adoption

[A Nested Model of Visualization Design and Validation. Munzner. IEEE TVCG 15(6):921-928, 2009 (Proc. InfoVis 2009).]

Validation solution: use methods from appropriate fields at each level

- avoid mismatches between level and validation

anthropology/ ethnography

Domain situation
Observe target users using existing tools

problem-driven work

Data/task abstraction

design

Visual encoding/interaction idiom
Justify design with respect to alternatives

computer science

Algorithm
Measure system time/memory
Analyze computational complexity

technique-driven work

cognitive psychology

Analyze results qualitatively
Measure human time with lab experiment (*lab study*)

anthropology/ ethnography

Observe target users after deployment (*field study*)
Measure adoption

[A Nested Model of Visualization Design and Validation. Munzner. IEEE TVCG 15(6):921-928, 2009 (Proc. InfoVis 2009).]

Visualization: Angles of attack

problem-driven work

Visualization: Angles of attack

problem-driven work ↔ **technique-driven work**

Visualization: Angles of attack

problem-driven work ↔ **technique-driven work**

evaluation

Visualization: Angles of attack

problem-driven work ↔ **technique-driven work**

theoretical foundations

evaluation

Visualization: Angles of attack

problem-driven work ↔ **technique-driven work**

theoretical foundations

evaluation

Problem-driven visualization: Design studies

Problem driven visualization: Design studies

"A design study is a project in which visualization researchers analyze a specific real-world problem faced by domain experts..."

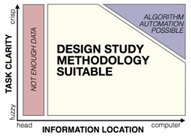
Problem driven visualization: Design studies

"A design study is a project in which visualization researchers analyze a specific real-world problem faced by domain experts, design a visualization system that supports solving this problem..."

Problem driven visualization: Design studies

"A design study is a project in which visualization researchers analyze a specific real-world problem faced by domain experts, design a visualization system that supports solving this problem, validate the design, and reflect about lessons learned in order to refine visualization design guidelines."

[Design Study Methodology: Reflections from the Trenches and the Stacks. Sedlmair, Meyer & Munzner. IEEE TVCG 18(12): 2431-2440, 2012 (Proc. InfoVis 2012).]



Michael Sedlmair

Miriah Meyer

Design Study Methodology

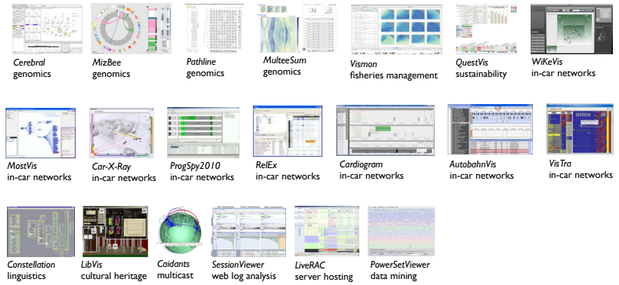
Reflections from the Trenches and from the Stacks

Tamara Munzner

http://www.cs.ubc.ca/labs/imager/tr/2012/dsm/

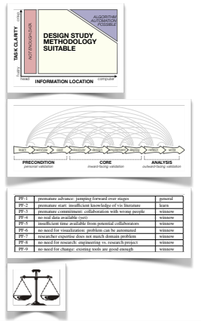
Design Study Methodology: Reflections from the Trenches and from the Stacks.
Sedlmair, Meyer, Munzner. IEEE Trans. Visualization and Computer Graphics 18(12): 2431-2440, 2012 (Proc. InfoVis 2012).

