Characterization of Information Visualization Systems

Tamara Munzner
Department of Computer Science
University of British Columbia

Conf. on Quantification in Visual Computing
Stuttgart, Germany
9 Oct 2018

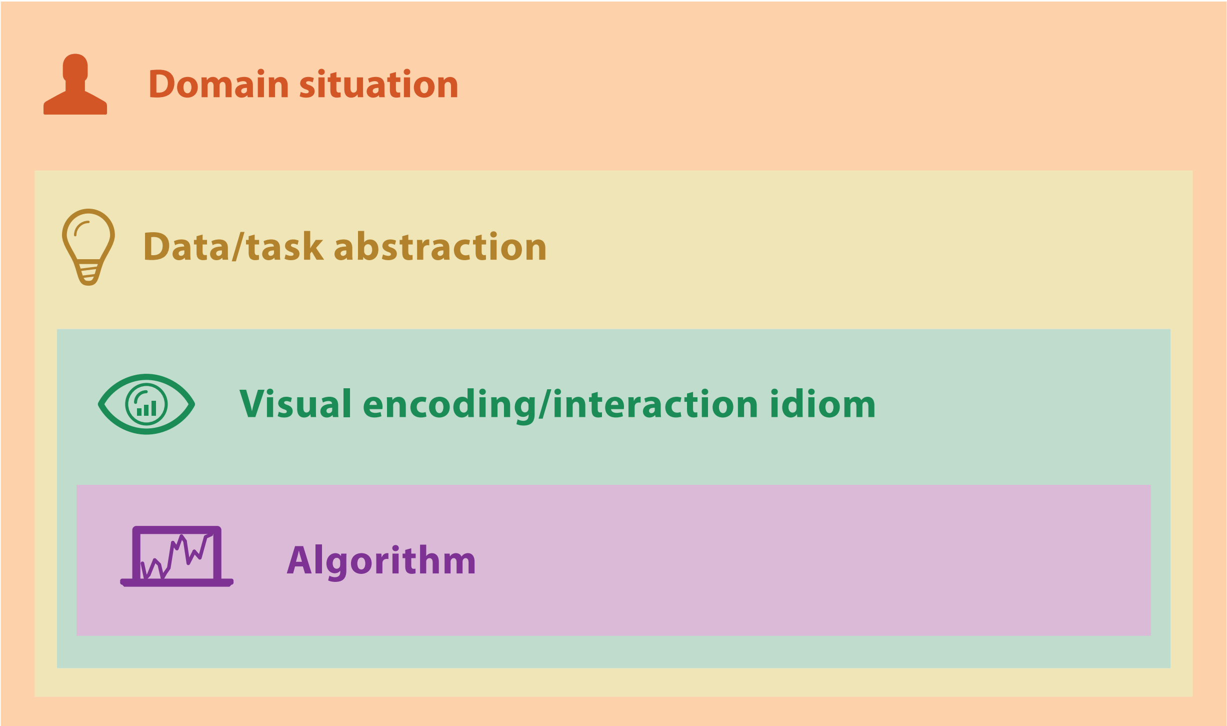
www.cs.ubc.ca/~tmm/talks.html#stuttgart18

@tamaramunzner

Quantification and visualization: Challenges

- When to use what methods for evaluating visualization designs?
  - Formalism: Nested model
- What role can qualitative methods play in developing quantitative metrics?
- How can we evaluate quantitative metrics beyond significance testing?
  - In-depth case study: Search sets for path tracing in node-link graphs

When to use what methods?



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.

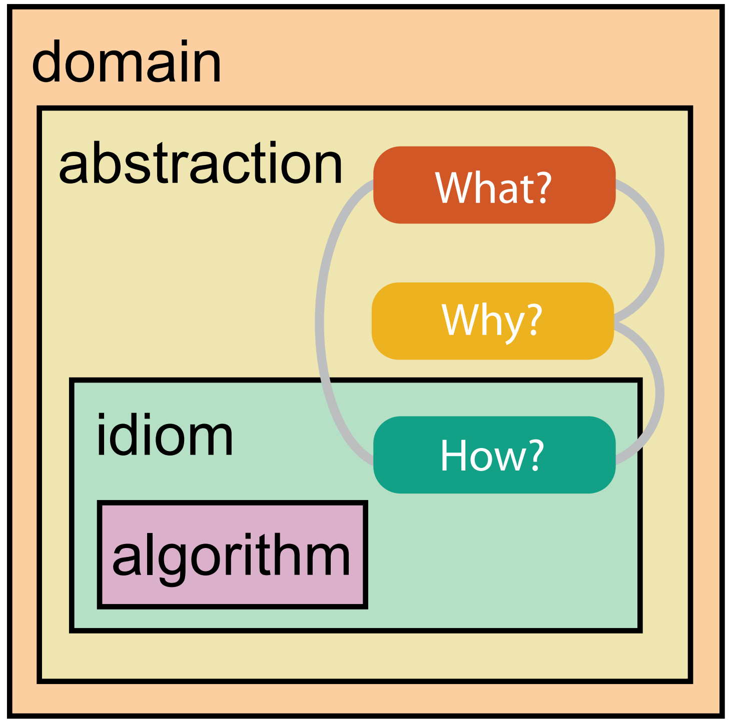
How to evaluate a visualization: So many methods, how to pick?

- Computational benchmarks?
  - quant: system performance, memory
- User study in lab setting?
  - quant: (human) time and error rates, preferences
  - qual: behavior/strategy observations
- Field study of deployed system?
  - quant: usage logs
  - qual: interviews with users, case studies, observations
- Analysis of results?
  - quant: metrics computed on result images
  - qual: consider what structure is visible in result images
- Justification of choices?
  - qual: perceptual principles, best practices

Nested model: Four levels of visualization design

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

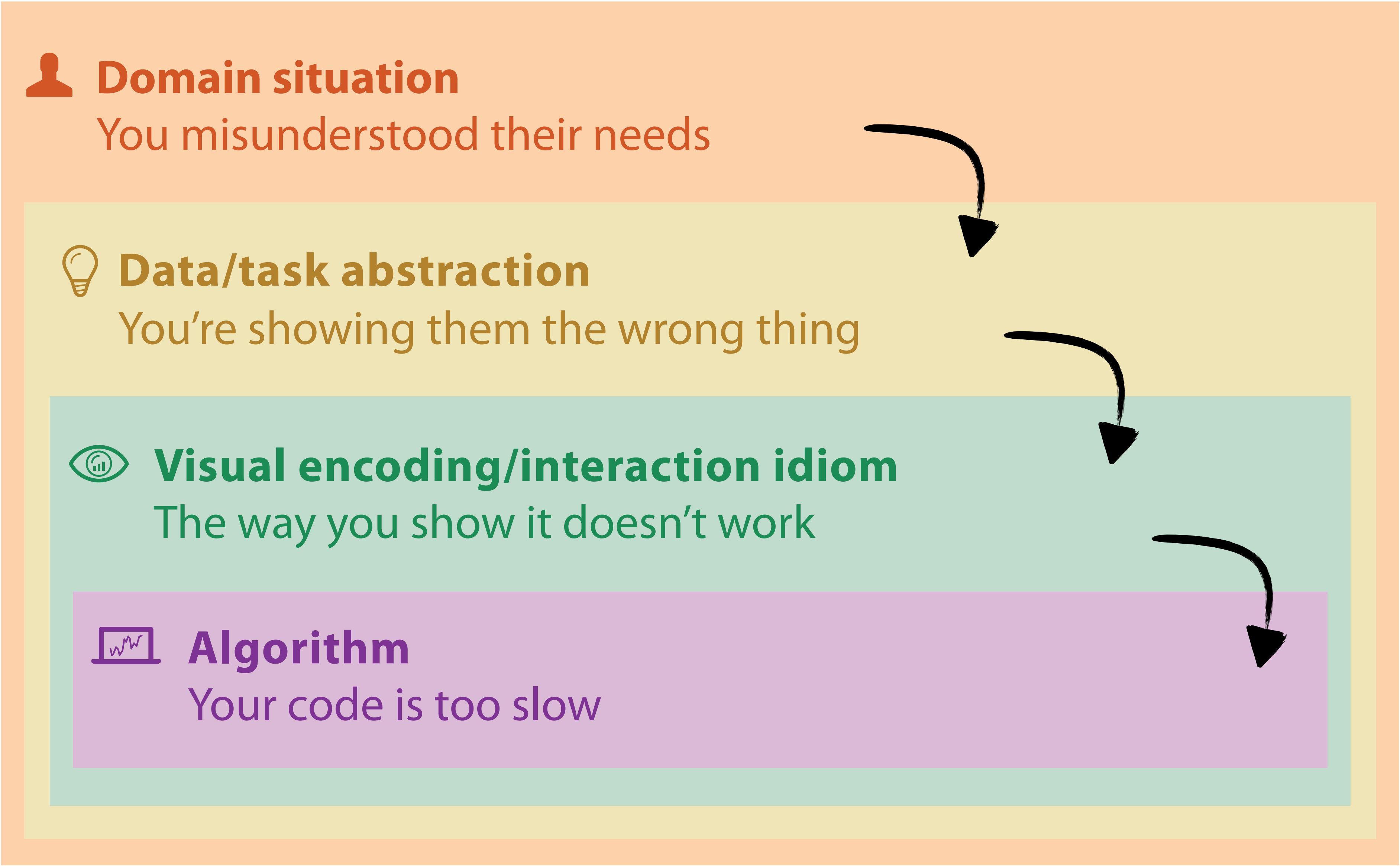
- *domain situation*
  - who are the target users?
- *abstraction*
  - translate from specifics of domain to vocabulary of visualization
    - **what** is shown? **data** abstraction
    - **why** is the user looking at it? **task** abstraction
- *idiom*
  - **how** is it shown?
    - **visual encoding** idiom: how to draw
    - **interaction** idiom: how to manipulate
- *algorithm*
  - efficient computation



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

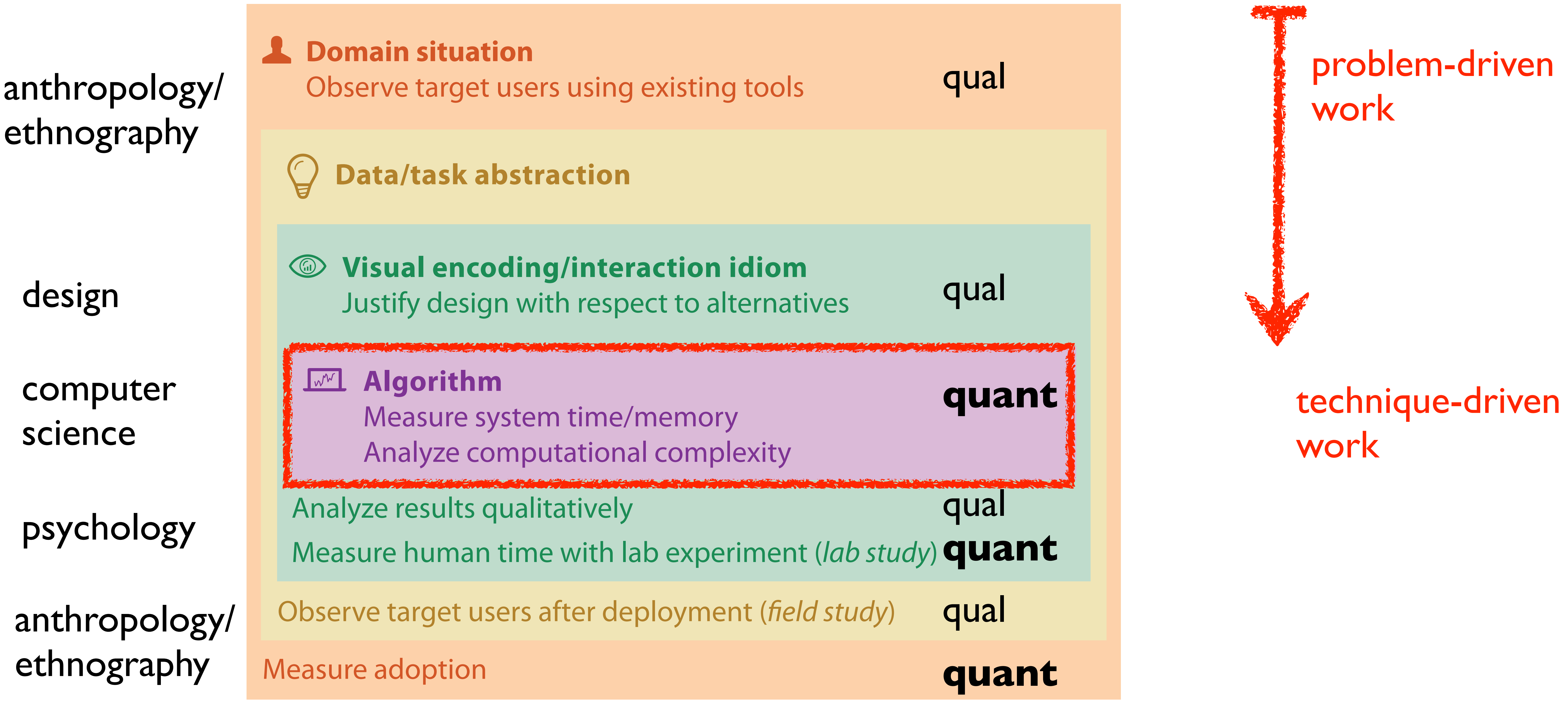
Different threats to validity at each level

- cascading effects downstream



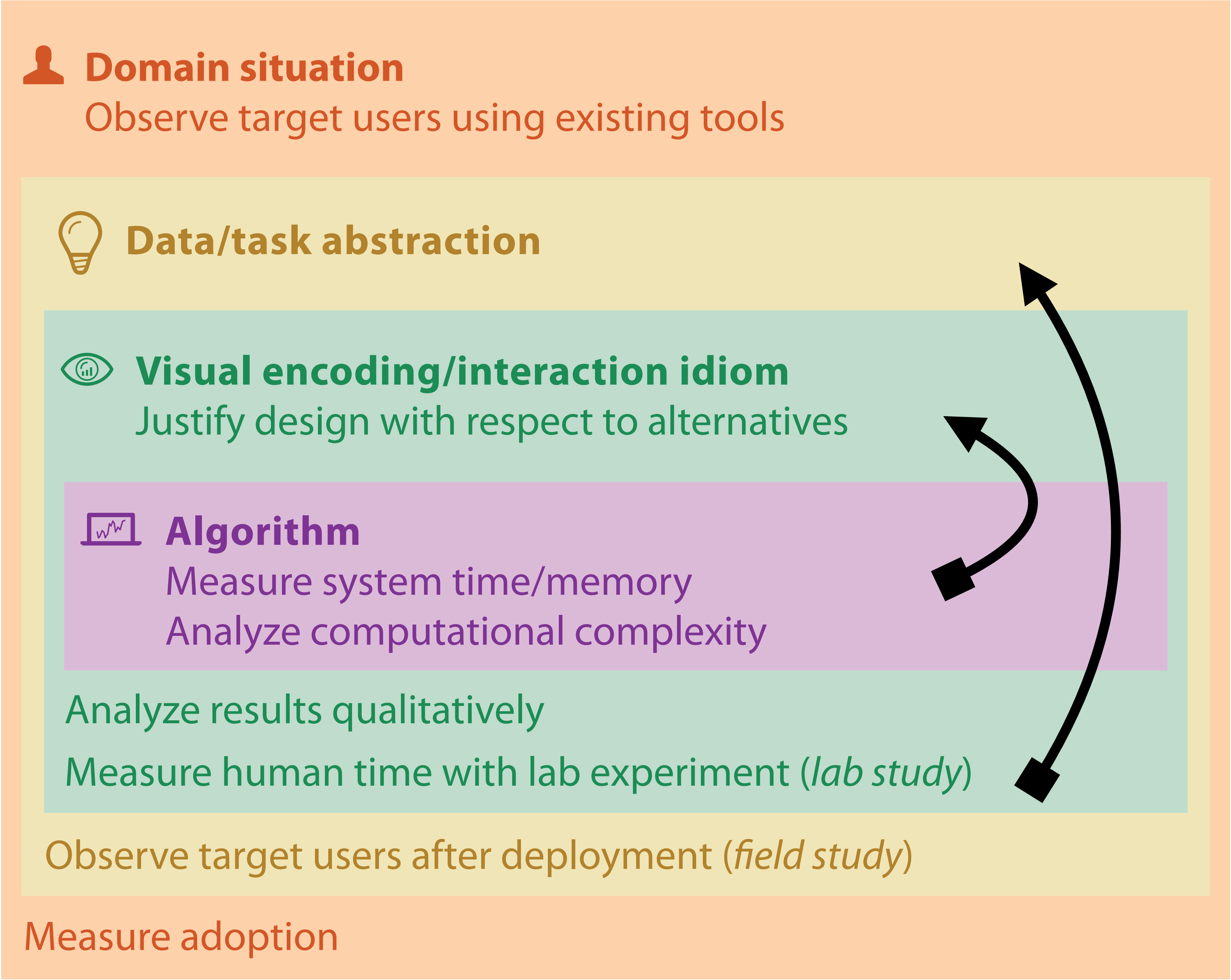
Interdisciplinary: need methods from different fields at each level

- mix of qual and quant approaches (typically)



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

Mismatches: Common problem



benchmarks can't confirm design

lab studies can't confirm task abstraction

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

Analysis examples: Single paper includes only subset of methods

MatrixExplorer. Henry and Fekete. InfoVis 2006.

- observe and interview target users
- justify encoding/interaction design
- measure system time/memory
- qualitative result image analysis

LiveRAC. McLachlan, Munzner, Koutsofios, and North. CHI 2008.

- observe and interview target users
- justify encoding/interaction design
- qualitative result image analysis
- field study, document deployed usage

An energy model for visual graph clustering. (LinLog) Noack. Graph Drawing 2003

- qualitative/quantitative image analysis

Effectiveness of animation in trend visualization. Robertson et al. InfoVis 2008.

- lab study, measure time/errors for operation

Interactive visualization of genealogical graphs. McGuffin and Balakrishnan. InfoVis 2005.

- justify encoding/interaction design
- qualitative result image analysis
- test on target users, get utility anecdotes

Flow map layout. Phan et al. InfoVis 2005.

- justify encoding/interaction design
- computational complexity analysis
- measure system time/memory
- qualitative result image analysis

Role of quant methods in qual metrics?
How to eval quant metrics?

Jessica Dawson

A search-set model of path tracing in graphs

Joanna McGrenere

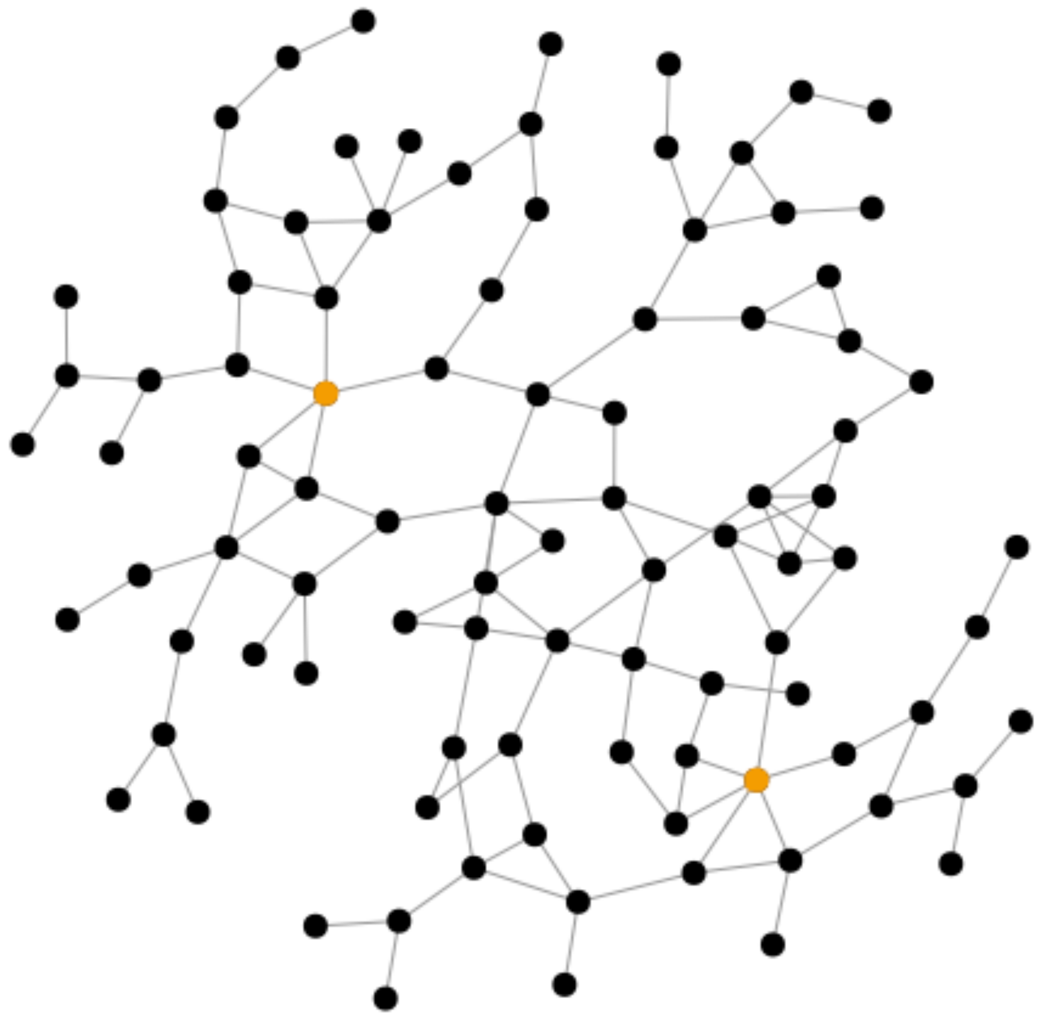
joint work with:
Jessica Q. Dawson, Joanna McGrenere

http://www.cs.ubc.ca/labs/imager/tr/2014/SearchSet

A search-set model of path tracing in graphs.
Dawson, Munzner, McGrenere. Information Visualization, 14(4):308-338 2015.