Lessons learned from the trenches: 20+ between us



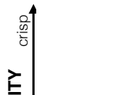
Methodology for problem-driven work

- definitions
- 9-stage framework
- 32 pitfalls & how to avoid them
- comparison to related methodologies



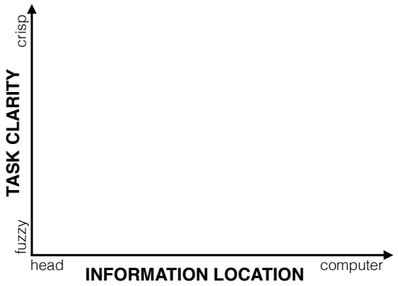
http://www.cs.ubc.ca/~tmm/talks.html#g21

Design study methodology: definitions



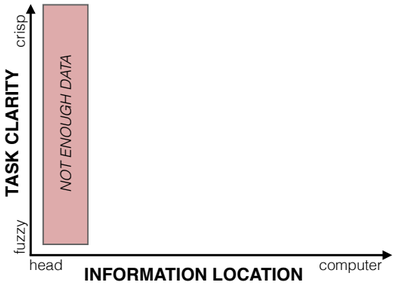
[Design Study Methodology: Reflections from the Trenches and the Stacks. Sedlmair, Meyer & Munzner. IEEE TVCG 18(12): 2431-2440, 2012 (Proc. InfoVis 2012).]

Design study methodology: definitions



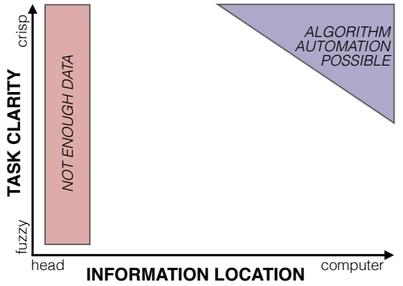
[Design Study Methodology: Reflections from the Trenches and the Stacks. Sedlmair, Meyer & Munzner. IEEE TVCG 18(12): 2431-2440, 2012 (Proc. InfoVis 2012).]

Design study methodology: definitions



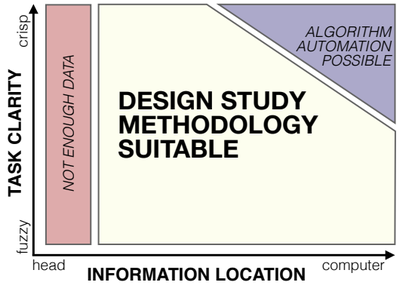
[Design Study Methodology: Reflections from the Trenches and the Stacks. Sedlmair, Meyer & Munzner. IEEE TVCG 18(12): 2431-2440, 2012 (Proc. InfoVis 2012).]

Design study methodology: definitions



[Design Study Methodology: Reflections from the Trenches and the Stacks. Sedlmair, Meyer & Munzner. IEEE TVCG 18(12): 2431-2440, 2012 (Proc. InfoVis 2012).]

Design study methodology: definitions



[Design Study Methodology: Reflections from the Trenches and the Stacks. Sedlmair, Meyer & Munzner. IEEE TVCG 18(12): 2431-2440, 2012 (Proc. InfoVis 2012).]

32 pitfalls & how to avoid them

32 pitfalls & how to avoid them

learn → winnow → cast

PF-1	premature advance: jumping forward over stages	general
PF-2	premature start: insufficient knowledge of vis literature	learn
PF-3	premature commitment: collaboration with wrong people	winnow
PF-4	no real data available (yet)	winnow
PF-5	insufficient time available from potential collaborators	winnow
PF-6	no need for visualization: problem can be automated	winnow
PF-7	researcher expertise does not match domain problem	winnow
PF-8	no need for research: engineering vs. research project	winnow
PF-9	no need for change: existing tools are good enough	winnow
PF-10	no real/important/recurring task	winnow
PF-11	no rapport with collaborators	winnow
PF-12	not identifying front line analyst and gatekeeper before start	cast
PF-13	assuming every project will have the same role distribution	cast
PF-14	mistaking fellow tool builders for real end users	cast

32 pitfalls & how to avoid them

learn → winnow → cast → discover → design → implement → deploy

PF-1	premature advance: jumping forward over stages	general	PF-21	mistaking technique-driven for problem-driven work	design
PF-2	premature start: insufficient knowledge of vis literature	learn	PF-22	nonrapid prototyping	implement
PF-3	premature commitment: collaboration with wrong people	winnow	PF-23	usability: too little / too much	implement
PF-4	no real data available (yet)	winnow	PF-24	premature end: insufficient deploy time built into schedule	deploy
PF-5	insufficient time available from potential collaborators	winnow	PF-25	usage study not case study: non-real task/data/user	deploy
PF-6	no need for visualization: problem can be automated	winnow	PF-26	*liking* necessary but not sufficient for validation	deploy
PF-7	researcher expertise does not match domain problem	winnow			
PF-8	no need for research: engineering vs. research project	winnow			
PF-9	no need for change: existing tools are good enough	winnow			
PF-10	no real/important/recurring task	winnow			
PF-11	no rapport with collaborators	winnow			
PF-12	not identifying front line analyst and gatekeeper before start	cast			
PF-13	assuming every project will have the same role distribution	cast			
PF-14	mistaking fellow tool builders for real end users	cast			
PF-15	ignoring practices that currently work well	discover			
PF-16	expecting *just talking* or *fly on wall* to work	discover			
PF-17	experts focusing on visualization design vs. domain problem	discover			
PF-18	learning their problems/language: too little / too much	discover			
PF-19	abstraction: too little	design			
PF-20	premature design commitment: consideration space too small	design			

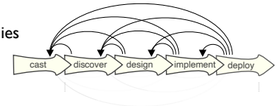
32 pitfalls & how to avoid them

learn → winnow → cast → discover → design → implement → deploy → reflect → write

PF-1	premature advance: jumping forward over stages	general	PF-21	mistaking technique-driven for problem-driven work	design
PF-2	premature start: insufficient knowledge of vis literature	learn	PF-22	nonrapid prototyping	implement
PF-3	premature commitment: collaboration with wrong people	winnow	PF-23	usability: too little / too much	implement
PF-4	no real data available (yet)	winnow	PF-24	premature end: insufficient deploy time built into schedule	deploy
PF-5	insufficient time available from potential collaborators	winnow	PF-25	usage study not case study: non-real task/data/user	deploy
PF-6	no need for visualization: problem can be automated	winnow	PF-26	*liking* necessary but not sufficient for validation	deploy
PF-7	researcher expertise does not match domain problem	winnow	PF-27	failing to improve guidelines: confirm, refine, reject, propose	reflect
PF-8	no need for research: engineering vs. research project	winnow	PF-28	insufficient writing time built into schedule	write
PF-9	no need for change: existing tools are good enough	winnow	PF-29	no technique contribution ≠ good design study	write
PF-10	no real/important/recurring task	winnow	PF-30	too much domain background in paper	write
PF-11	no rapport with collaborators	winnow	PF-31	story told chronologically vs. focus on final results	write
PF-12	not identifying front line analyst and gatekeeper before start	cast	PF-32	premature end: win race vs. practice music for debut	write
PF-13	assuming every project will have the same role distribution	cast			
PF-14	mistaking fellow tool builders for real end users	cast			
PF-15	ignoring practices that currently work well	discover			
PF-16	expecting *just talking* or *fly on wall* to work	discover			
PF-17	experts focusing on visualization design vs. domain problem	discover			
PF-18	learning their problems/language: too little / too much	discover			
PF-19	abstraction: too little	design			
PF-20	premature design commitment: consideration space too small	design			

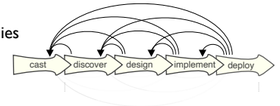
Design studies & user-centered design

- user-centered design: well-known HCI methodology
  - iterative refinement & deployment
  - evaluation through case studies & field studies



Design studies & user-centered design

- user-centered design: well-known HCI methodology
  - iterative refinement & deployment
  - evaluation through case studies & field studies



- what's specific to visualization?
  - discovering task and data **abstractions**
  - designing visual encoding & interaction **idioms** that map to abstractions

Two case studies of visualizing imperfect models

- NLP for temporal data

- ML for graph data

time by Wayne Middleton from the Noun Project machine learning by Eli Magaziner from the Noun Project

Two case studies of visualizing imperfect models

- NLP for temporal data

- ML for graph data

time by Wayne Middleton from the Noun Project machine learning by Eli Magaziner from the Noun Project