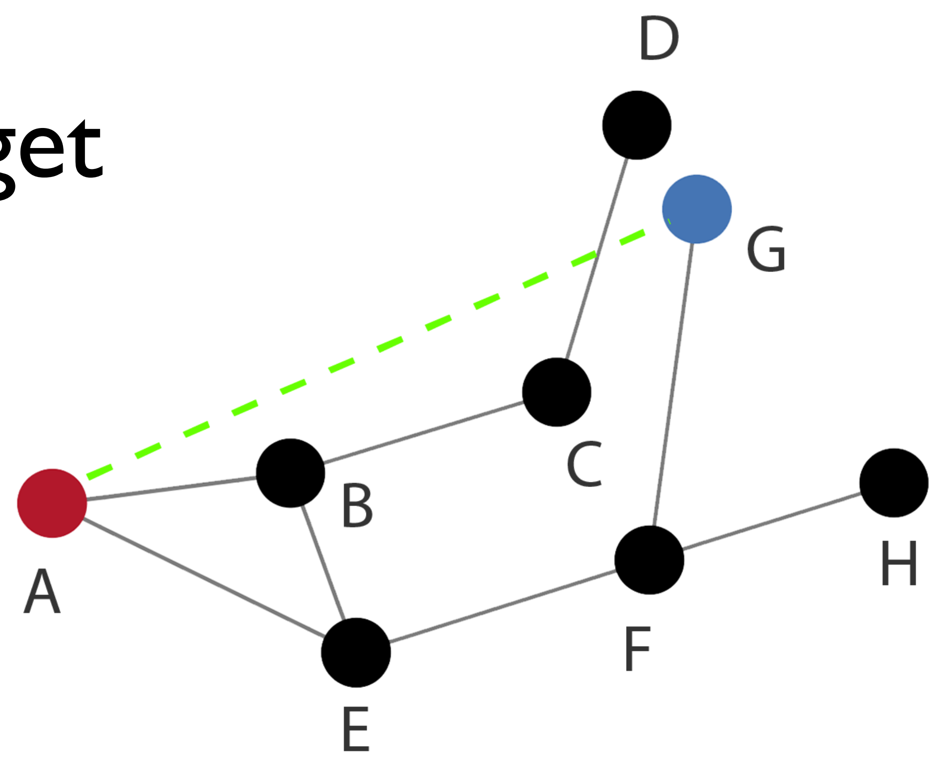
Path tracing in node-link graphs

- widely studied abstract task in previous work
  [Ghoniem et al 2002, Comparison of the Readability of Graphs Using Node-Link and Matrix-Based Representations]
  [Lee et al 2006, Task Taxonomy for Graph Visualization]
- common concrete task in real-world contexts
  - movie domain:
    How much distance between me and Kevin Bacon?
  - epidemiology domain:
    How many potential disease transmission paths between two people?



Human behavior & graph readability

- previous work observing human behaviour when interacting with graphs
  - identify new metrics
    [van Ham & Rogowitz, 2008] [Dwyer et al., 2009] [Purchase et al., 2012]
  - understand how metrics operate through eye tracking
    [Körner, 2004] [Huang, Eades, Hong 2009] [Huang, 2013]
- one eye tracking study led to identification of a path tracing behavior:
  geodesic tendency
  people look along straight line towards target
  [Huang, Eades, and Hong. 2009
  A Graph Reading Behavior: Geodesic-Path Tendency]



Geodesic tendency

1. First try closest to geodesic:

   A B

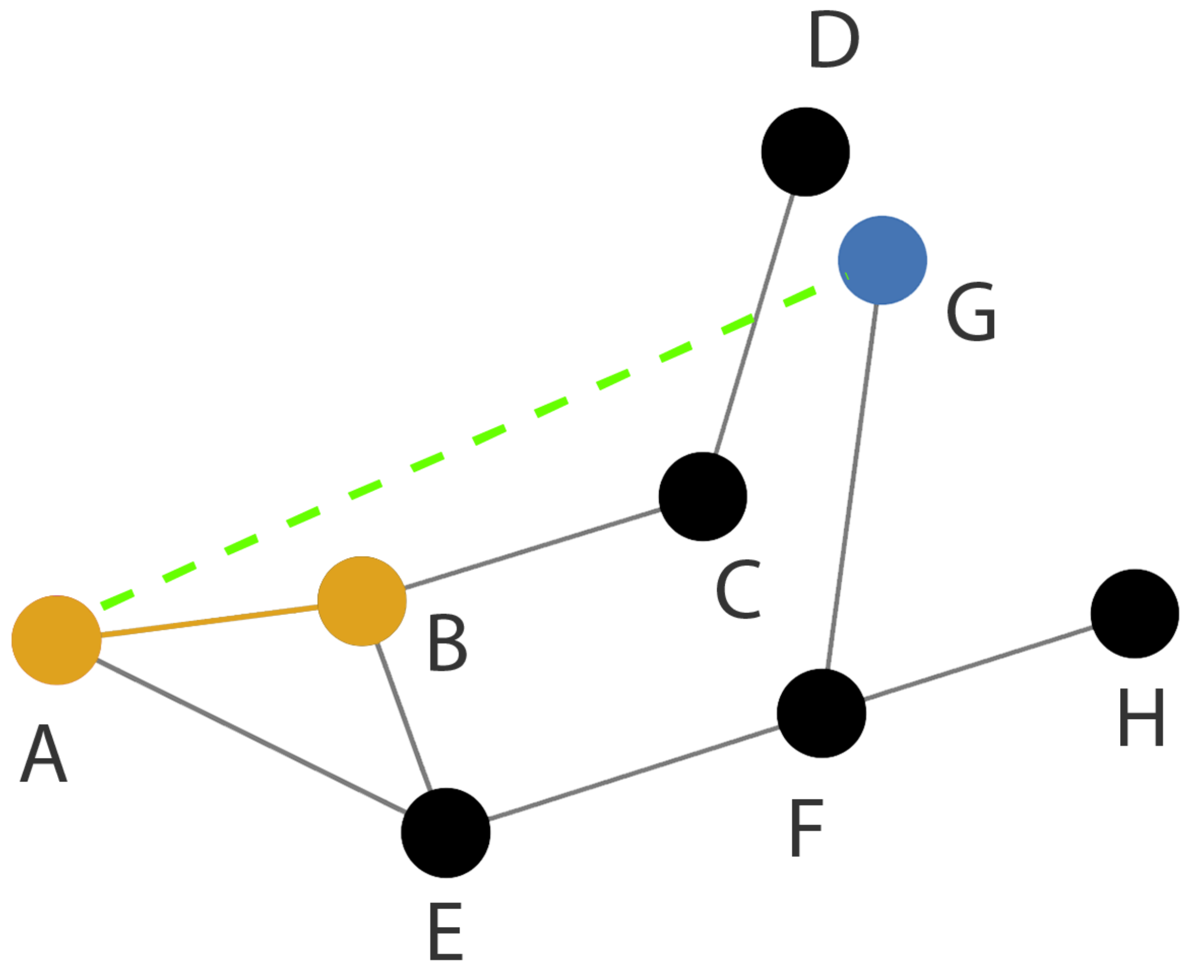


Image redrawn from Huang et al. 2013

Geodesic tendency

1. A B C D
   Doesn't pan out, try again

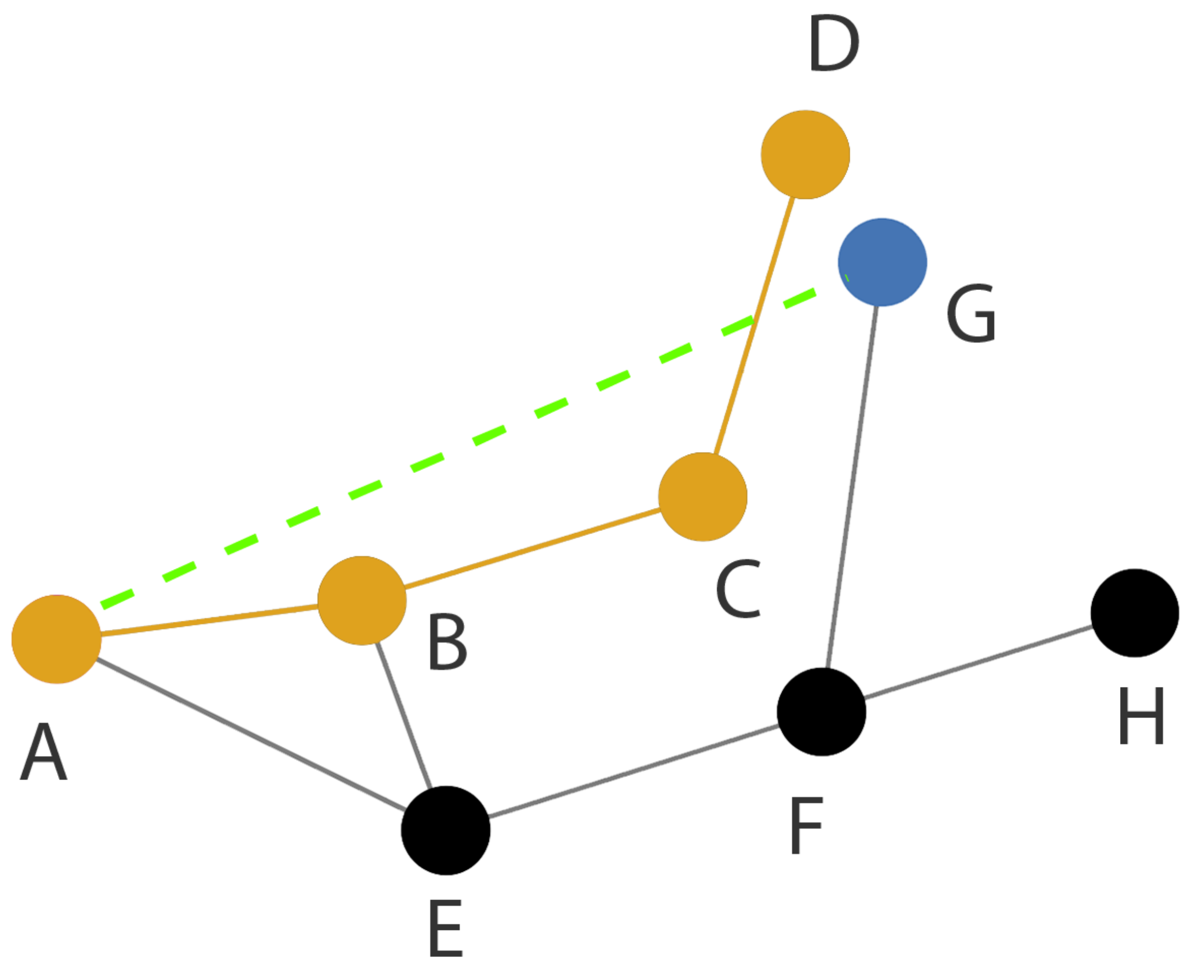


Image redrawn from Huang et al. 2013

Geodesic tendency

2. Next try, diverge further from geodesic:

A E

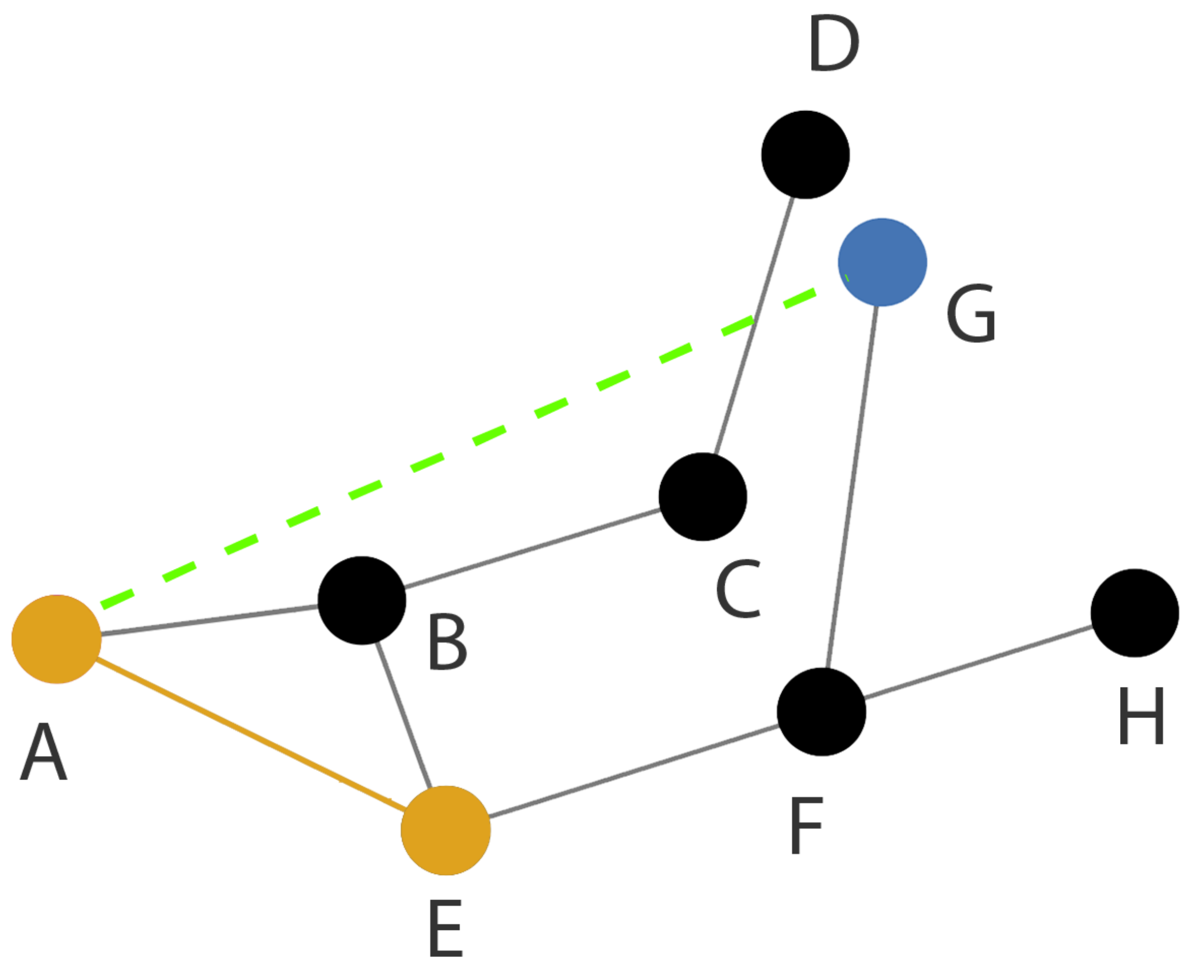


Image redrawn from Huang et al. 2013

Geodesic tendency

2. A E F G

Success!

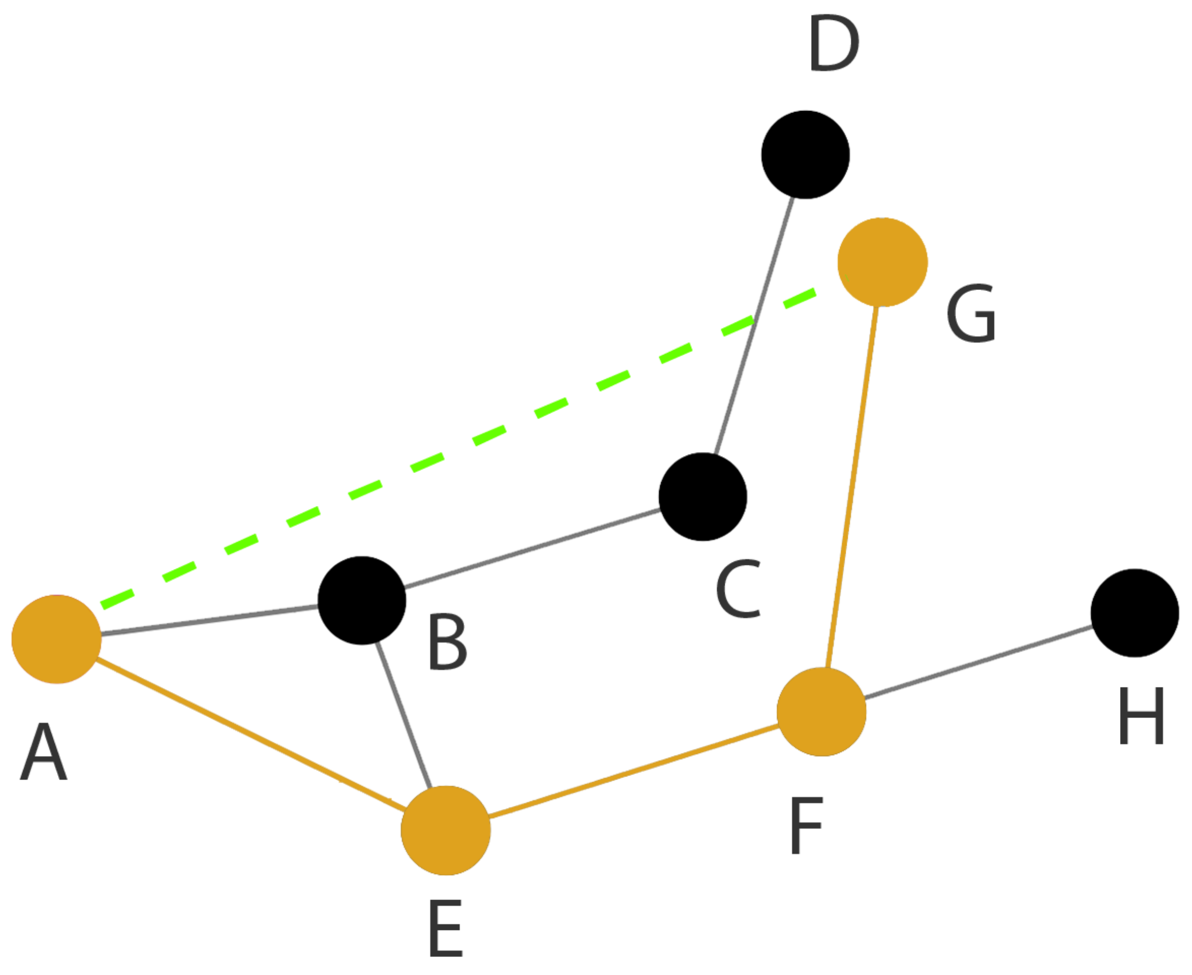


Image redrawn from Huang et al. 2013

Geodesic tendency

Set of likely paths searched:

1. A B C D
2. A E F G

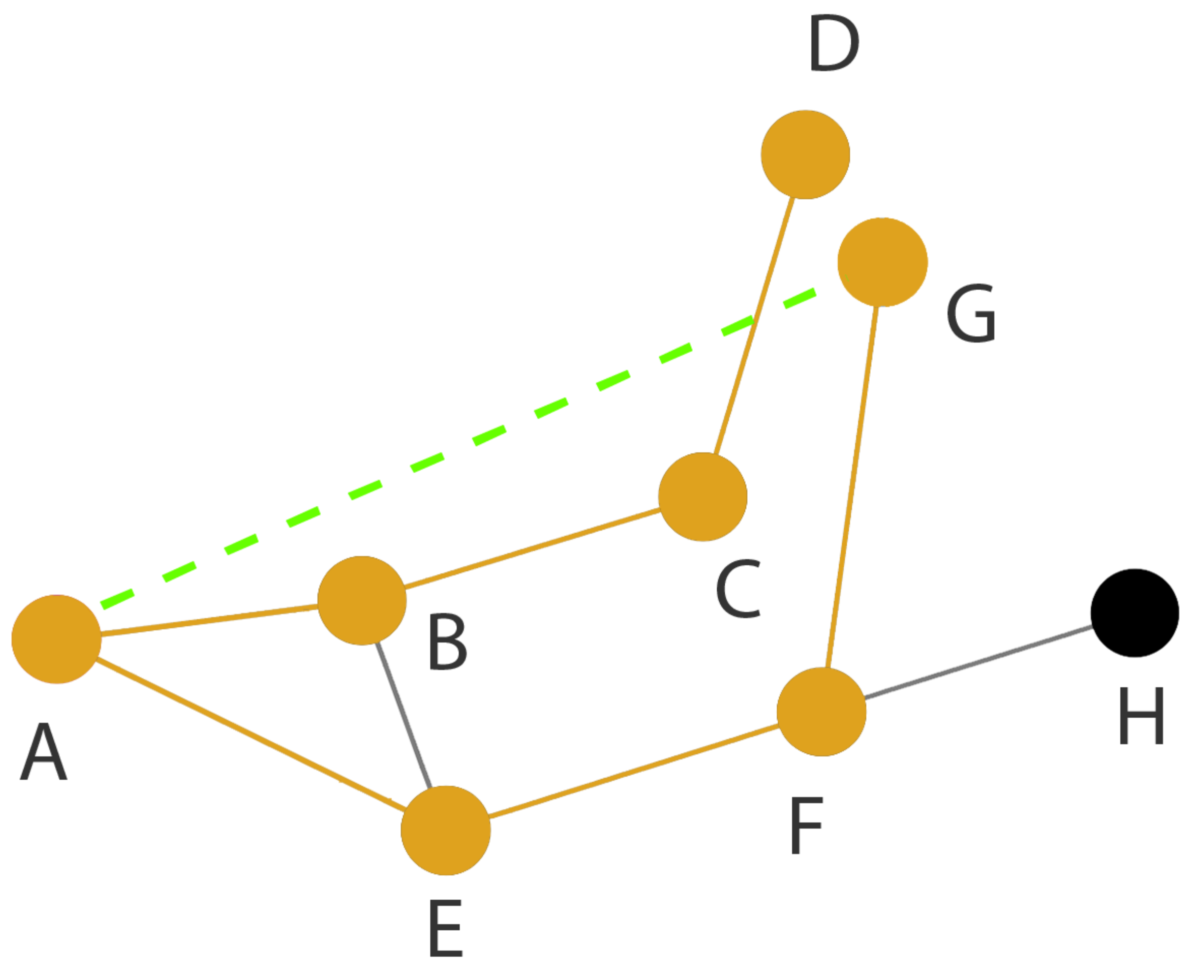


Image redrawn from Huang et al. 2013

Geodesic tendency

But our early piloting showed geodesic tendency only part of story…

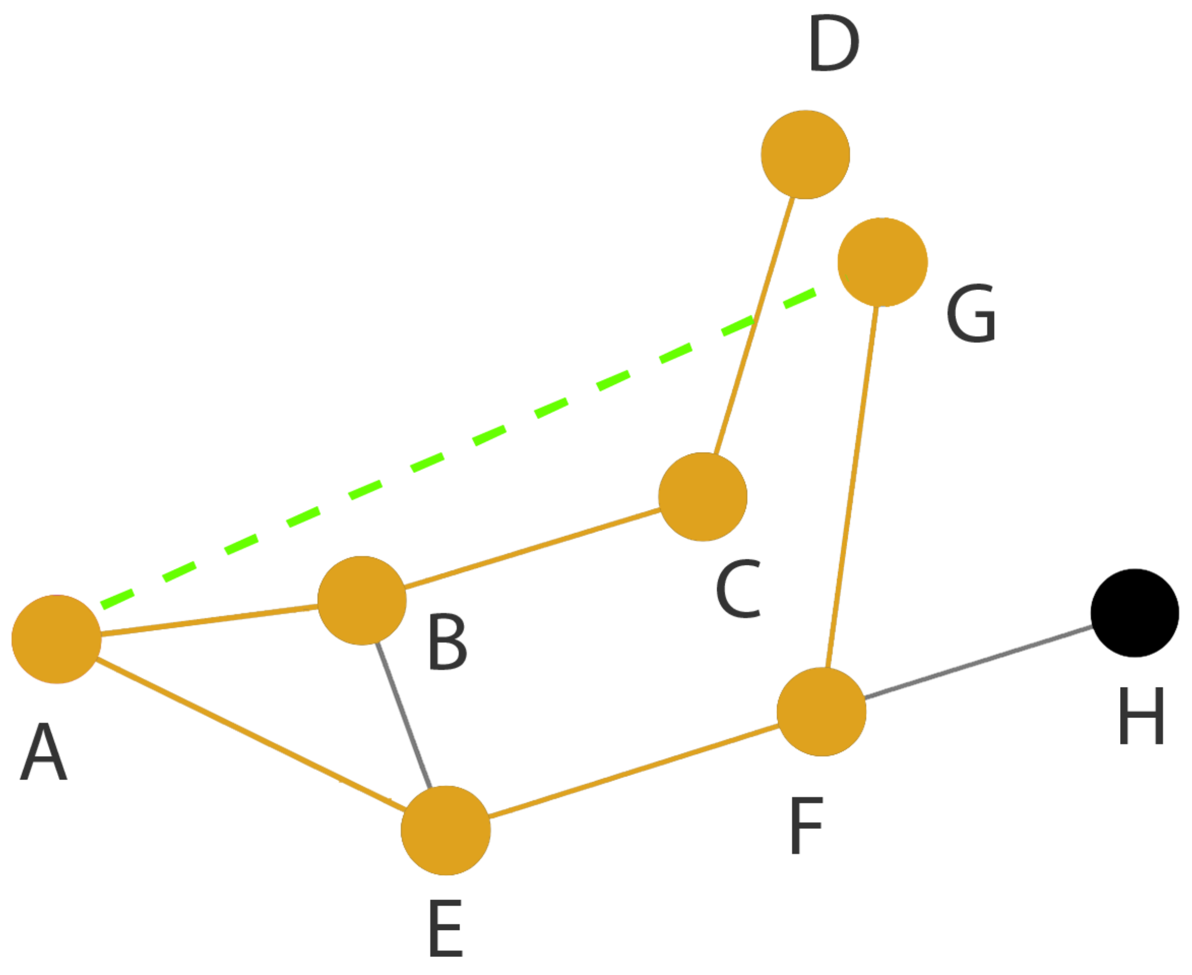


Image redrawn from Huang et al. 2013

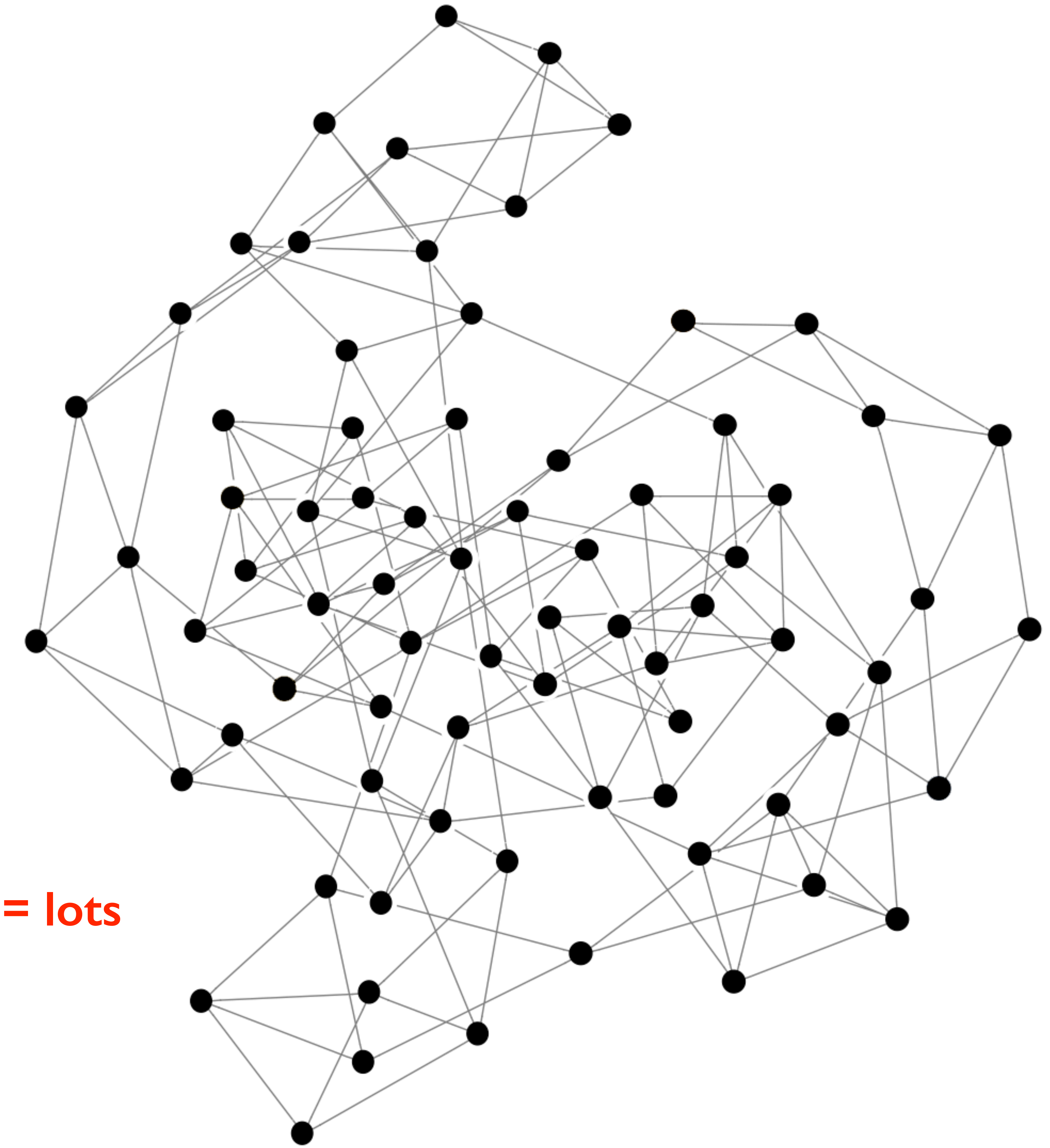
Can layout quality provide an answer?

- layout quality in graph drawing judged with **quantitative *readability* metrics**:
  - minimize edge-edge crossings, minimize total edge lengths, maximize angular resolution of edges at nodes, ...
- early algorithmic work based on metrics easy to compute
  - typically used in optimization context
  - derived through introspection, assumed to be appropriate
- subsequent empirical work investigated how metrics impact graph readability for humans
  - controlled experiments in lab setting
    [Purchase et al, 1995] [Purchase, 1997] [Purchase, 2002] [Körner, 2004] [Huang et al, 2005] [van Ham & Rogowitz, 2008] [Dwyer et al, 2009] [Huang, 2011] [Huang & Huang, 2011] [Körner, 2011] [Purchase et al, 2012] …
  - despite mixed findings, edge-edge crossings often considered as most important

Global vs local metrics

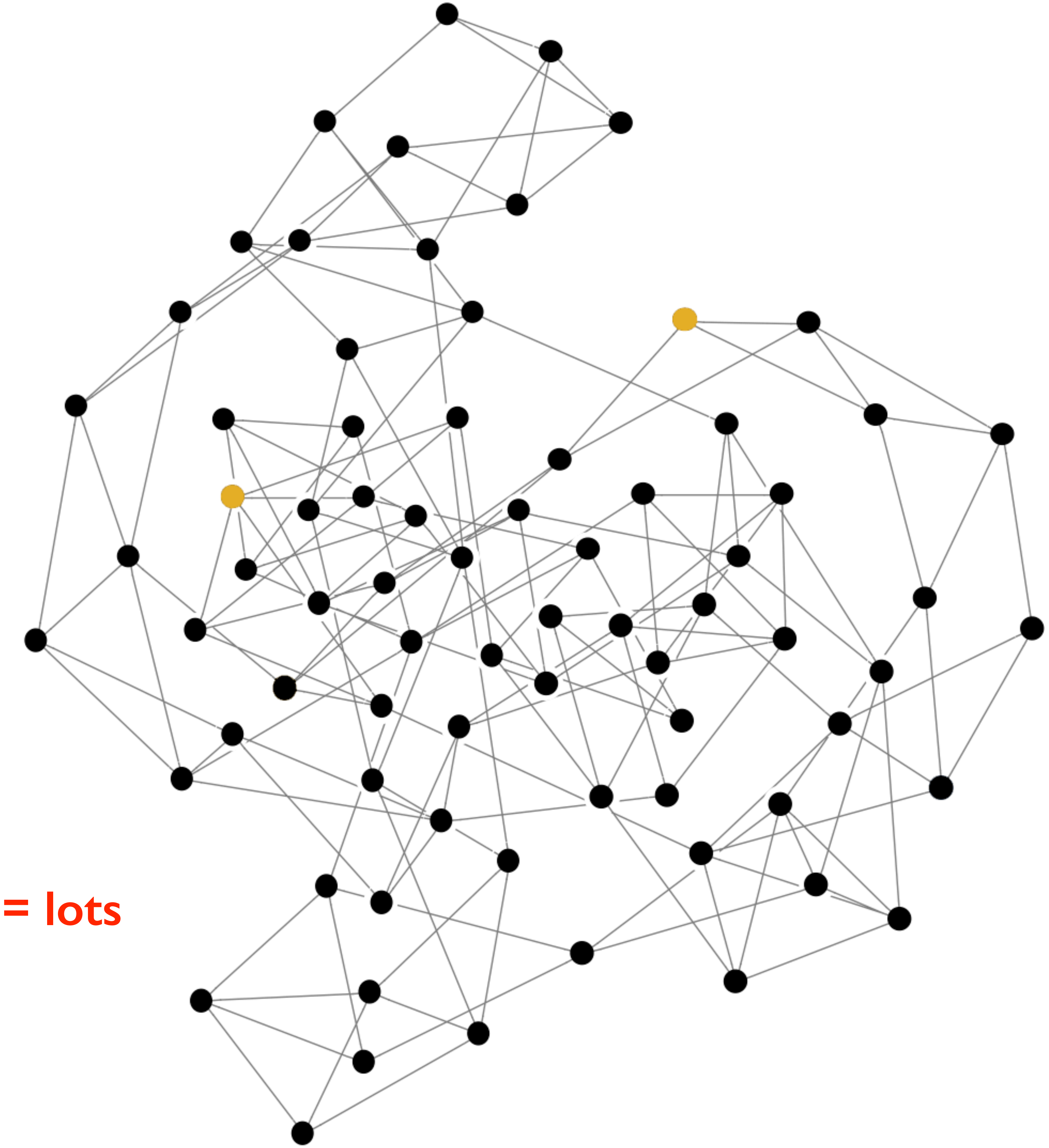
- poorly understood:
  when is one path harder to follow than another?
- metrics typically used and evaluated globally

global edge-edge crossings = lots



Global vs local metrics

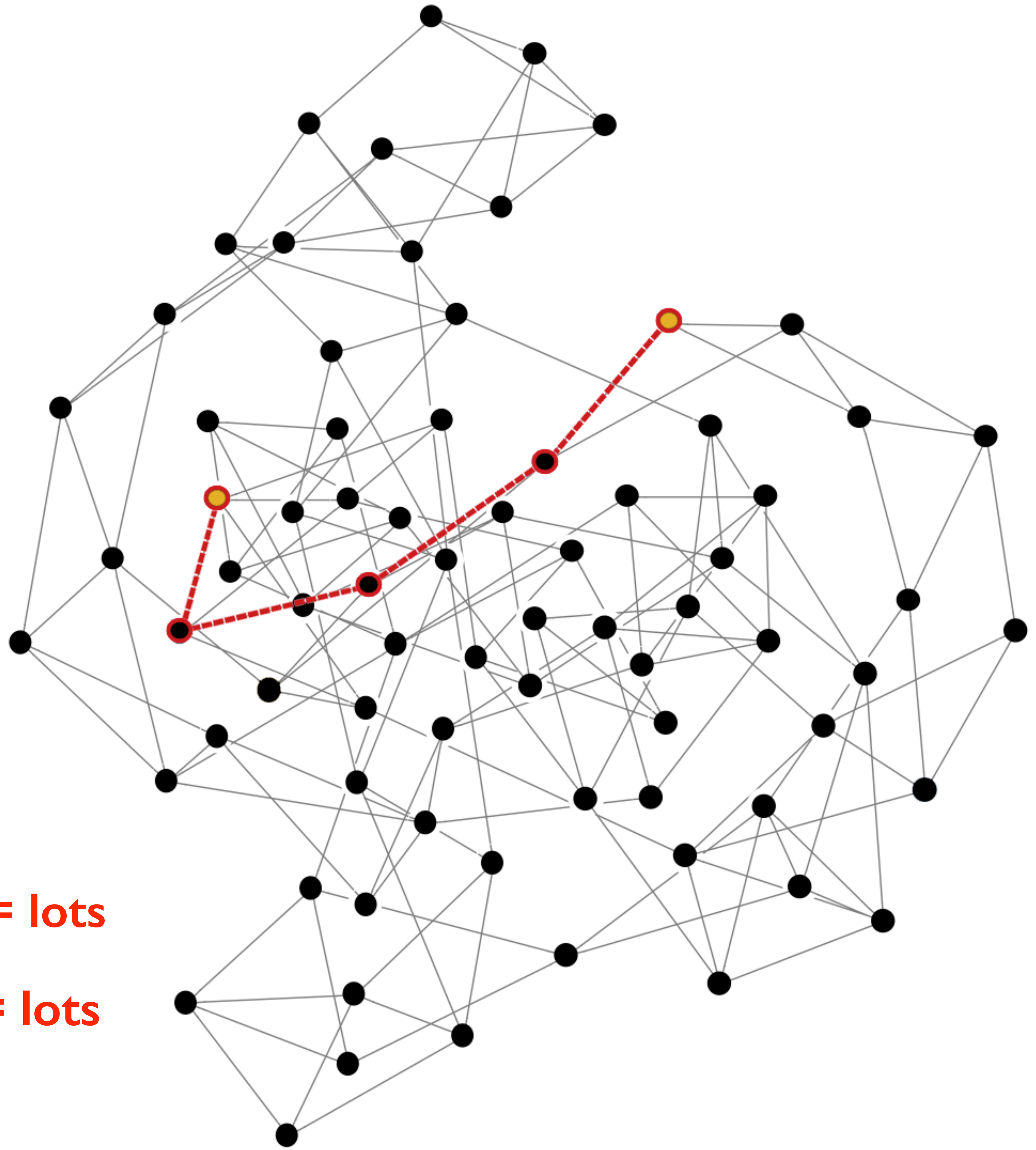
- poorly understood: when is one path harder to follow than another?
- metrics typically used and evaluated globally



global edge-edge crossings = lots

Global vs local metrics

- poorly understood:
  when is one path harder to follow than another?
- metrics typically used and evaluated globally

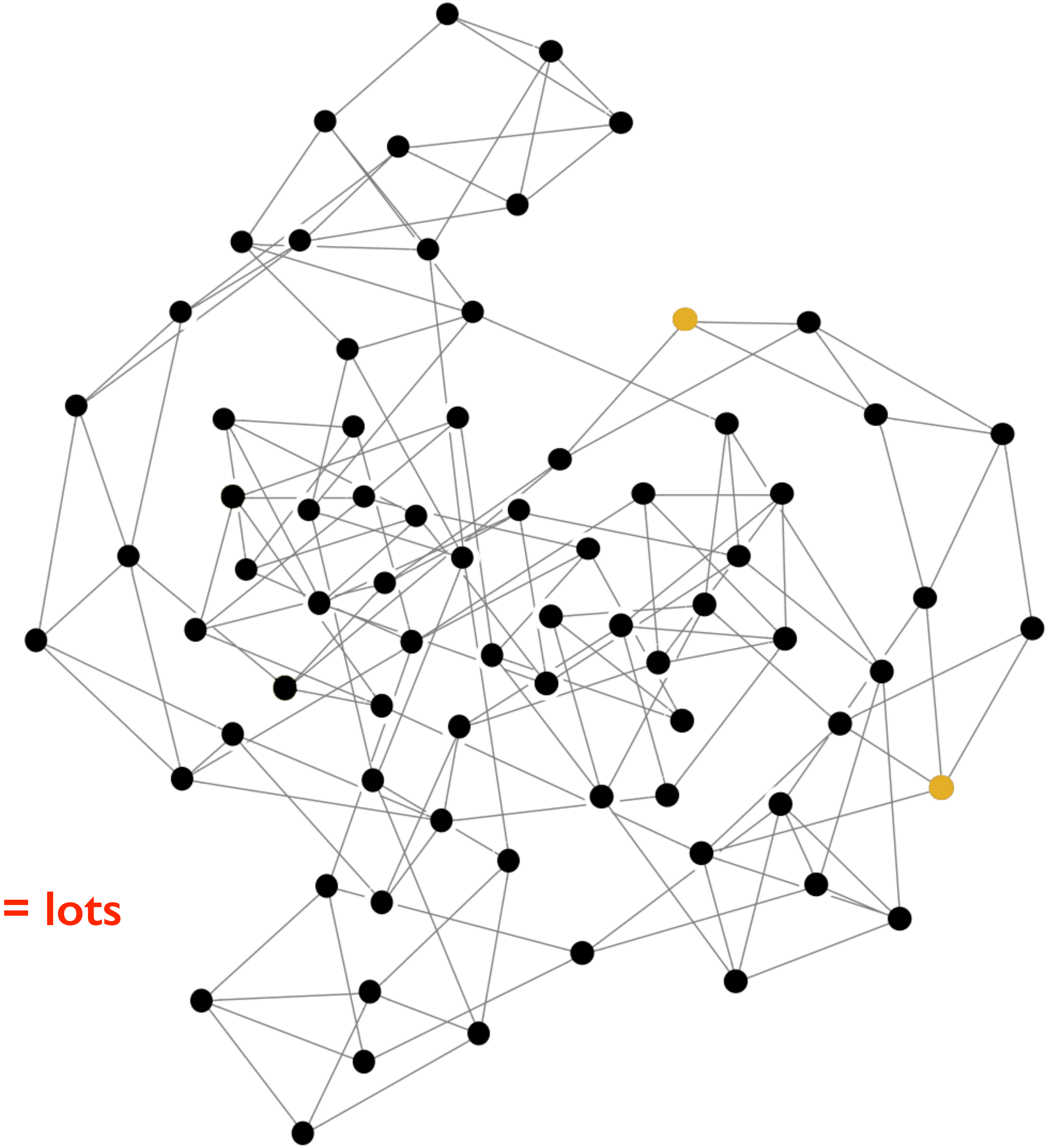


global edge-edge crossings = lots

local edge-edge crossings = lots

Global vs local metrics

- poorly understood:
  when is one path harder to follow than another?
- metrics typically used and evaluated globally



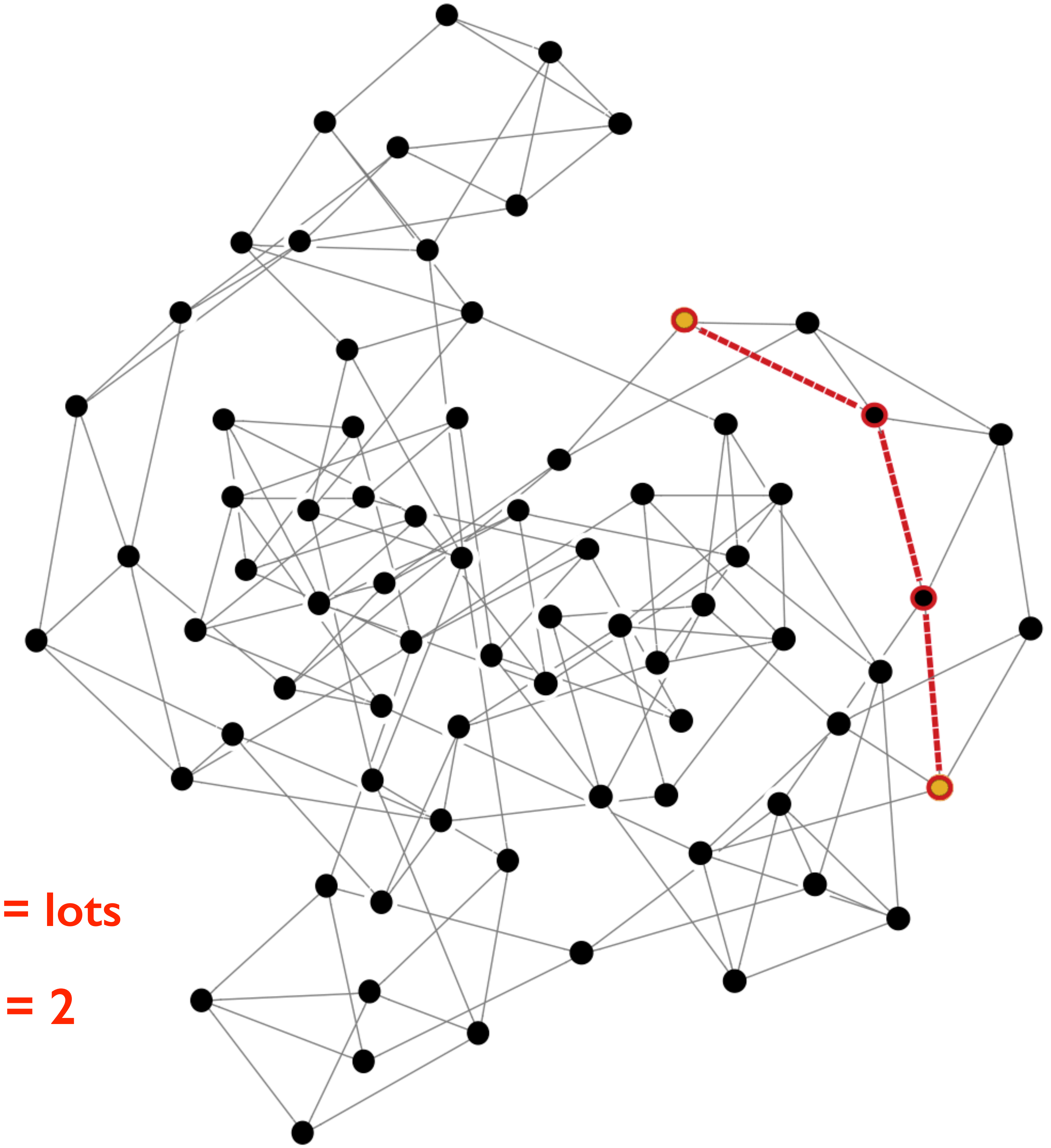
global edge-edge crossings = lots

Global vs local metrics

- poorly understood:
  when is one path harder to follow than another?
- metrics typically used and evaluated globally

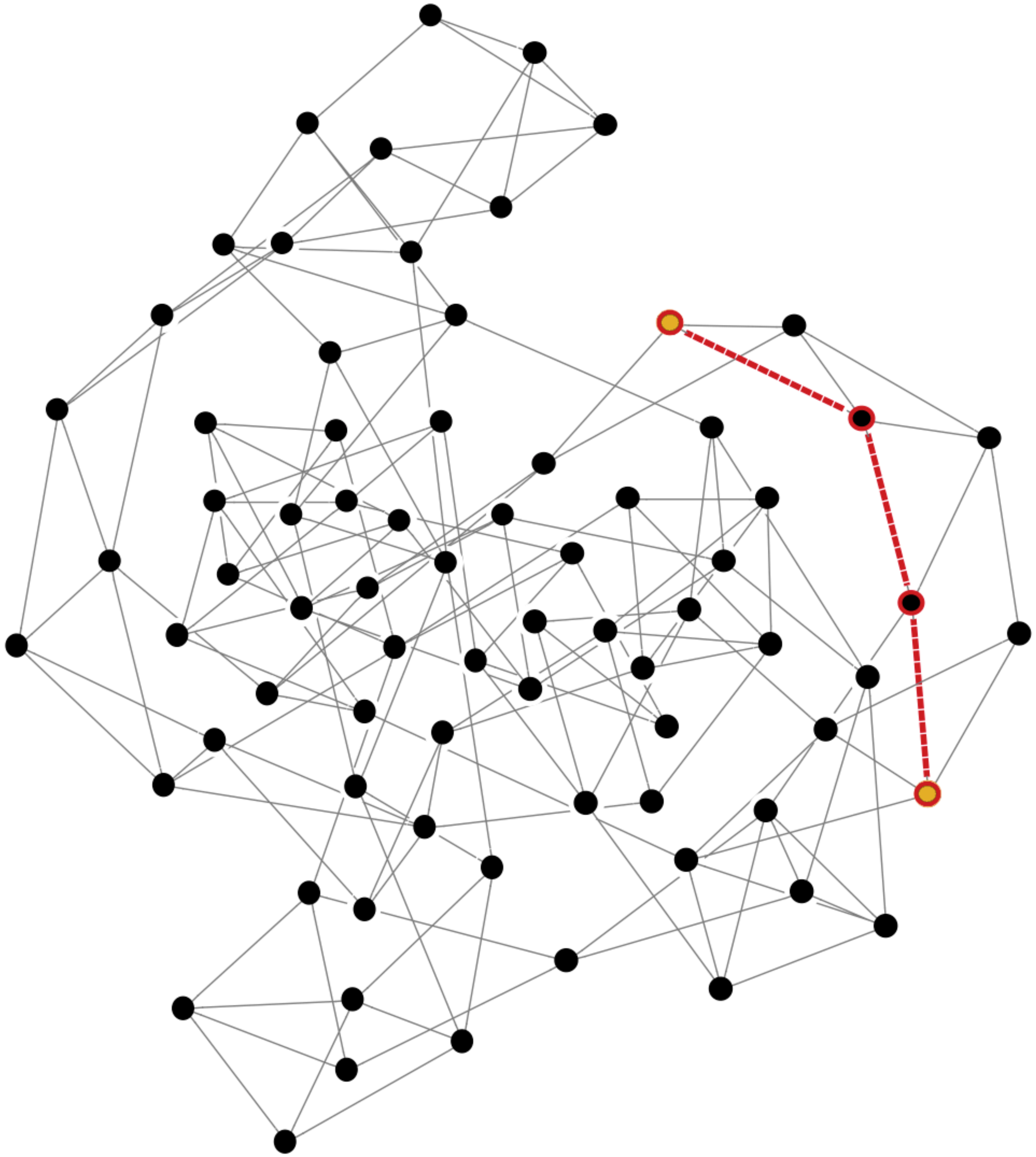
global edge-edge crossings = lots

local edge-edge crossings = 2



Global vs local metrics

- poorly understood:
  when is one path harder to follow than another?
- metrics typically used and evaluated globally
- finding: metrics along **local *solution path*** were much better predictors of difficulty
  [Ware, Purchase, Colpoys, McGill 2002. Cognitive Measurements of Graph Aesthetics]



Our "Goldilocks" observation

- global computation often takes too much into account
- but computing only along solution path may take too little into account!
  - overly local: does not account for everything relevant to task
- what would be just right?
  - measure metrics on the full set of paths a user searches while completing a task!
- we identified novel goal
  - predict set of paths that a user is likely to search while path tracing: **search set**
  - would be good for
    - designing new interaction techniques & automatic graph layout algorithms
    - characterizing how users read graphs
    - improving measurement of metrics that affect graph readability

Multi-stage project

- *introduce concept of the search set*
- observational study:
  - quantitative data collection
  - qualitative analysis: open coding observational video of path tracing on "training" data
  - result: detailed characterization of path tracing behaviours
- model development: a predictive model of a search set
  - algorithmic implementation
  - quantitative assessment (preliminary)
- quantitative study:
  - use search set to measure metrics that affect graph readability
  - quantitative assessment: multiple regression analysis on (reserved) test data

The search set concept: Research questions

- (Q1) can we identify distinct path tracing behaviours?
- (Q2) how common are these behaviours?
- (Q3) can we predict a search set based on these behaviours?
- (Q4) how much improvement from measuring metrics on search set?

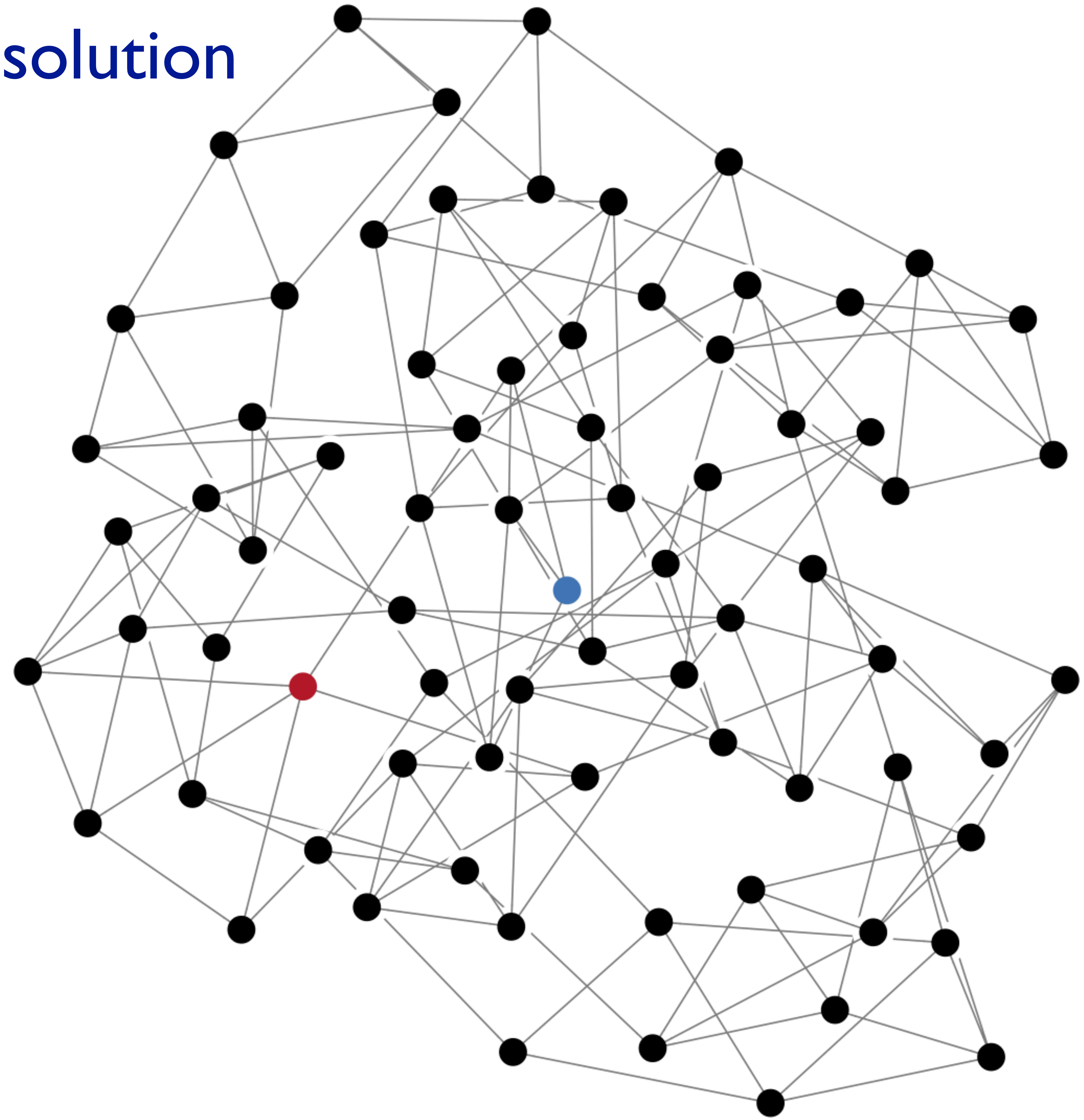
Search Set Case Study: Qualitative Study

Observational user study

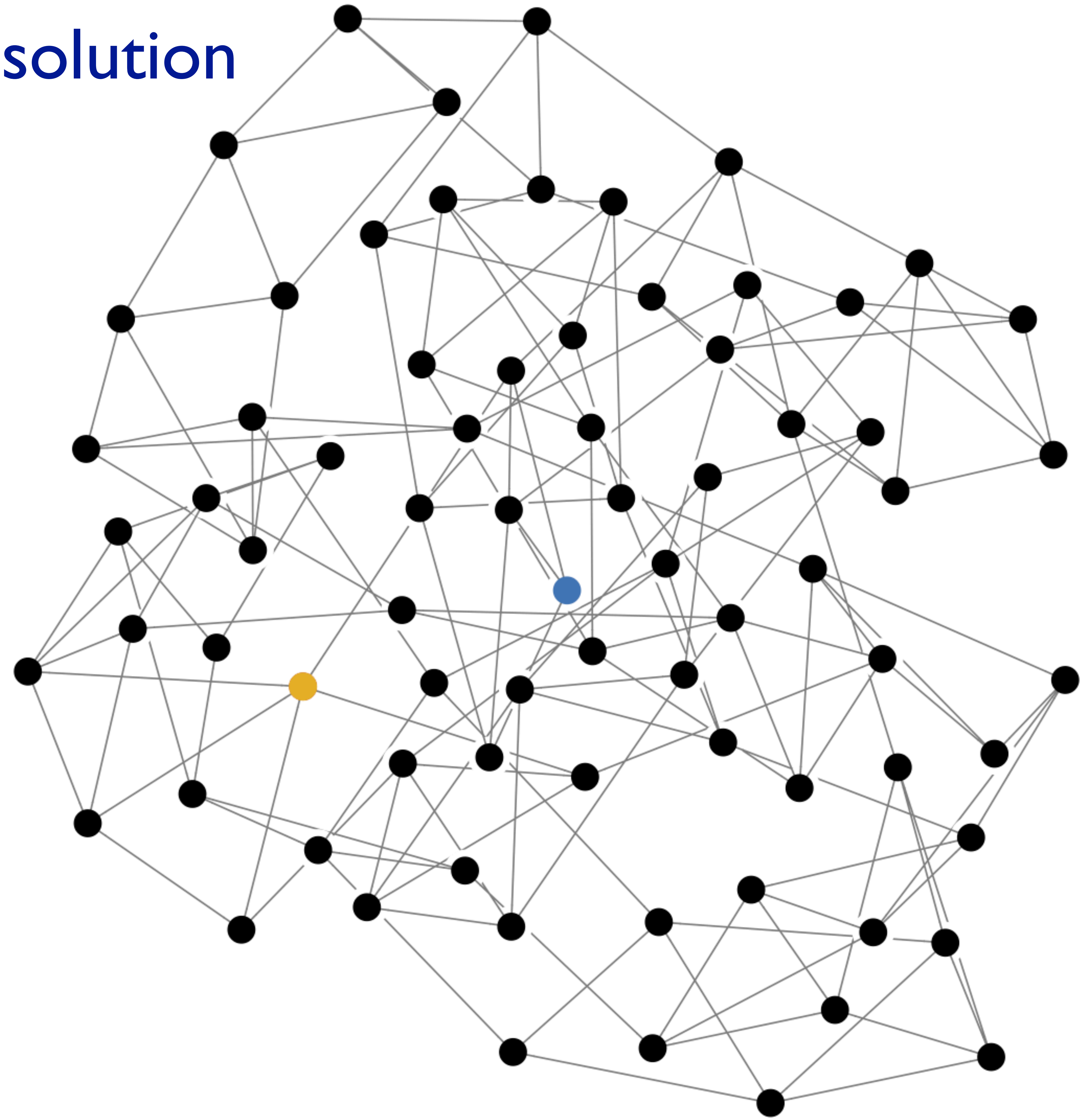
- 12 participants
- interface:
  graphs displayed on Cintiq tablet
- primary task:
  find shortest path between
  red and blue nodes
- secondary task:
  trace progress: hover nodes with tablet pen

- 144 trials, split into two sessions (~1.5 hours each)
  - 1 unique graph shown per trial
- one shortest path in each graph
- two phases: 1) find then 2) demonstrate solution path