TimeLineCurator

Interactive Authoring of Visual Timelines from Unstructured Text

Johanna Fulda
@jofu_

Matthew Brehmer
@mattbrehmer

Tamara Munzner
@tamaramunzner

http://about.timelinecurator.org
http://timelinecurator.org

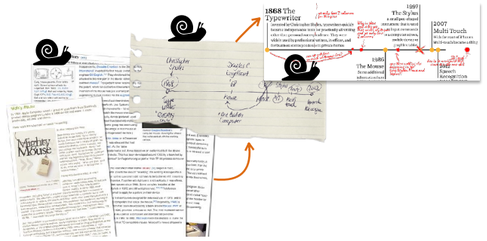
TimeLineCurator: Interactive Authoring of Visual Timelines from Unstructured Text.
Fulda, Brehmer, Munzner. IEEE Trans. Visualization and Computer Graphics (Proc IEEE VAST 2015) 22(1):300-309, 2015.

TimeLineCurator
visual & browser-based

https://vimeo.com/jofu/tlc

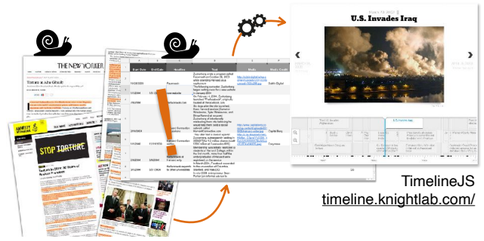
Manual creation process

Browse → Extract → Format → Show → Update



Structured creation process

Browse → Extract → Format → Show → Update



TimelineJS
timeline.knightlab.com/

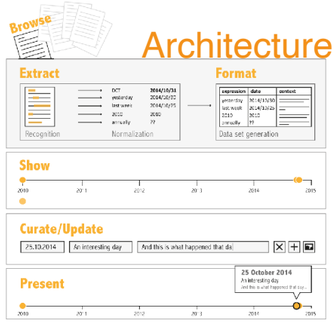
Timeline authoring model

- time required for each task

	Browse	Extract	Format	Show	Update
Manual Drawing	slow	slow		slow	slow
Structured Creation	slow	slow	slow	automated	fast
TimeLine Curator	fast	automated	automated	fast	fast

The general case for curation

- build for human in the loop as continuing need
  - automatic processing to accelerate not replace
  - **assume computational results good but not perfect**
    - for the indefinite future!
  - visual feedback to accelerate



The importance of being brisk

- cool use case: eureka moment
  - success: enable what was impossible before
  - vis tools for new insights & discoveries
- workhorse use case: workflow speedup
  - success: vis tools accelerate your prior workflow
    - sometimes enables the previously infeasible
- TLC use cases
  - started with speedup use case, for presentation
    - make this doc into a timeline now!
  - two other use cases nudge towards exploration
    - comparison between multiple timelines
    - speculative browsing

TimeLineCurator: Speculative Browsing

speculative browsing

https://vimeo.com/jofu/tlc

Visualizing Graph Neural Networks with CorGIE:

Corresponding a Graph to Its Embedding

https://arxiv.org/abs/2106.12839

Zipeng Liu
UBC/Beihang

Yang Wang
Uber/Facebook

Jürgen Bernard
UBC/Zurich

Tamara Munzner
UBC

Visualizing Graph Neural Networks with CorGIE: Corresponding a Graph to Its Embedding.
Liu, Wang, Bernard, Munzner. Under review.

Two case studies of visualizing imperfect models

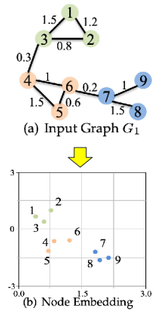
- NLP for temporal data

- ML for graph data

time by Wayne Middleton from the Noun Project machine learning by Eli Magaziner from the Noun Project

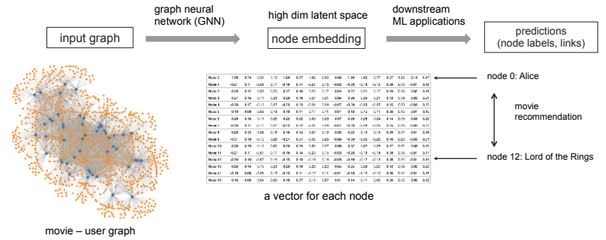
Graph neural network (GNN)