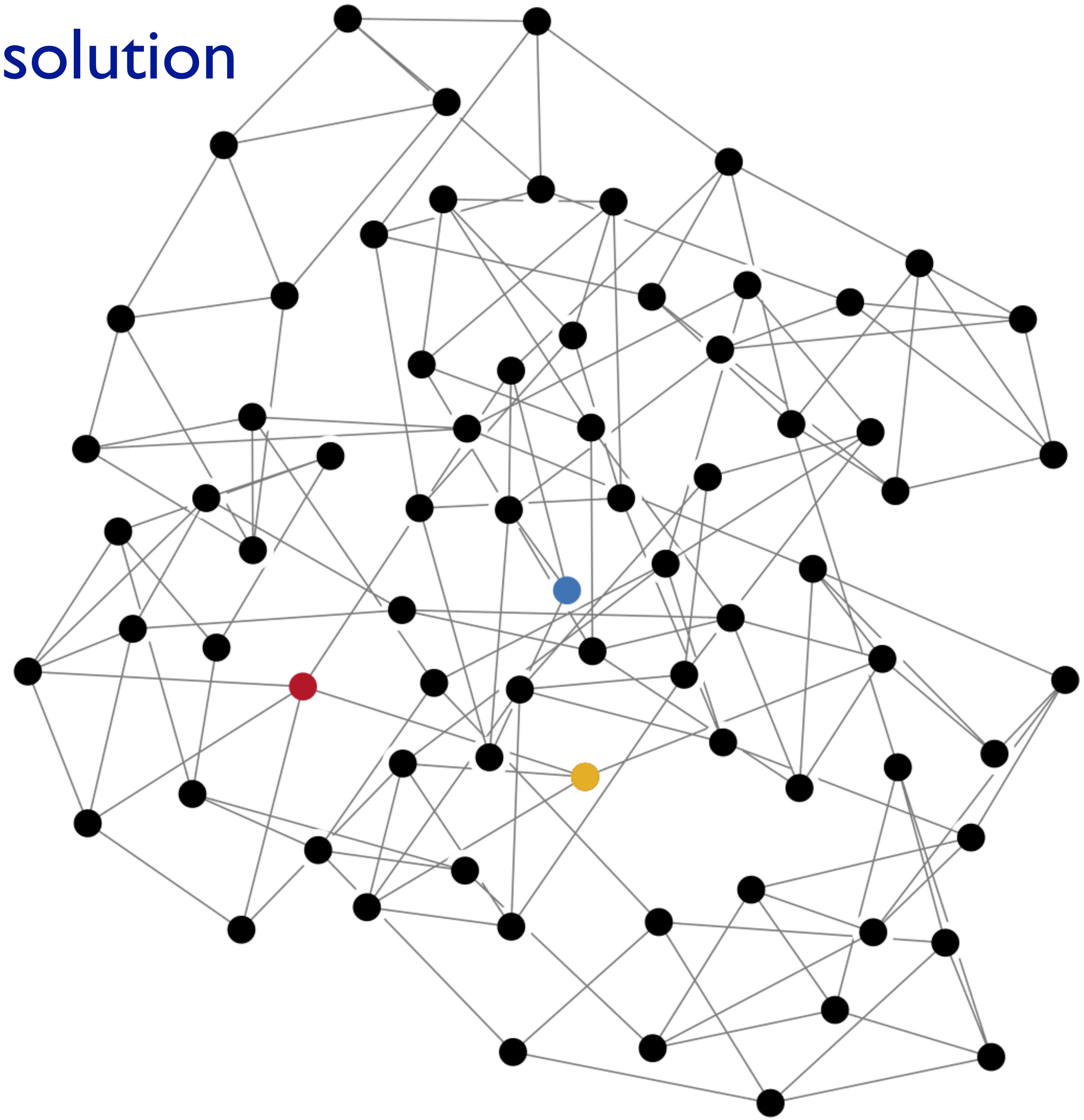
Searching for solution



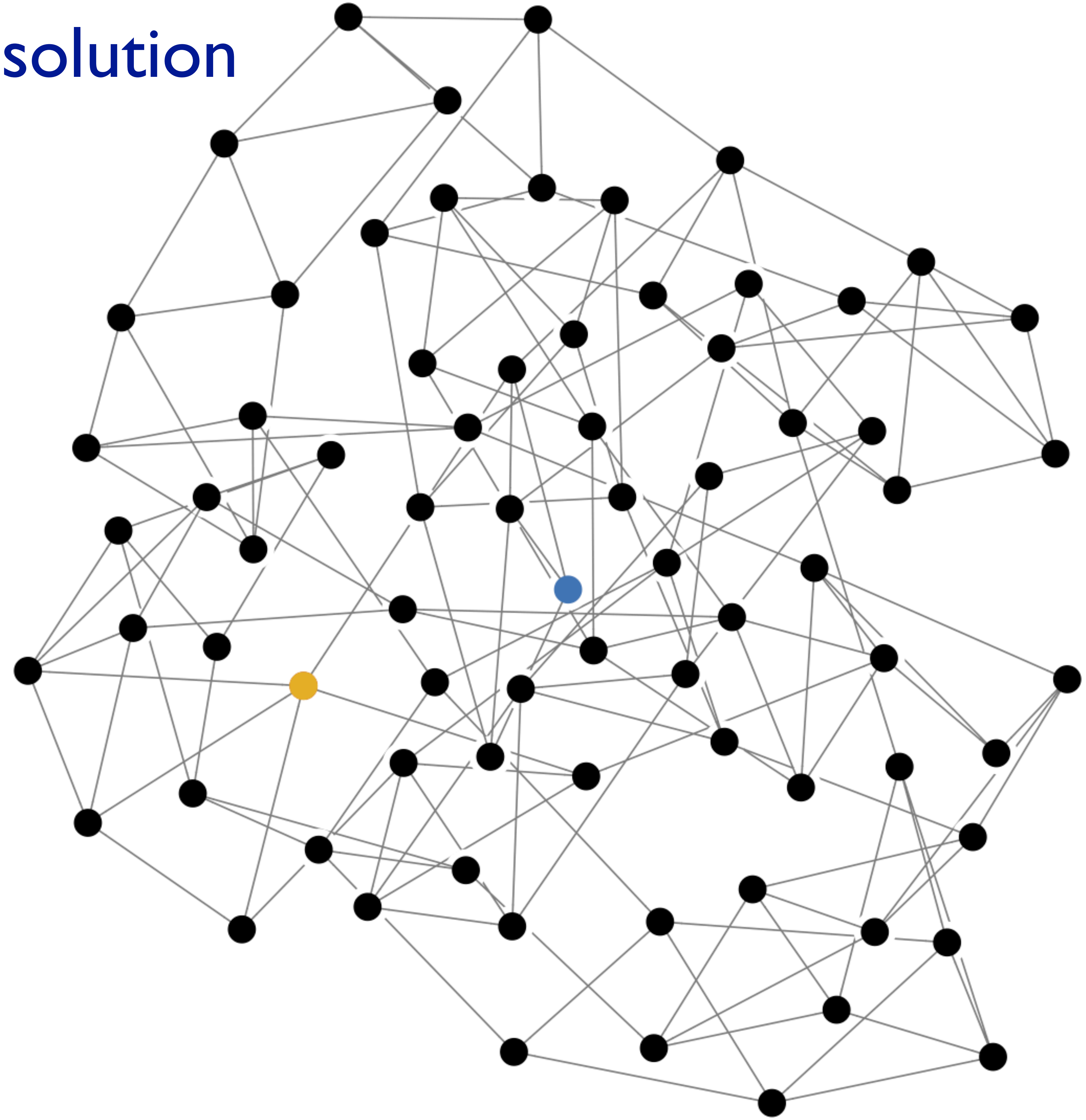
Searching for solution



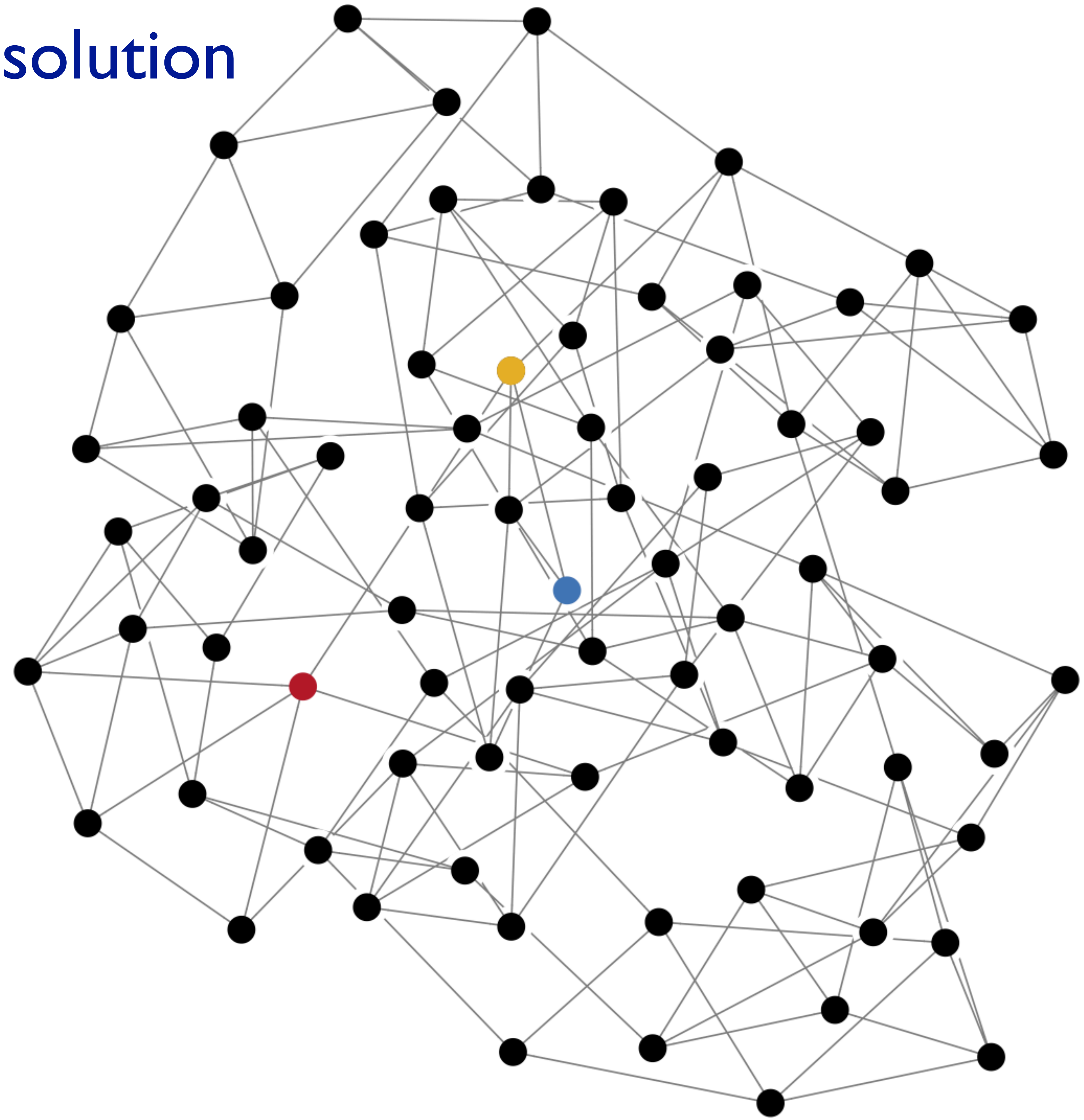
Searching for solution



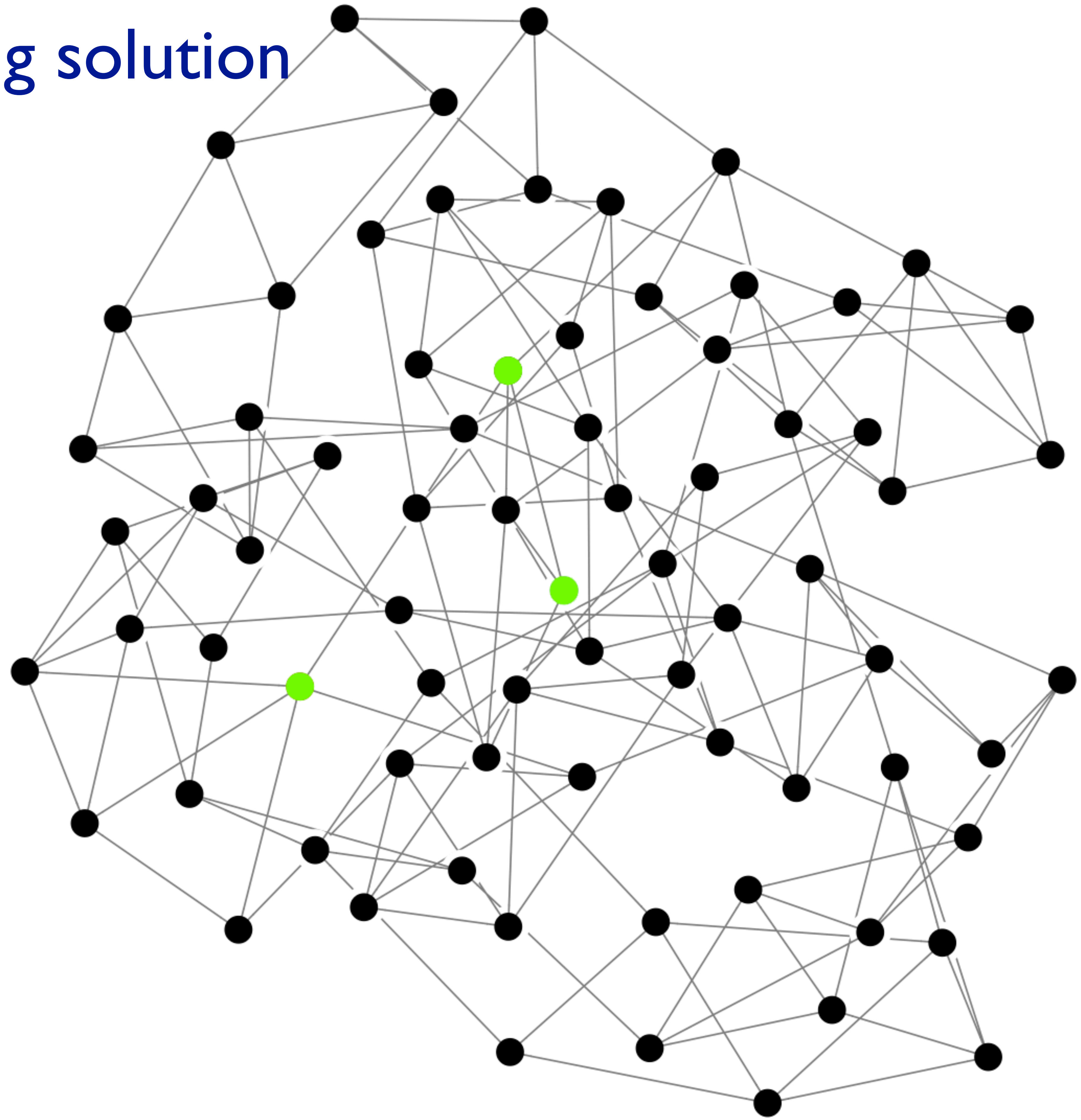
Searching for solution



Searching for solution



Demonstrating solution

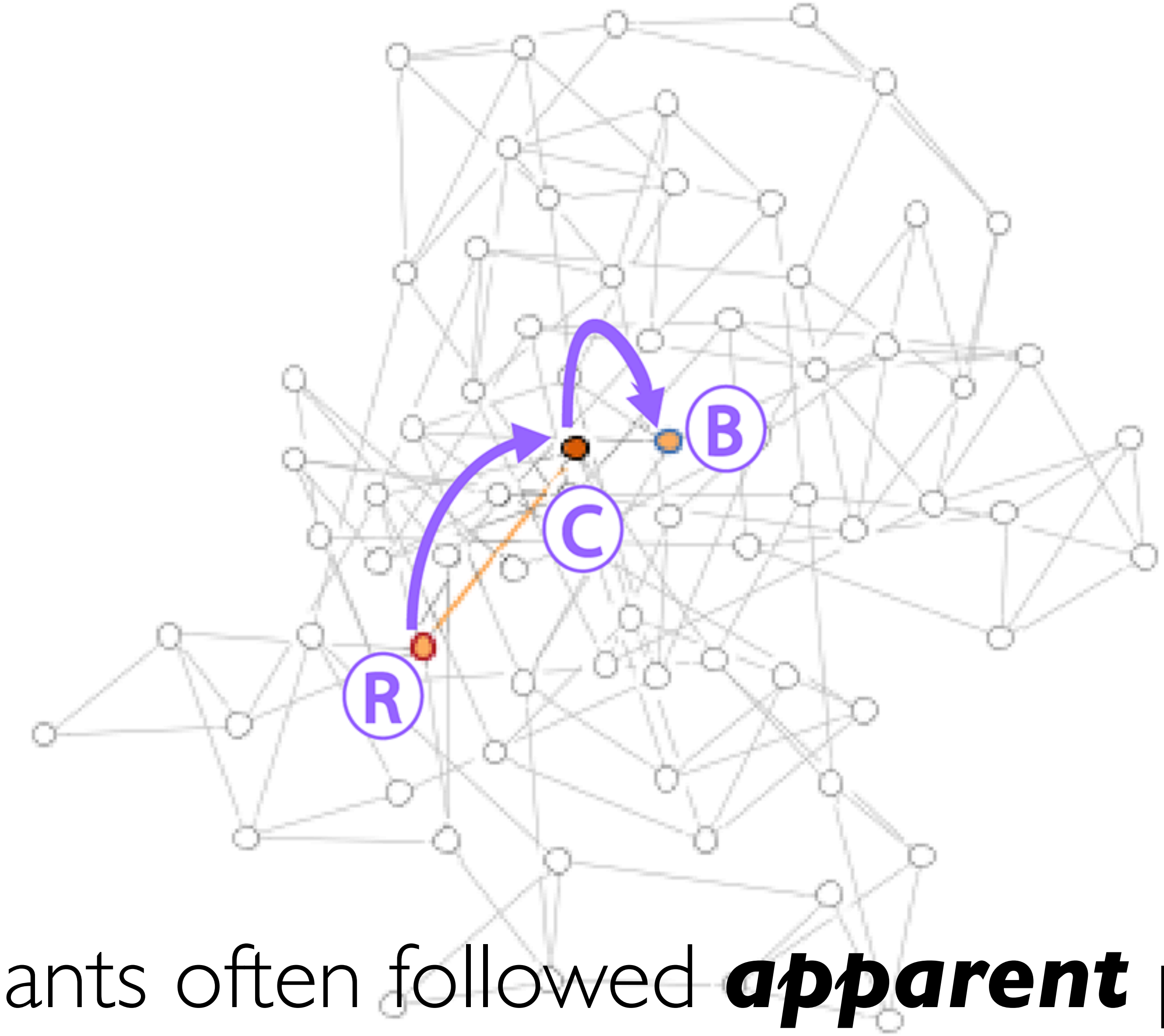


Observational user study

- primary quantitative collected data
  - Sequences of node hovers along paths for each trial
  - Response time to complete trial
  - Error rate (correct/incorrect solution path)
- analysis approach: split into three parts
  - qualitative analysis of path tracing behaviors
    - for "training" data
  - developing a predictive search set model and algorithmically instantiating it
  - multiple regression analysis comparing metrics with/without search set
    - on reserved test data

Qualitative analysis: Method

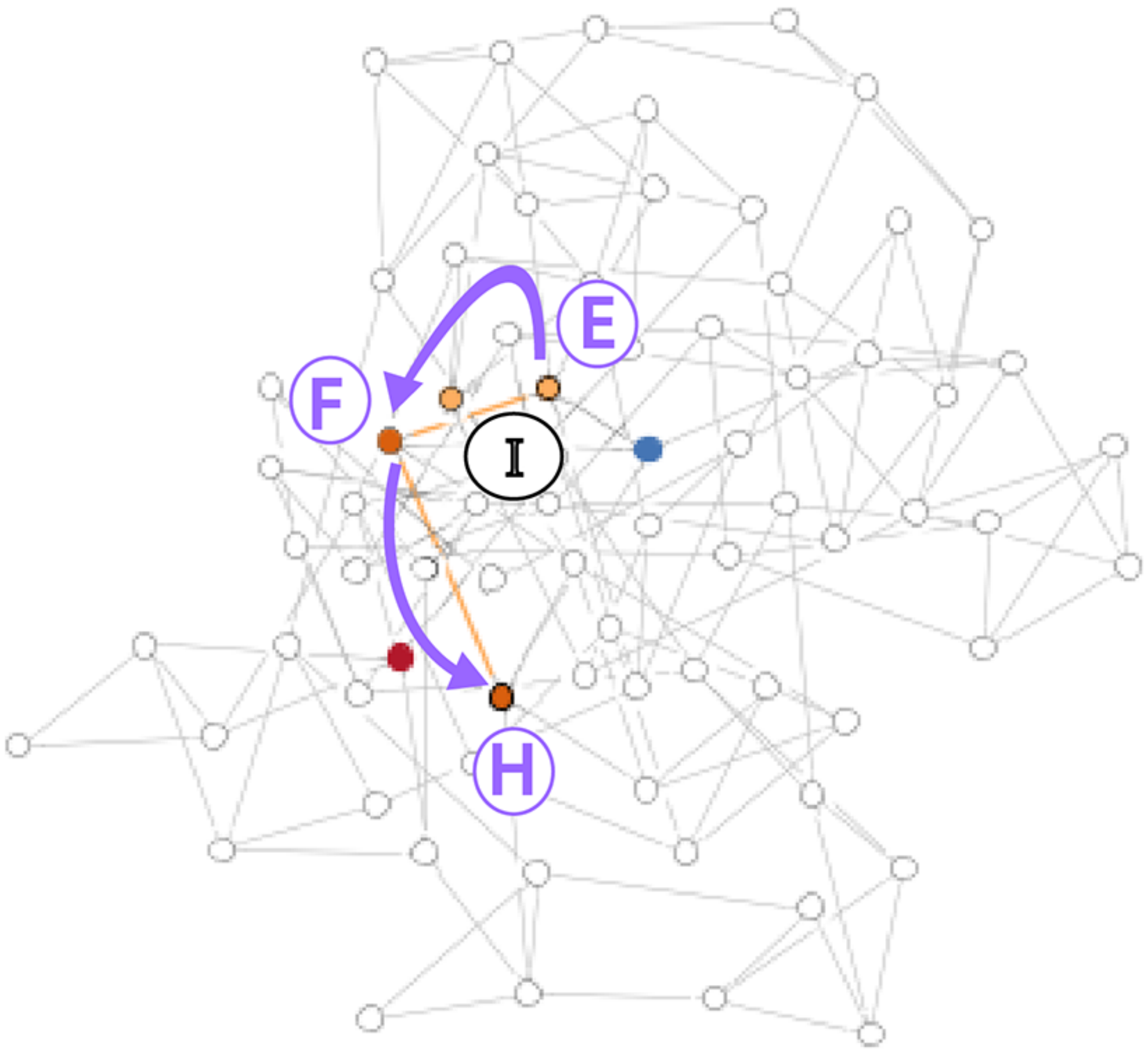
Manually coded paths because…



…participants often followed ***apparent*** paths

Qualitative analysis: Method

Manually coded paths because…

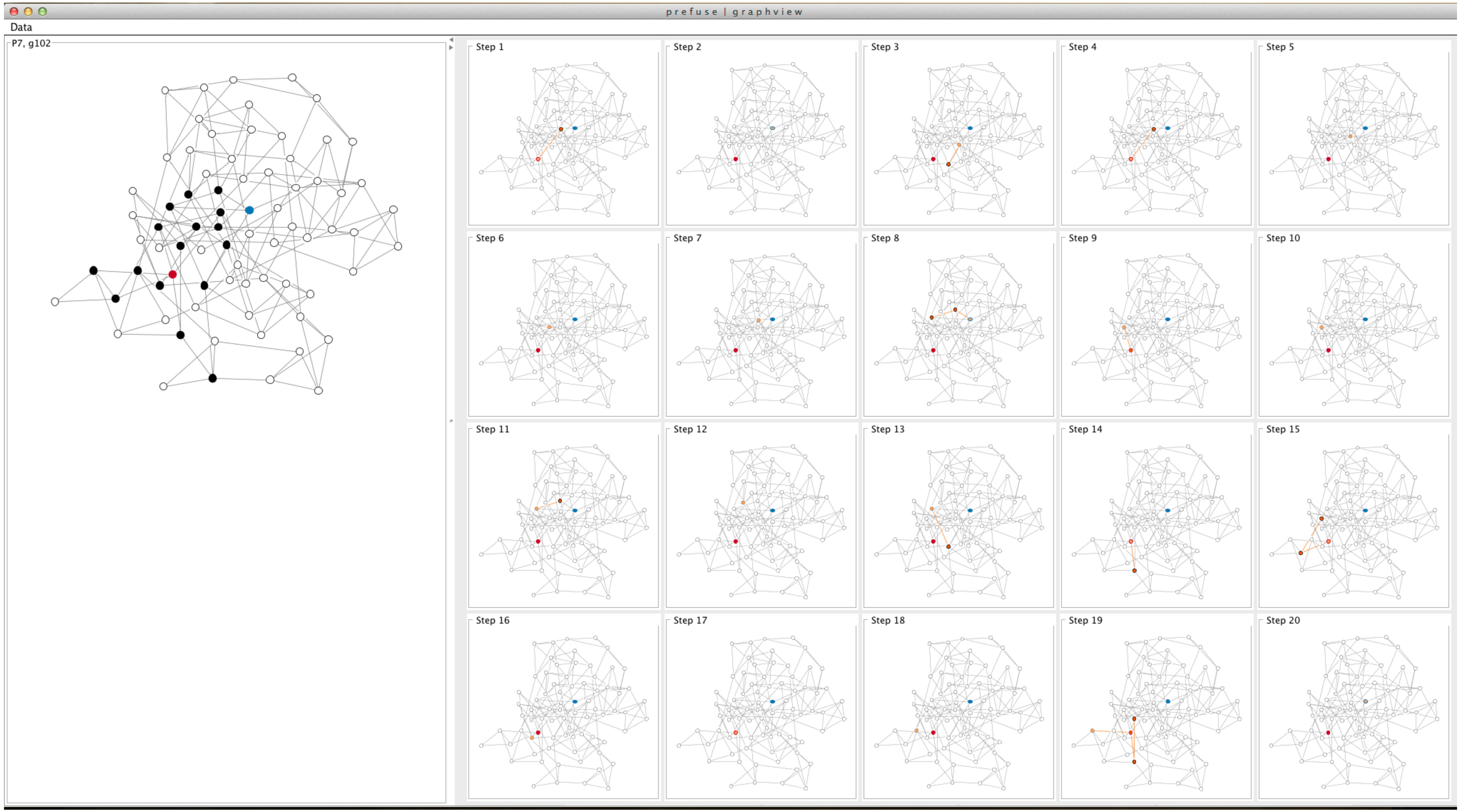


… some nodes were just in the way

Qualitative analysis: Method

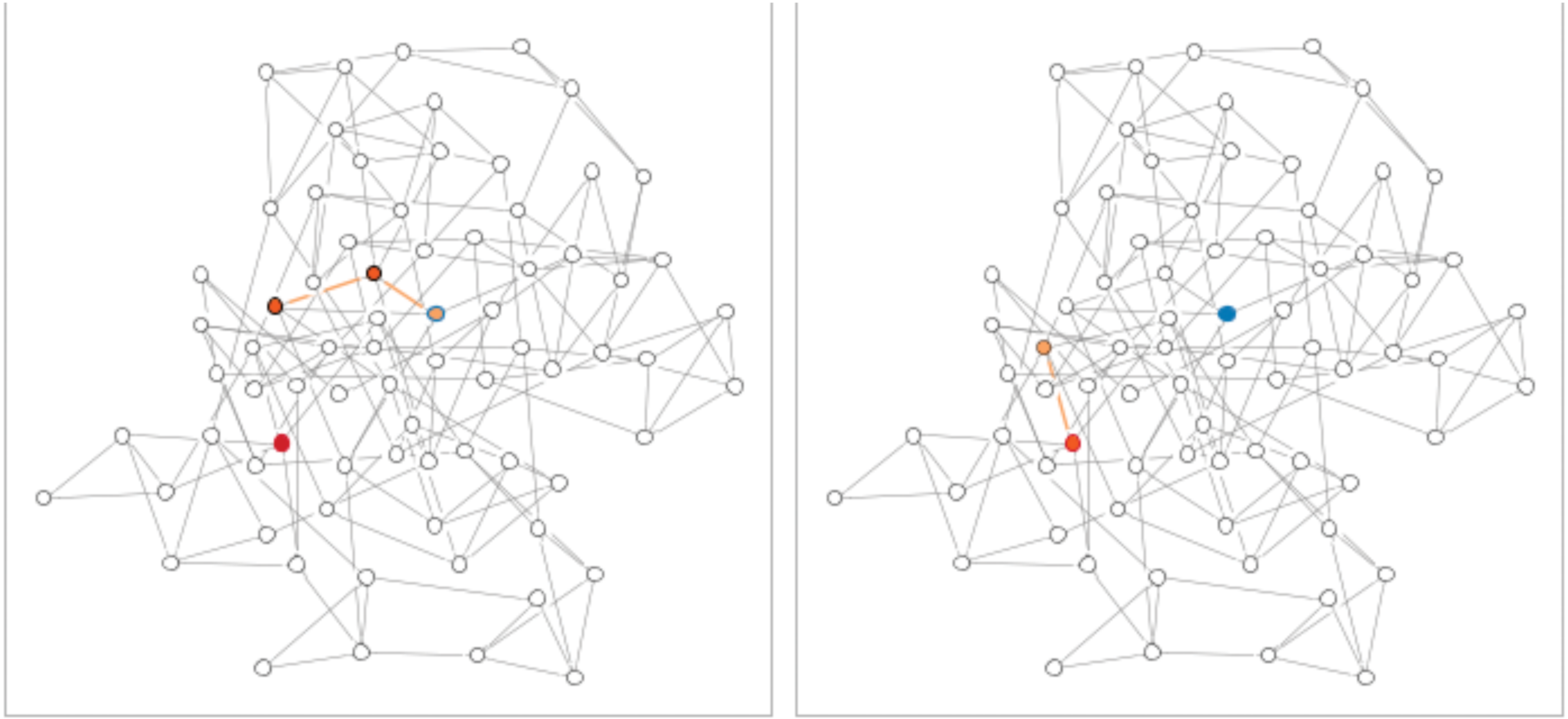
- training set of 24 study graphs analyzed
  - reserved other 120 graphs as validation set
  - 12 participant trials per graph
  - for a total of 288 trials coded
- one investigator performed this coding solo
  - with some automatic support via visualization interface