- machine learning (ML) models for graphs
  - like CNN for images
  - like Transformer for text
- many real-world graph-related applications
  - node classification
    - examples: fraud detection, disease classification
  - link prediction
    - examples: product recommendation, protein interactions



[Cai et al. TKDE'18]

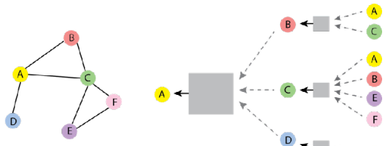
Graph neural network (GNN)



Graph neural network (GNN)

input graph → graph neural network (GNN) → high dim latent space: node embedding → downstream ML applications → predictions (node labels, links)

node features are aggregated / passed through **topological neighborhood**



Remake from https://snap-stanford.github.io/cs224w-notes/machine-learning-with-networks/graph-neural-networks

Evaluate GNN

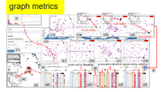
Two big-picture questions

- "Are we there yet?": should we train / tune more?
- "Are we lost?": does it behave as we expect?

Evaluate GNN: Previous approaches

input graph → graph neural network (GNN) → high dim latent space: node embedding → downstream ML applications → predictions (node labels, links)

graph metrics



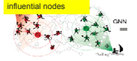
Li et al. EmbeddingVis. VAST'18.

dimensional reduction (simple inspection)

quant metrics

e.g. accuracy, hit rate

influential nodes



Ying et al. GNNExplainer.

node classification (predictions → input graph)

Jin et al. GNNVis. Arxiv'20.

input graph & node embedding under-used!

Evaluate GNN: CorGIE idea

input graph → graph neural network (GNN) → high dim latent space: node embedding → downstream ML applications → predictions (node labels, links)

shared topo neighbors
similar node features

explore correspondences

nearby positions

Evaluate GNN: CorGIE idea

input graph → graph neural network (GNN) → high dim latent space: node embedding → downstream ML applications → predictions (node labels, links)

shared topo neighbors
similar node features

explore correspondences

nearby positions

Examples of correspondences:

Check [similar topology? Similar node features?] ← - - - Pick [a cluster]

Evaluate GNN: CorGIE idea

input graph → graph neural network (GNN) → high dim latent space: node embedding → downstream ML applications → predictions (node labels, links)

shared topo neighbors
similar node features

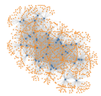
explore correspondences

nearby positions

Examples of correspondences:

Check [similar topology? Similar node features?] ← - - - Pick [a cluster]

Check [different topology? Different node features?] ← - - - Pick [two far-away clusters]



Evaluate GNN: CorGIE idea

input graph → graph neural network (GNN) → high dim latent space: node embedding → downstream ML applications → predictions (node labels, links)

shared topo neighbors
similar node features

explore correspondences

nearby positions

Examples of correspondences:

Check [similar topology? Similar node features?] ← - - - Pick [a cluster]

Check [different topology? Different node features?] ← - - - Pick [two far-away clusters]

Pick [two nodes sharing many topo neighbors] - - - → Check [how close the nodes are compared to others?]

Data and tasks

Topology space | **Latent space** | **Feature space**

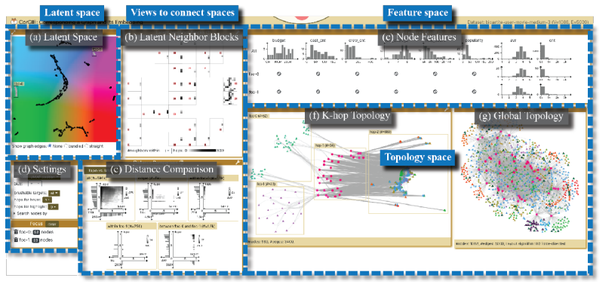
- data spaces

Data and tasks

Topology space (Specify) — Correspond — **Latent space** (Specify) — Correspond — **Feature space** (Specify)

- data spaces
- tasks
  - specify
  - correspond
- task iteration
  - levels in grouping structure of nodes

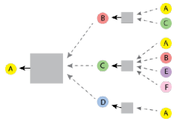
Contribution: Multi-view interactive interface

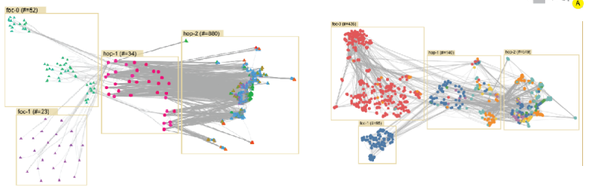


Video: https://osf.io/j56hu/

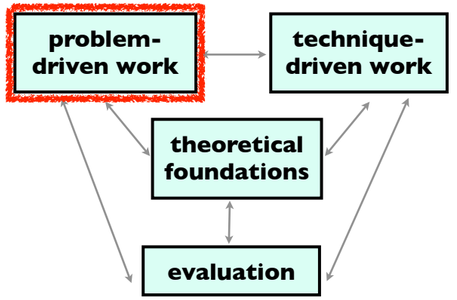
Contribution: K-hop layout

- show topo neighbors of user-specified node sets
  - mimic how info is aggregated in GNN
    - boxes from left to right: specified nodes, hop-1, hop-2, …
  - within box, cluster nodes by their topo connections





Problem-driven work: many domains



Problem-driven: Genomics

Aaron Barsky

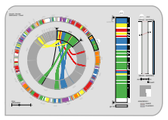
Jenn Gardy (Microbio)

Robert Kincaid (Agilent)

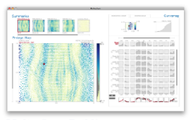
Cerebral
https://youtu.be/76HhG1FQngI

Miriah Meyer

Hanspeter Pfister (Harvard)



MizBee
https://youtu.be/86p7brwuz2g



MulteeSum, Pathline

Problem-driven: Genomics, fisheries

Joel Ferstay

Cydney Nielsen (BC Cancer)

Variant View
https://youtu.be/AHDnv_qMXxQ

Maryam Booshehrian

Torsten Moeller (SFU)

Vismon https://youtu.be/h0kHoS4VYmk

Problem-driven: Tech industry

Heidi Lam

Diane Tang (Google)

SessionViewer: web log analysis
https://youtu.be/T4MaTZd56G4

Peter McLachlan

Stephen North (AT&T Research)

LiveRAC: systems time-series
https://youtu.be/Id0c3H0VSkw

Problem-driven: Building energy mgmt, journalism

Matt Brehmer

Kevin Tate (Pulse/EnerNOC)

Energy Manager

Matt Brehmer

Stephen Ingram

Jonathan Stray (Assoc Press)

Overview
https://vimeo.com/71483614

Problem-driven: Data science

Kimberly Dextras-Romagnino

Segmentifier (Mobify)

e-commerce clickstreams

build tools for human-in-the-loop visual data analysis

https://youtu.be/TobYDFeISOg

Michael Oppermann

Ocupado (Sensible Building Science)

wifi proxy for real-time building occupancy

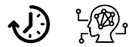
visual analytics for facilities management

https://youtu.be/KcwjVK8eUdw

Problem-driven visualization for imperfect models

- problem-driven methodology
  - translate domain problems into abstractions
    - before visual encoding idioms & algorithms
  - avoid collaboration pitfalls
    - understand roles, ensure aligned incentives

- interactive visualization supporting human-in-the-loop judgements about models
  - two cases: different data types



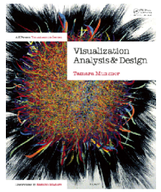
- overview: other problem-driven projects

More information

- this talk
  http://www.cs.ubc.ca/~tmm/talks.html#huawei22

- papers, videos, software, talks, courses
  http://www.cs.ubc.ca/group/infovis
  http://www.cs.ubc.ca/~tmm

- theoretical foundations: book
  (+ tutorial/course lecture slides)
  http://www.cs.ubc.ca/~tmm/vadbook



Visualization Analysis and Design.
Munzner.
AK Peters Visualization Series.
CRC Press, 2014.

@tamaramunzner

UBC

UBC Data Science Institute

DESIGNING for PEOPLE

CAIDA