Visualization interface for qualitative coding



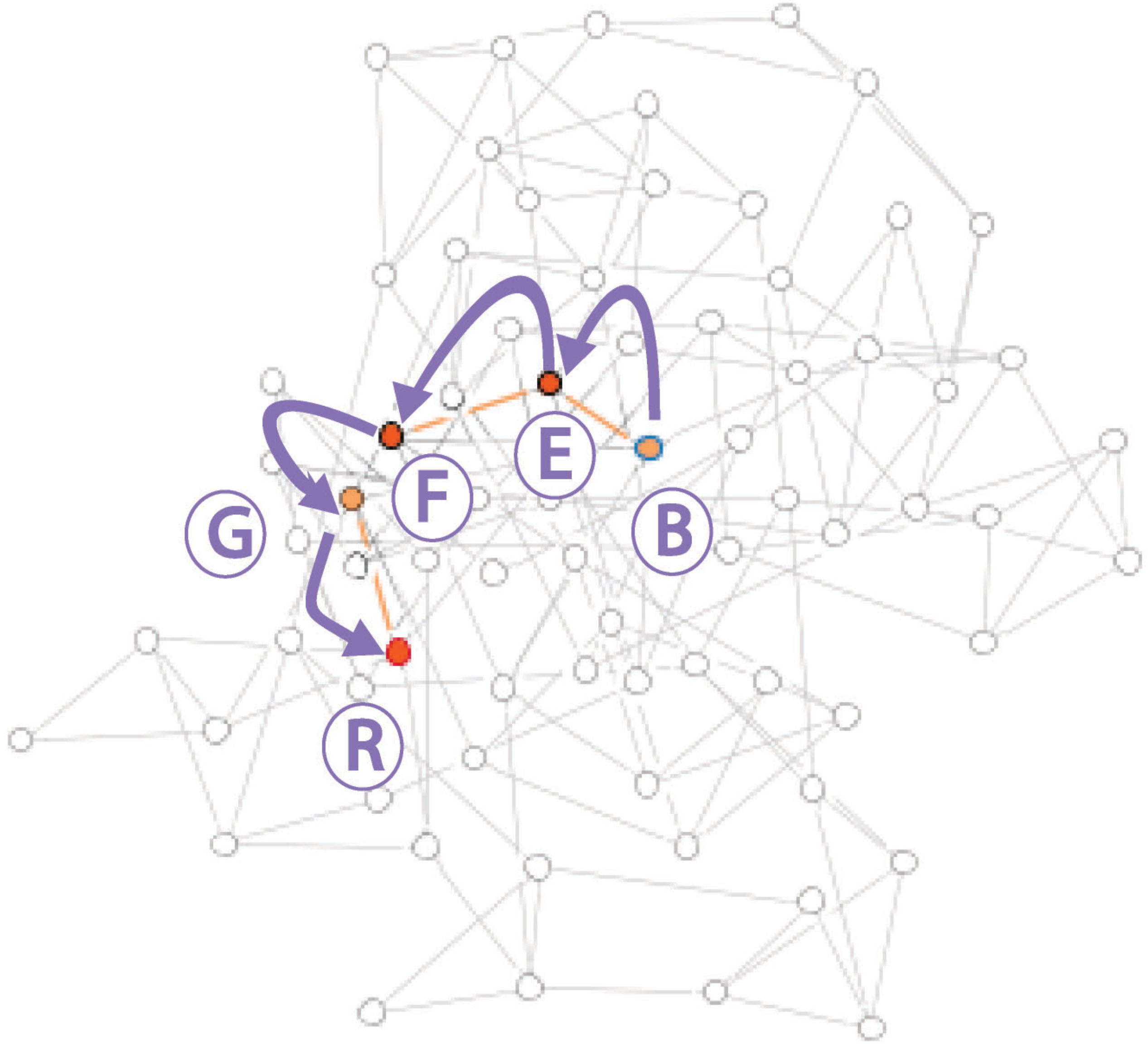
Visualization interface for qualitative coding

- Investigator looked at sequences of hovers . . .



Visualization interface for qualitative coding

- And created textual descriptions of full paths



Qualitative analysis

- many path dimensions recorded
  - anchor nodes where paths starts
  - target nodes that paths go towards
  - is a hop the closest to geodesic?
  - …
- also coded other interesting phenomenon
  - jumps between nodes
  - checks of node-edge crossings
  - …

Qualitative analysis: Key results

- It is possible to identify distinct path tracing behaviours (Q1)? Yes
  - investigator classified 96% of data examined with at least one code
- Many common path tracing behaviours emerged from coding (Q2)
  - use of both topological and apparent paths
  - repeated exploration of paths
  - when participants stop following paths
  - choice of nodes to search out from
  - interactions of geodesic tendency with continuity
  - prevalence of the geodesic tendency
  - likely directions for the first hop in a path

Qualitative analysis: Key results

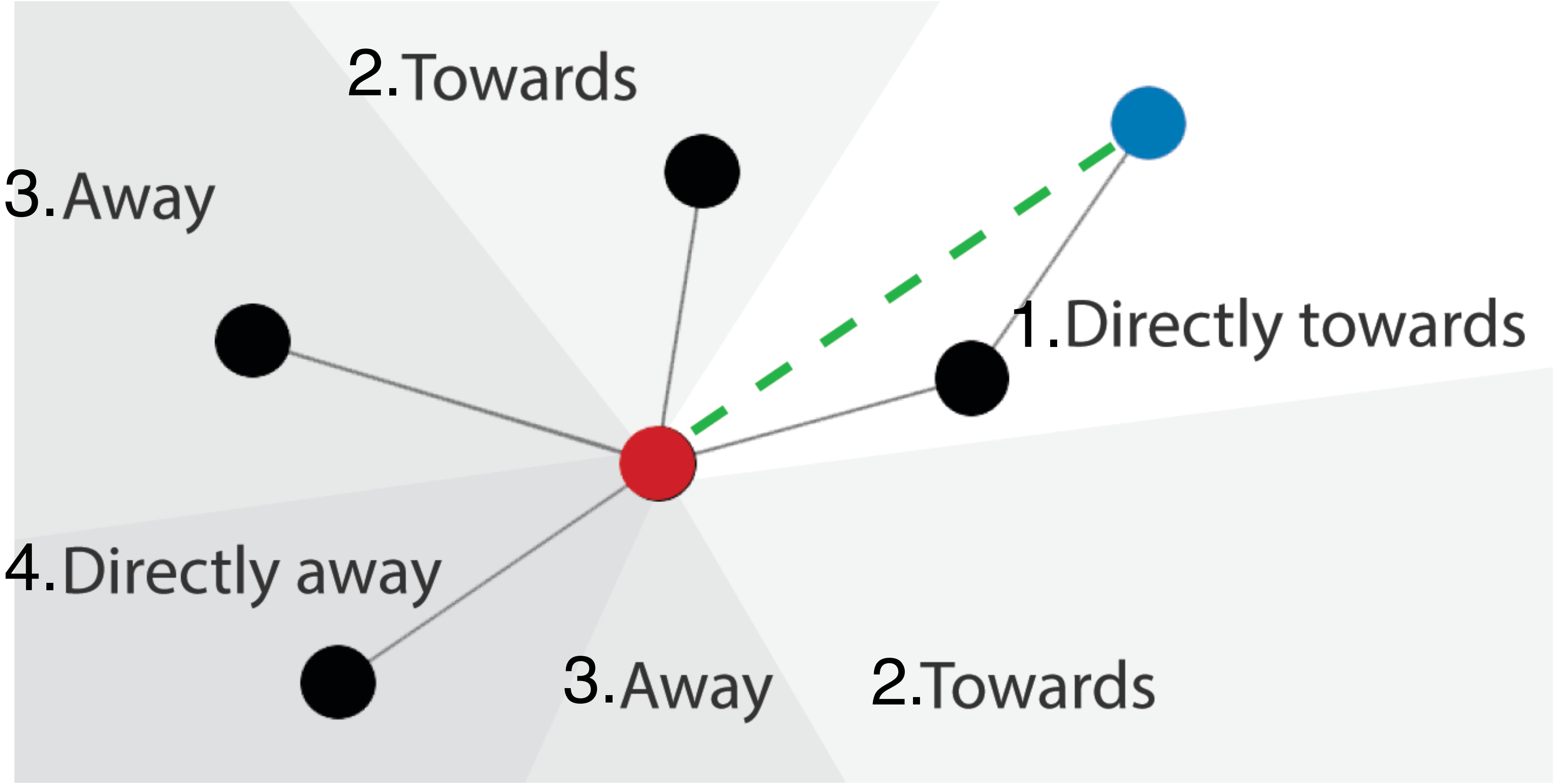
- It is possible to identify distinct path tracing behaviours (Q1)? Yes
  - investigator classified 96% of data examined with at least one code
- Many common path tracing behaviours emerged from coding (Q2)
  - use of both topological and apparent paths
  - repeated exploration of paths
  - when participants stop following paths
  - choice of nodes to search out from
  - interactions of geodesic tendency with continuity
  - **prevalence of the geodesic tendency**
  - **likely directions for the first hop in a path**

Selected behaviors: Prevalence of geodesic tendency

- participants often followed **closest to geodesic branches**
  - for all hops in a path, 40% of the time
  - for all but first or last hop, additional 26% of the time
- participants often aware of this behaviour
  - "*the [closest to geodesic] was more natural, it was harder to force myself to look away*" [P6]

Selected behaviors: Likelihood of first hop directions

- We found we could organize the direction of first hop into groups of similar likelihoods



Search Set Case Study: Predictive Model

From qualitative results to predictive model

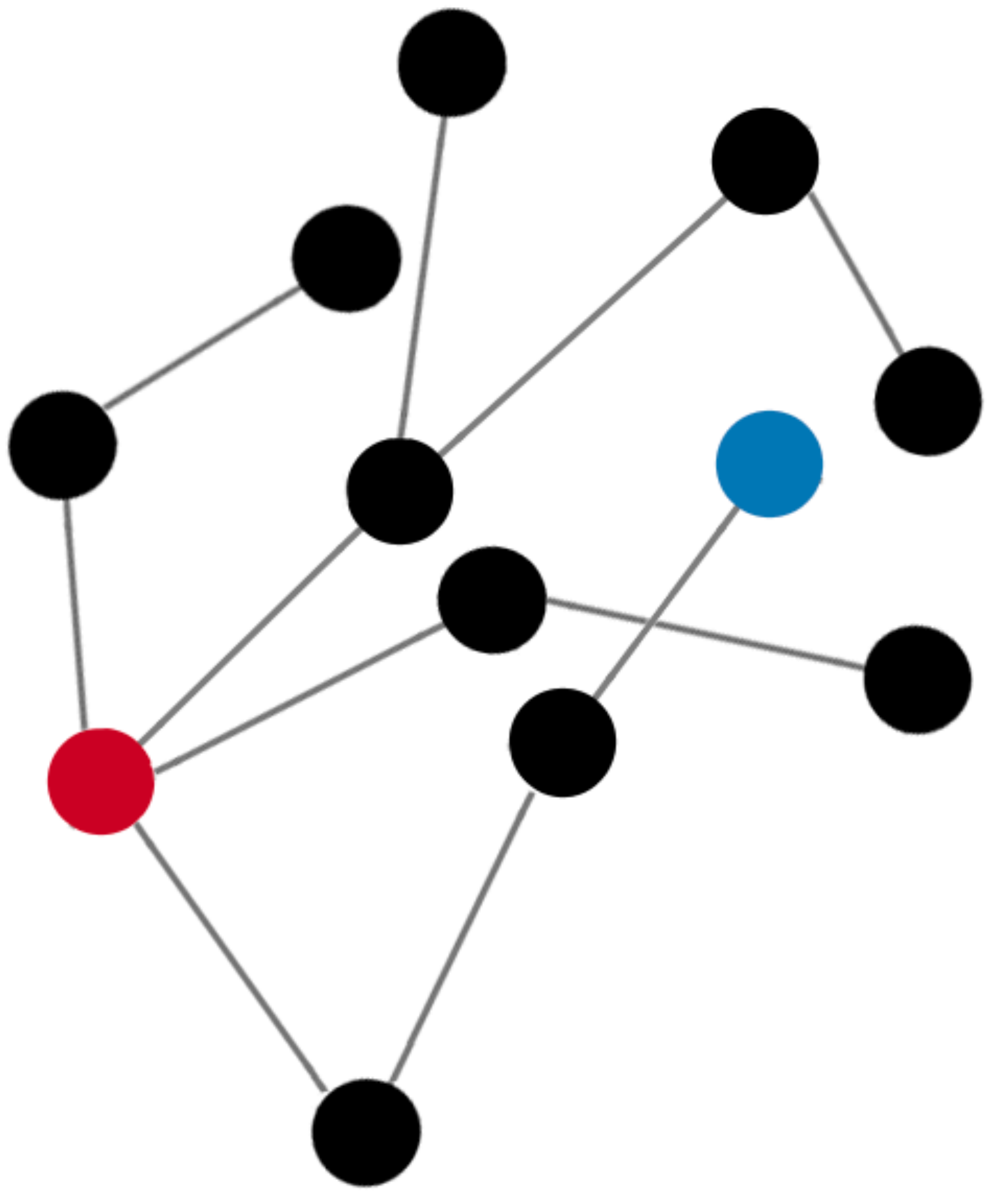
- to addresses third question:
  (Q3) can we predict a search set based on these behaviours?

- designed a 3-step, predictive model based on the characterized behaviours
  - input: a connected network with a unique solution between start/end nodes
  - output: ordered *batches* of paths that a user is likely to search
    - all paths in one batch similarly likely

3-Step search set model: Step 1

Generate batch of likely first-hop candidates

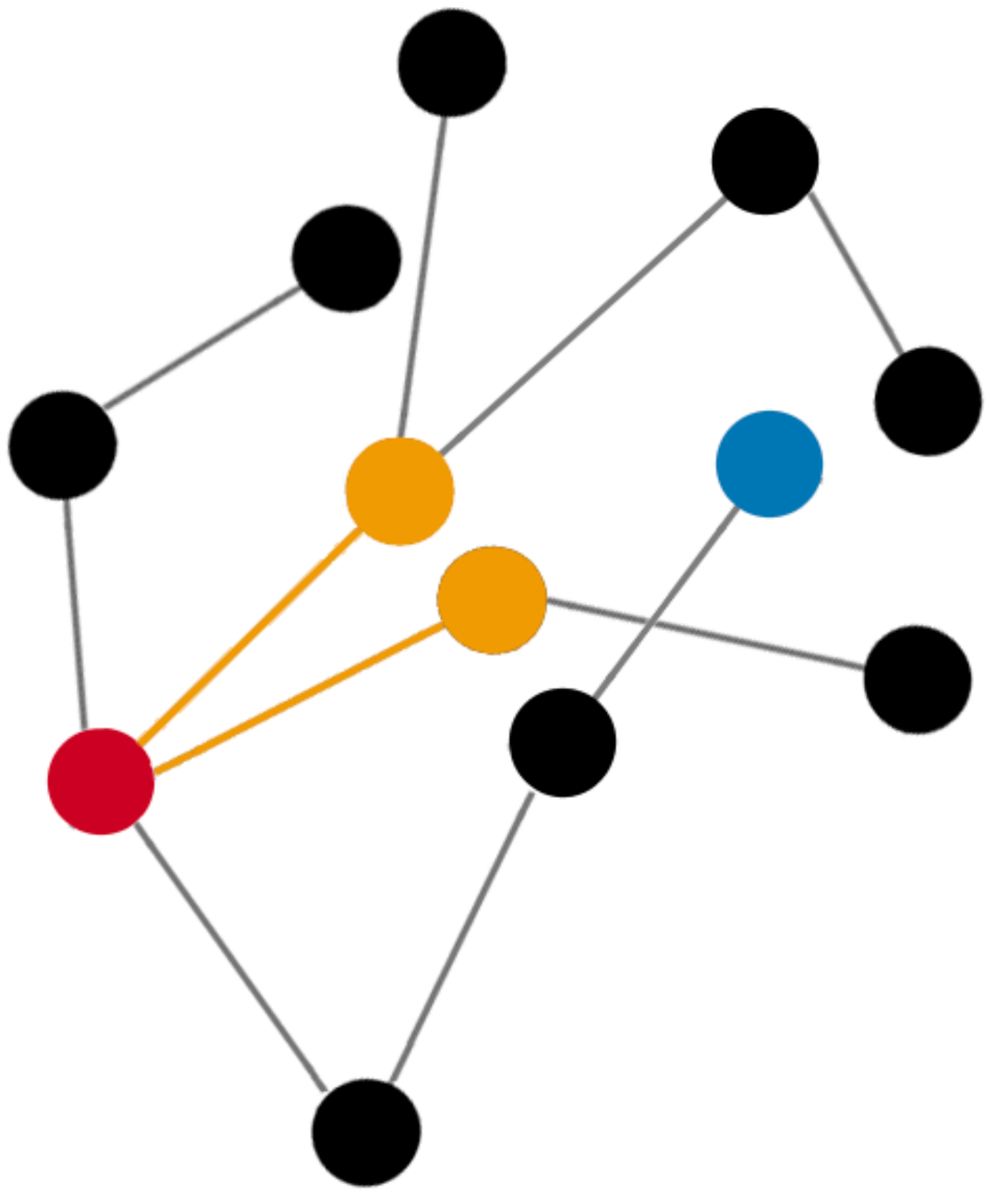
- Starting with directly towards



3-Step search set model: Step 1

Generate batch of likely first-hop candidates

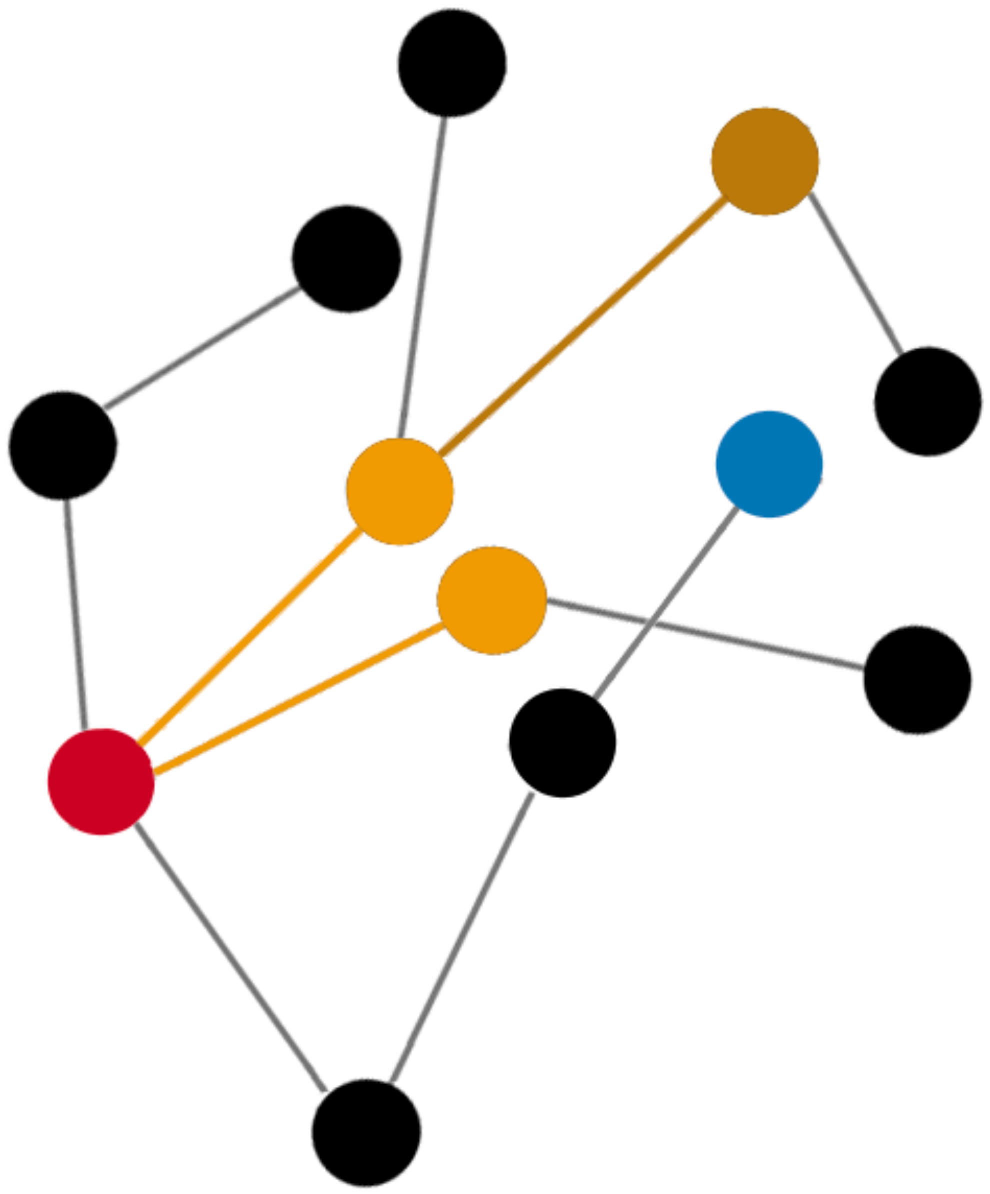
- Starting with directly towards



3-Step search set model: Step 2

From each candidate, follow geodesic shortest branches

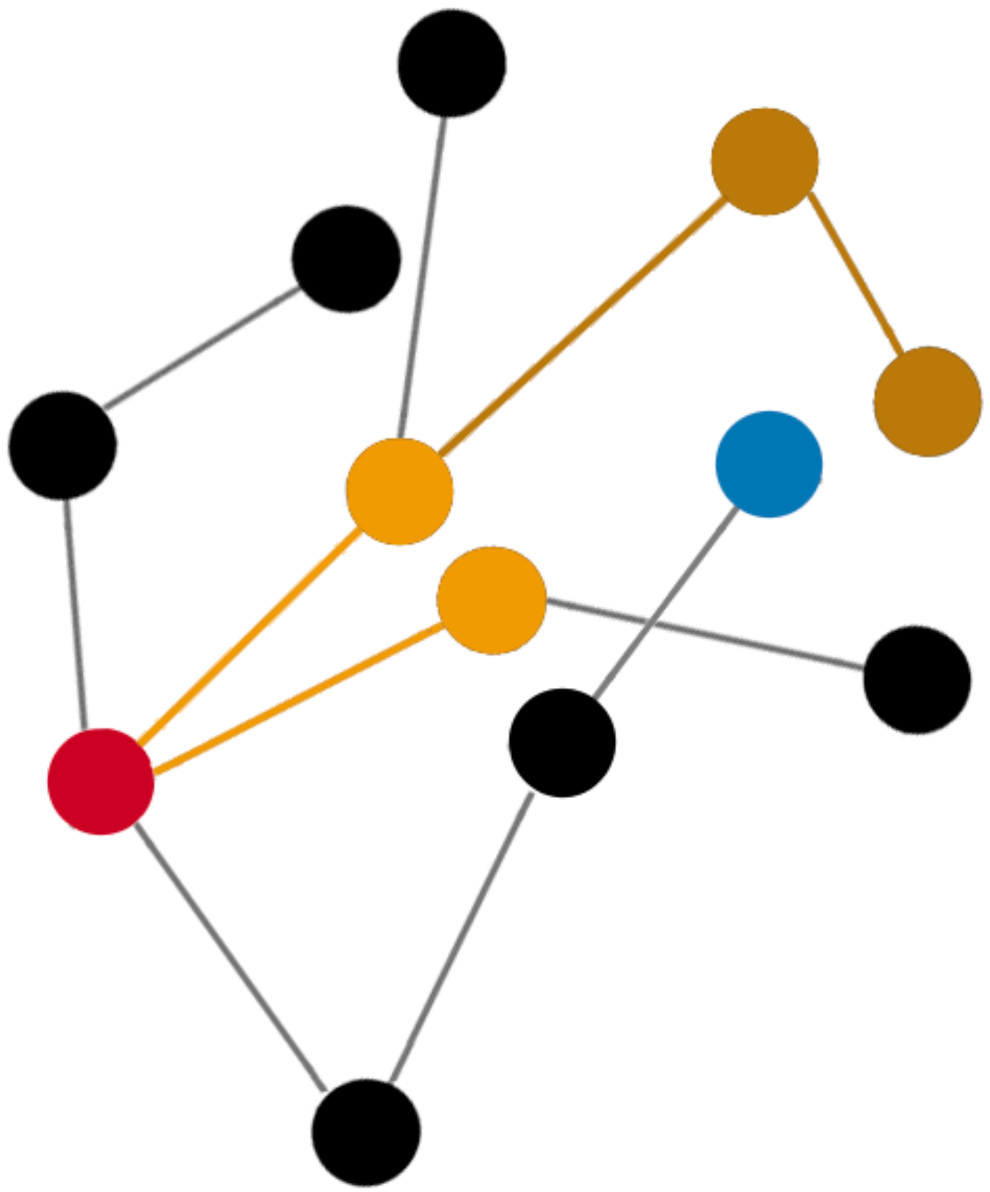
- Save path at each hop



3-Step search set model: Step 2

From each candidate, follow geodesic shortest branches

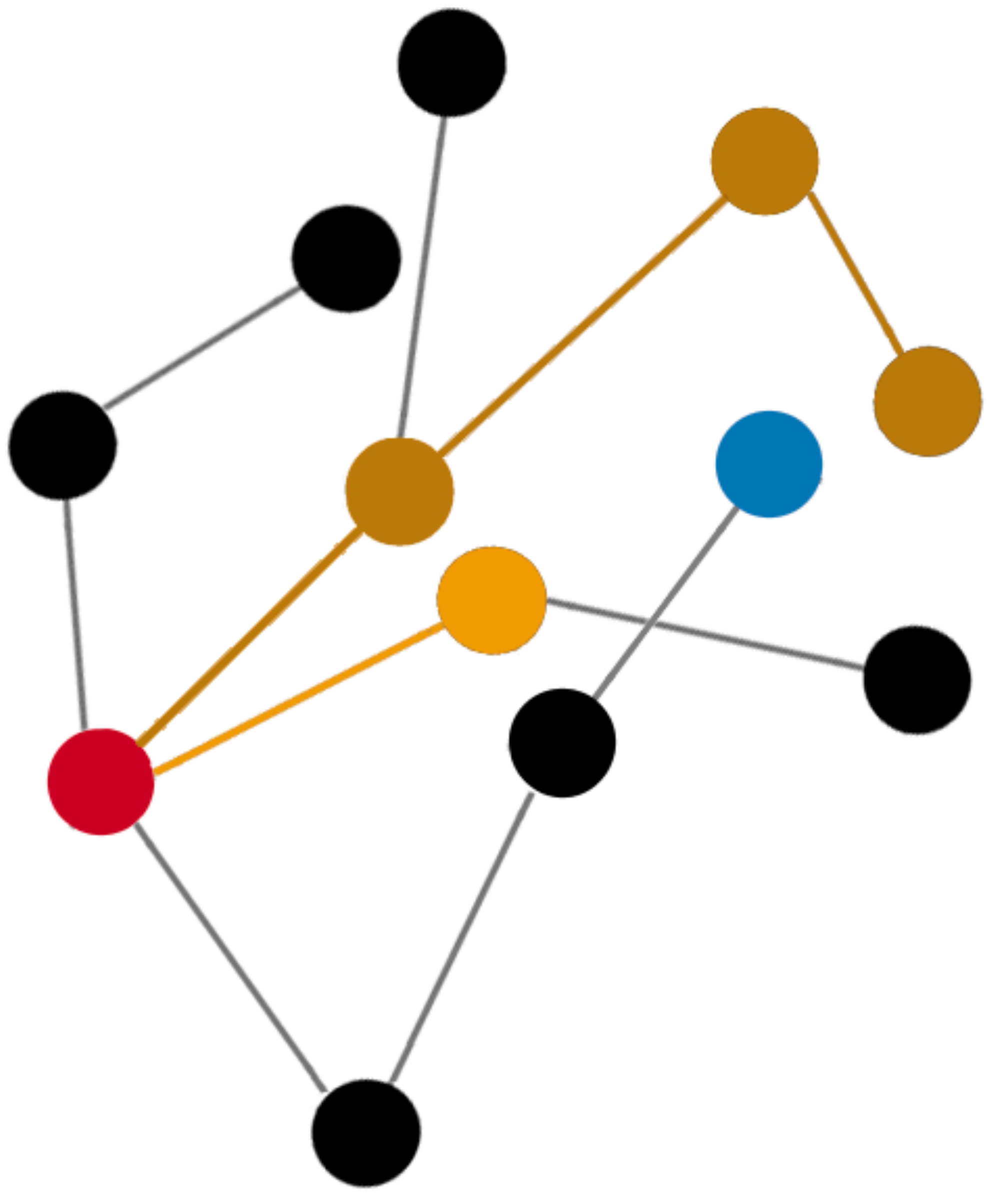
- Save path at each hop



3-Step search set model: Step 2

From each candidate, follow geodesic shortest branches

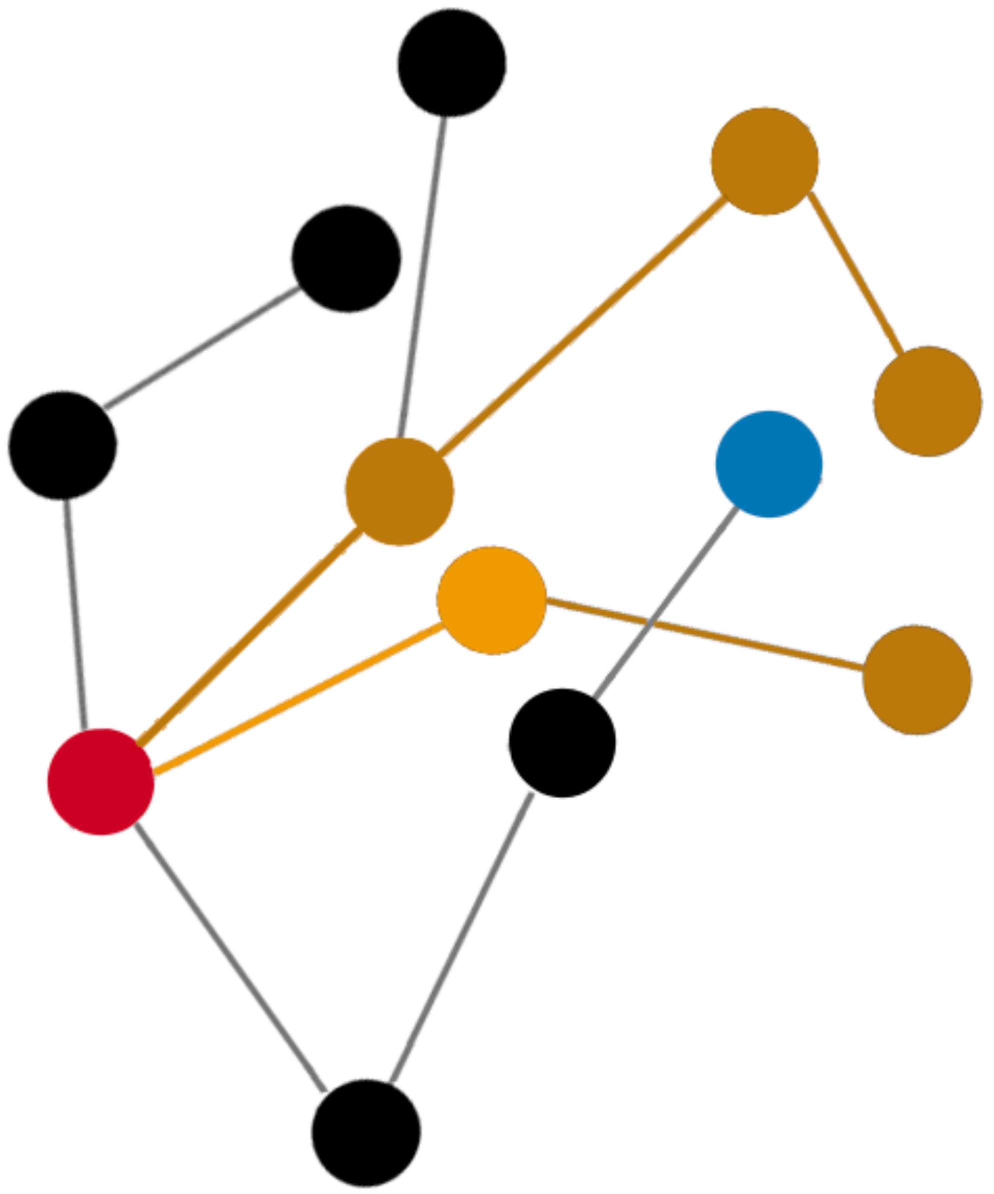
- Save path at each hop
- Go along path until stopping condition met



3-Step search set model: Step 2

From each candidate, follow geodesic shortest branches

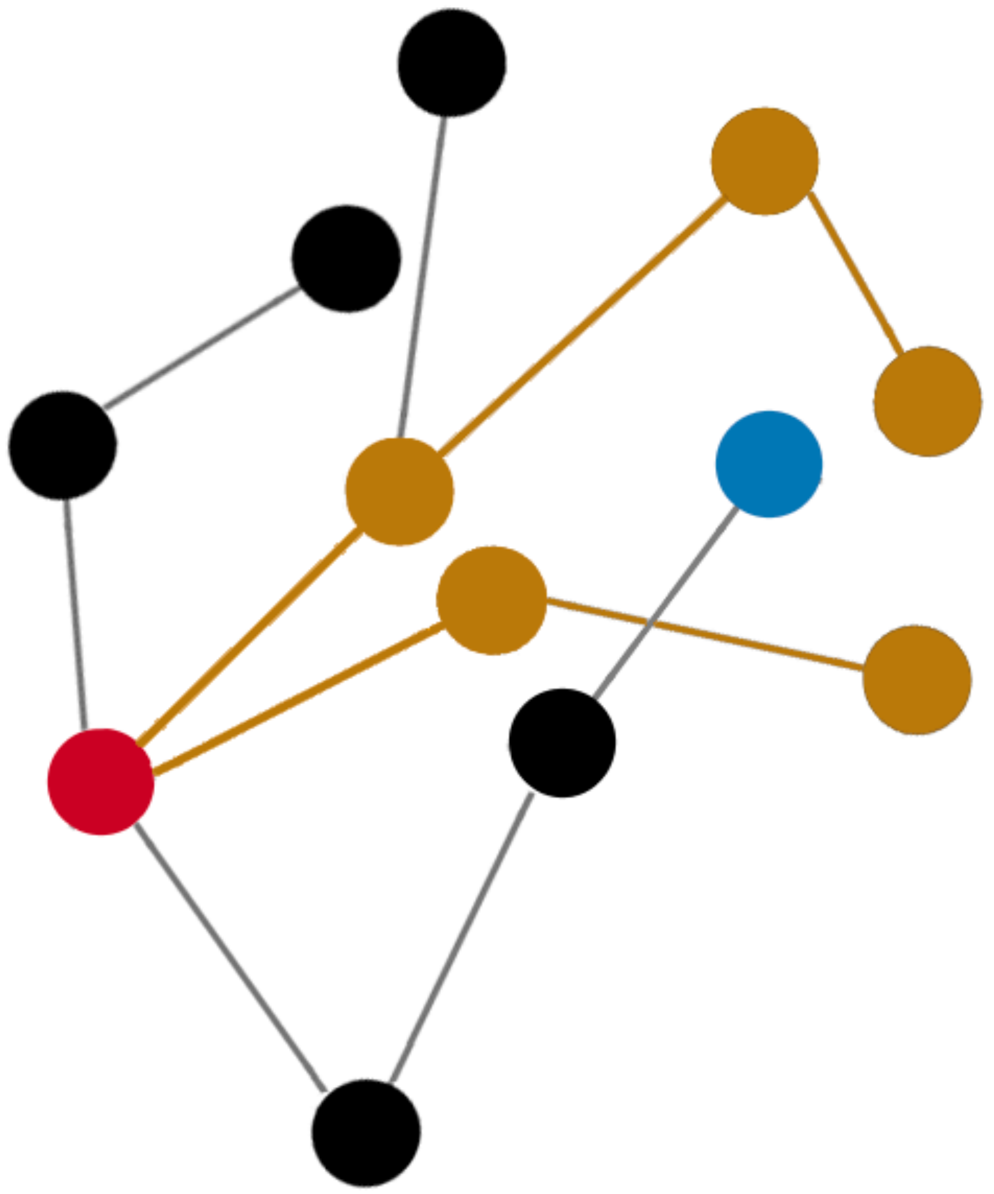
- Save path at each hop
- Go along path until stopping condition met



3-Step search set model: Step 2

From each candidate, follow geodesic shortest branches

- Save path at each hop
- Go along path until stopping condition met



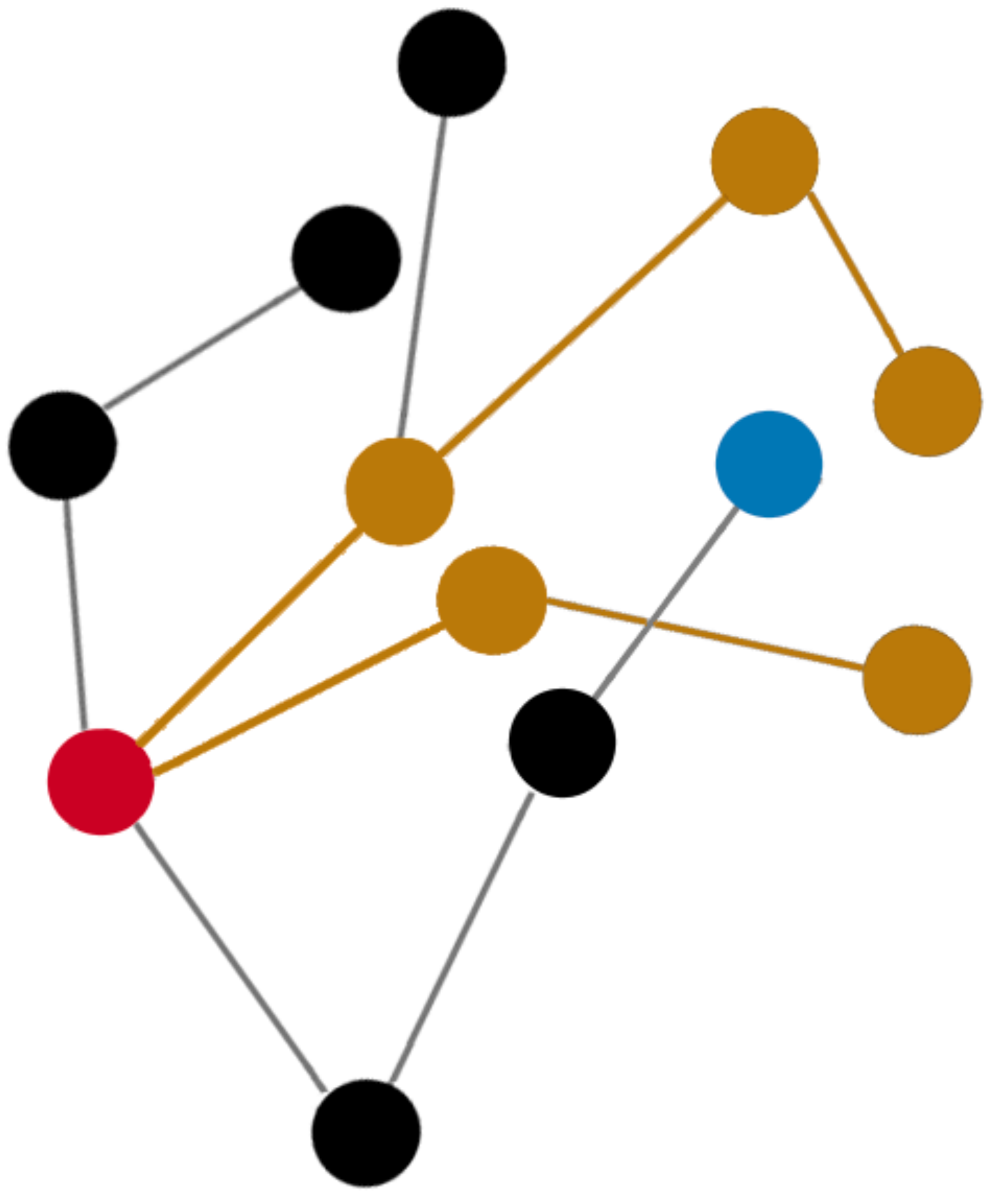
End of step 2:

- Batch of equally likely paths

3-Step search set model: Step 3

Does batch contains answer?

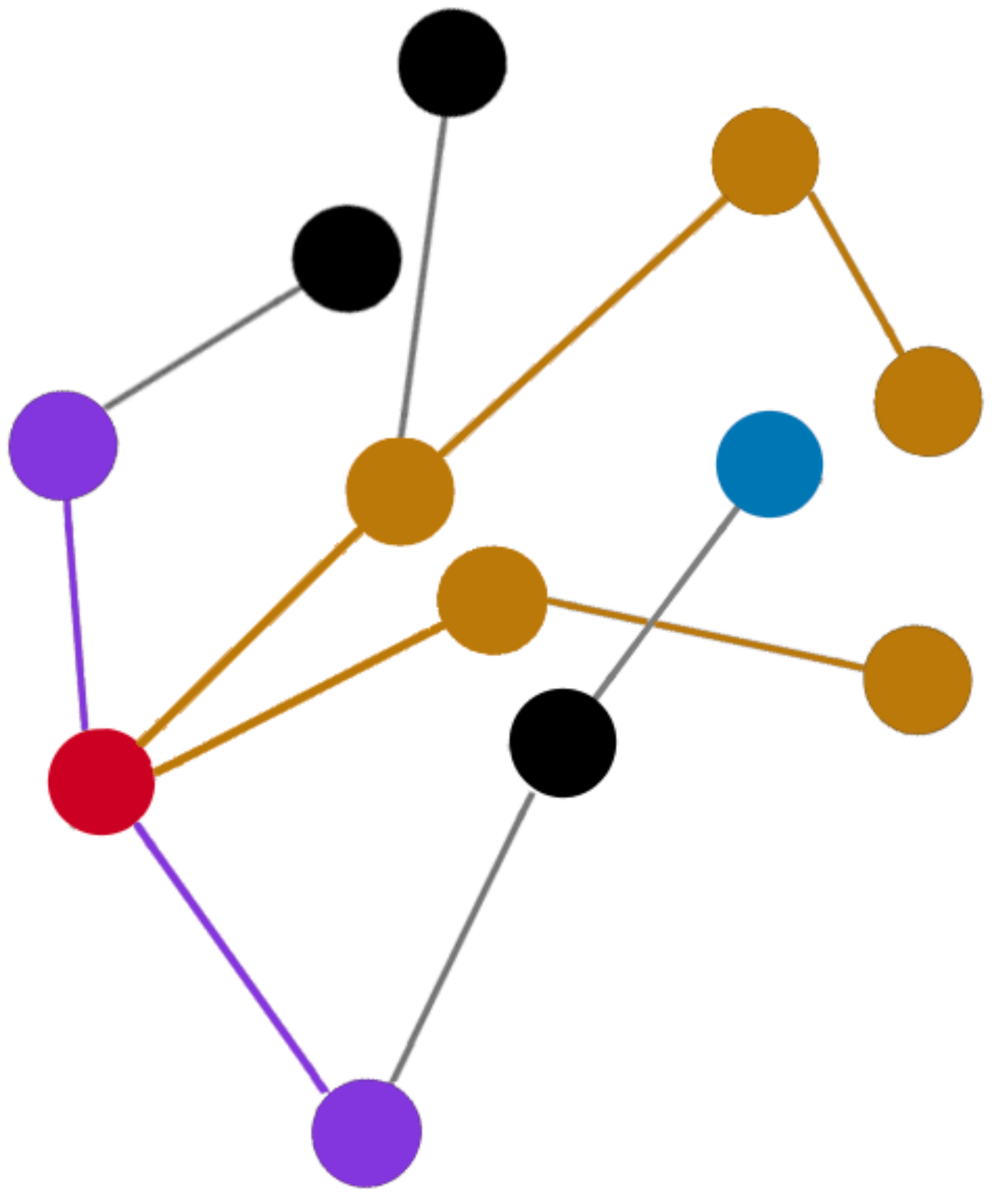
- If not: return to step 1



3-Step search set model: **Repeat** step 1

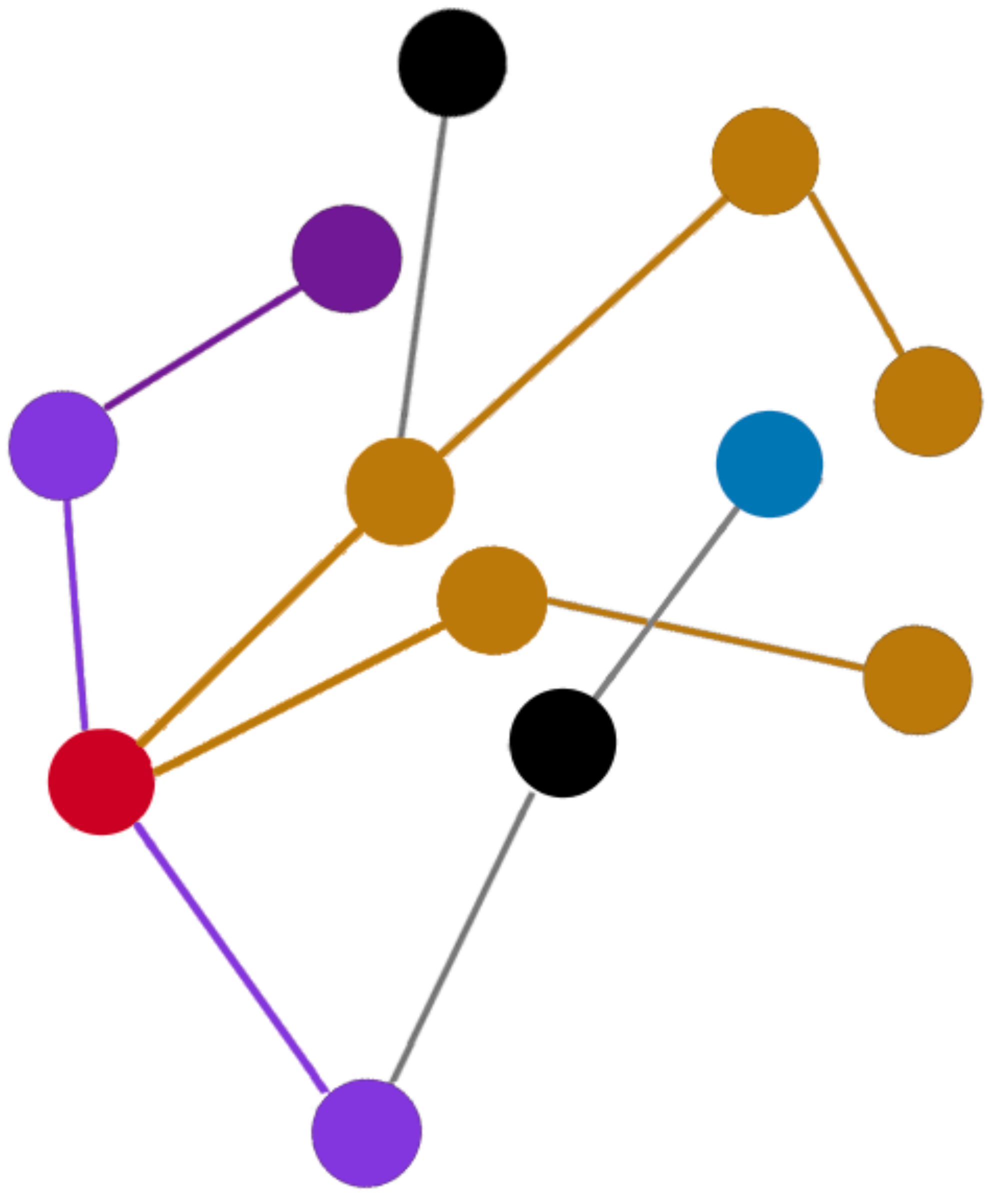
Generate batch of next most likely first-hop candidates

- Towards group



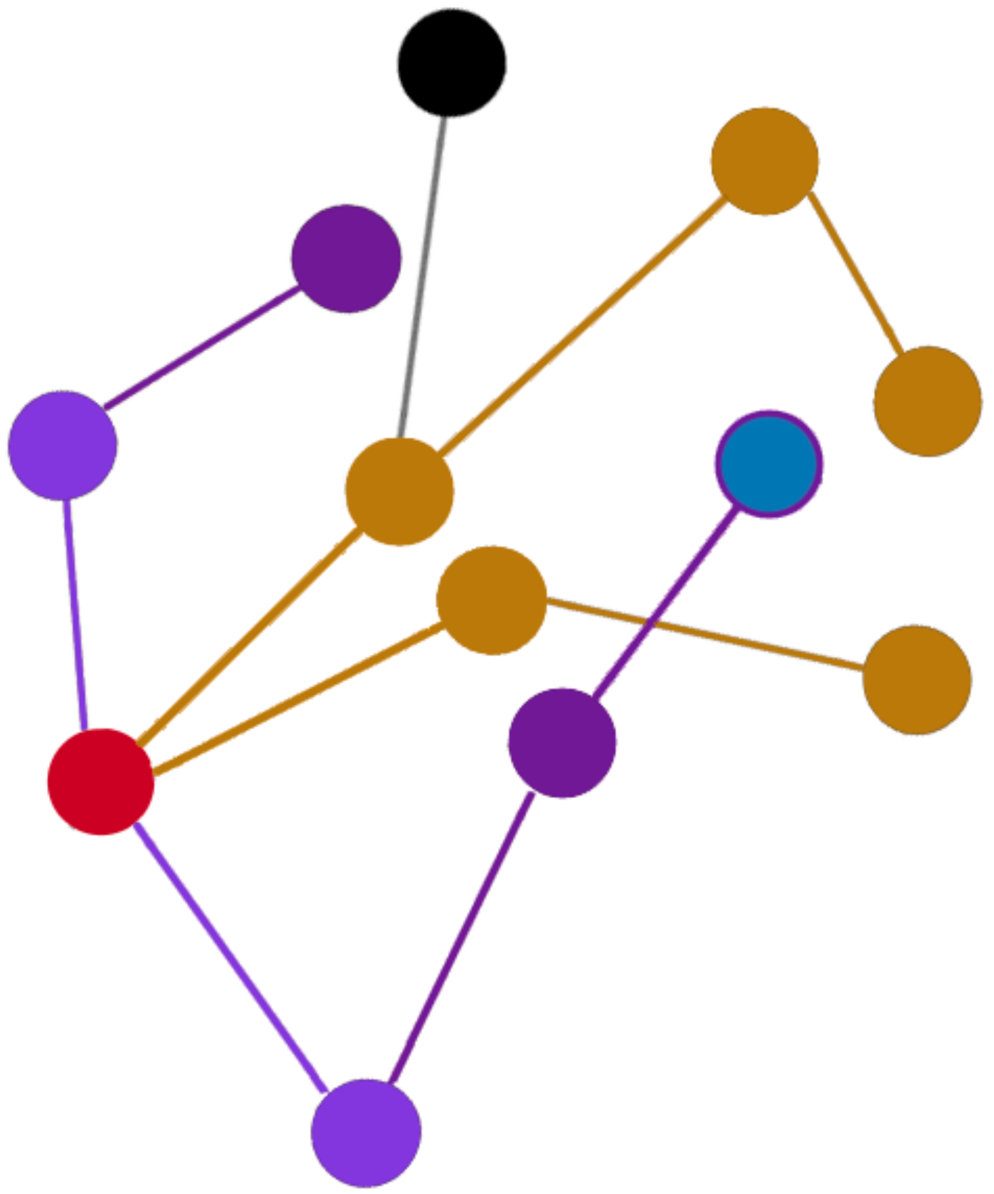
3-Step search set model: **Repeat** step 2

From each candidate, follow geodesic shortest branches



3-Step search set model: **Repeat** step 2

From each candidate, follow geodesic shortest branches

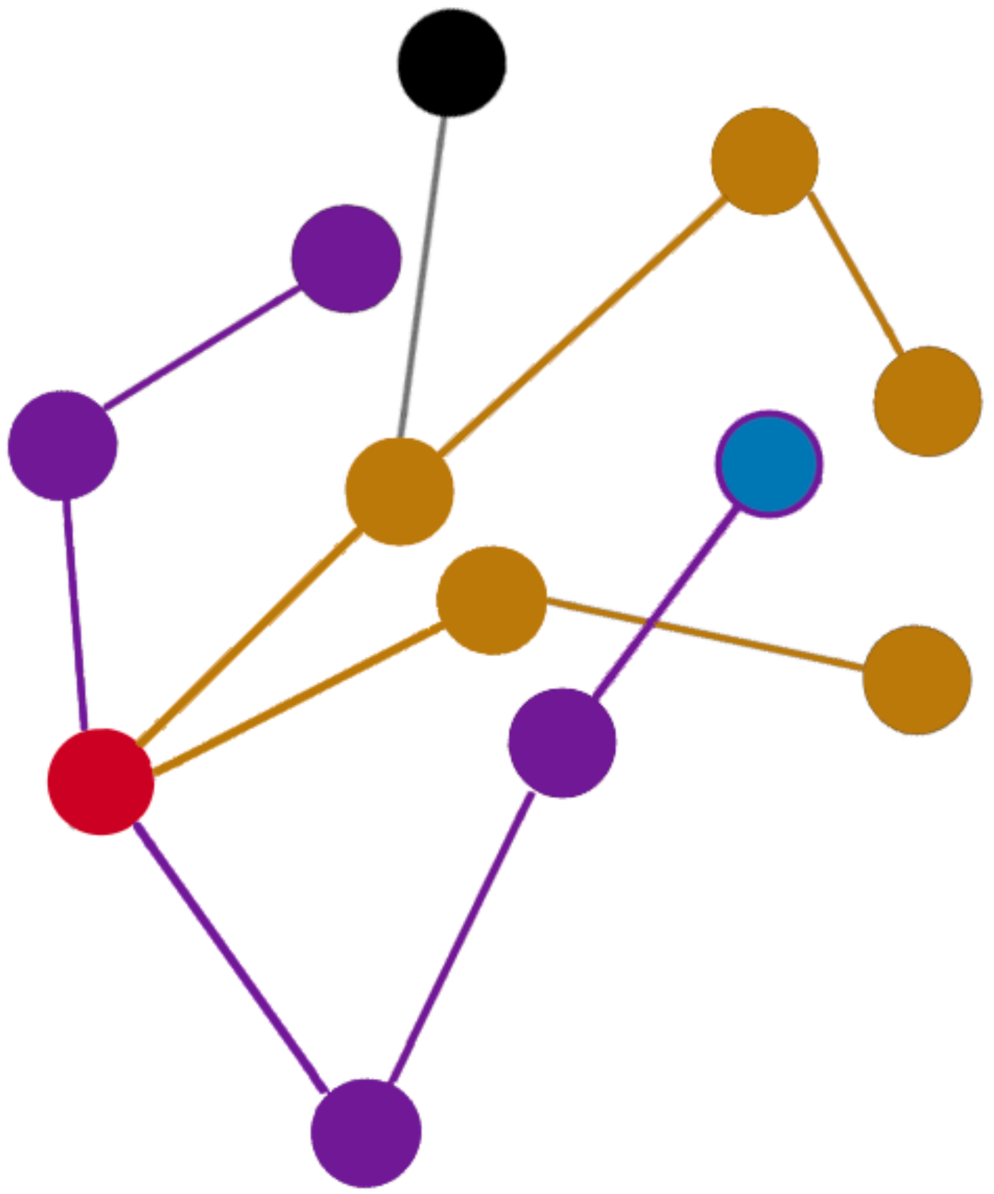


3-Step search set model: **Repeat** step 2

From each candidate, follow geodesic shortest branches

End of step 2:

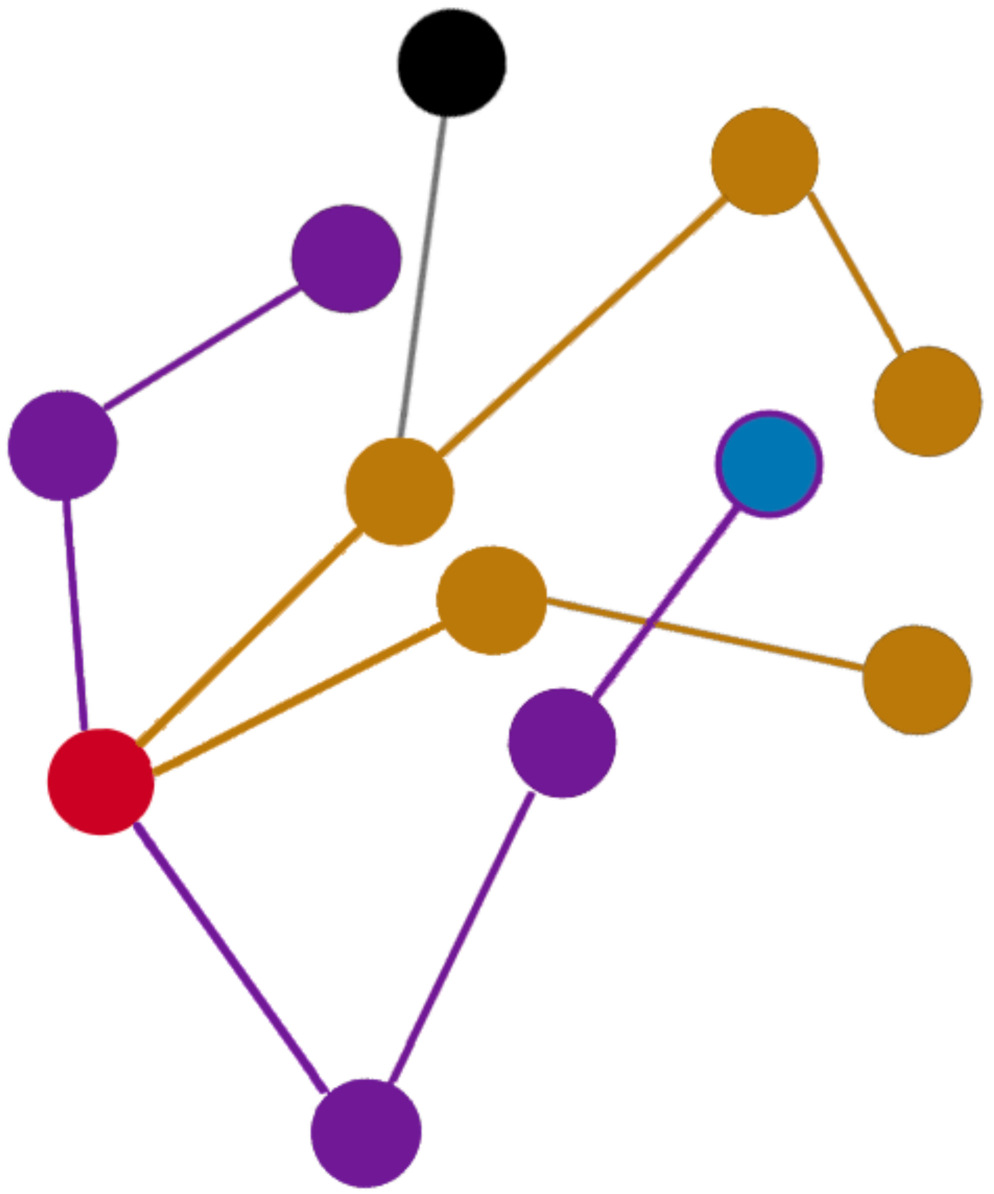
- Next batch of equally likely paths



3-Step search set model: **Repeat** step 3

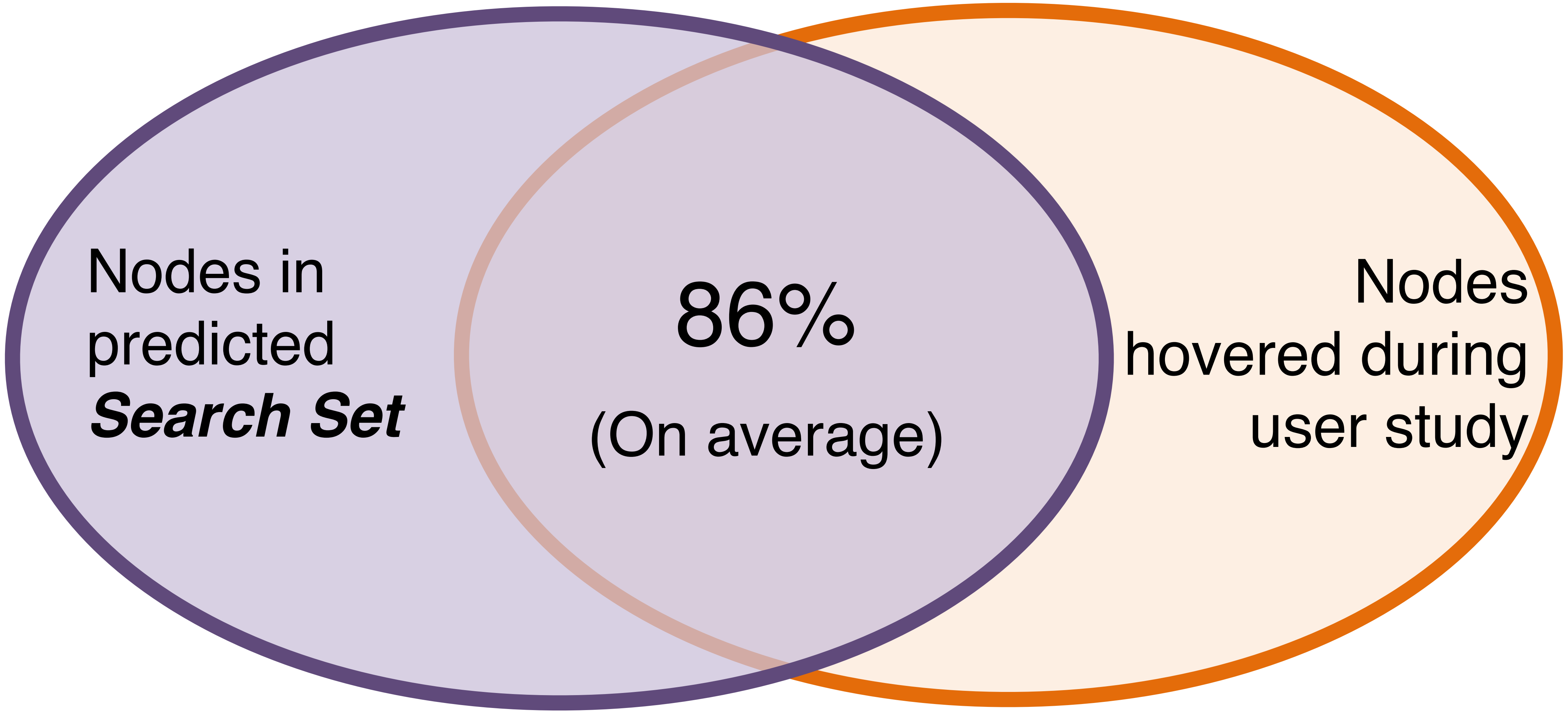
Does batch contains answer?

- Yup! So stop



Predictive model: Algorithmic implementation & results

- Implemented algorithm to run on actual graphs from study
  - Iterated on assigned parameters for angles, etc.
  - Used all (both training and test set) graphs to test model fit to data
- Results: Yes, can predict search set based on observed path tracing behaviours (Q3)



Search Set Case Study:
Multiple Regression Analysis

Further validation

- How much does this search set concept buy us?
  - (Q4) how much improvement from measuring metrics on search set?
    - one possible application of search set concept

Validation method

- vast majority of previous work uses NHST
  - null hypothesis significance testing
  - to determine a metric is important ("*edge crossings are significant, p<.05*")
- but we really want to know relative importance and overlap!
  - which metrics are correlated? proxies for the same underlying phenomenon?
  - multiple regression allows us to untangle how different metrics interact
- only two previous studies used regression
  - to compare relative importance of metrics
    [Ware et al., 2002] [Huang & Huang, 2011]
- also, only one previous study compared metrics between levels
  - edge-edge crossings at global vs. solution-path levels
    [Ware et al., 2002]

Hierarchical multiple regression experimental design

- compare metrics at three levels within graph
  - global (hypothesis: too big)
  - solution path (hypothesis: too small)
  - search set (hypothesis: just right)
- 9 metrics tested in total:
  - global:
    - node-edge & edge-edge crossings
  - search set
    - node-edge & edge-edge crossings
  - solution path
    - node-edge & edge-edge crossings
    - solution path length (# of hops)
    - solution path continuity (bendiness)
    - solution path branches (# of edges on each node)

Multiple regression experimental design

- some of these never previously studied
  - global:
    - **node-edge** & edge-edge crossings
  - search set
    - **node-edge & edge-edge crossings**
  - solution path
    - **node-edge** & edge-edge crossings
    - solution path length (# of hops)
    - solution path continuity (bendiness)
    - solution path branches (# of edges on each node)

Multiple regression details

- data sample
  - 120 graphs: the validation set, previously reserved
  - metrics measured on each graph
- dependent variables:
  - average response time
  - errors per graph (0 – 12)

Key results

- individual effects of metrics
  - replicated PW showing solution path metrics strongly correlated with response time
    - new result: same effect for error
  - search set edge-edge crossings strongly correlated with response time and error
  - global metrics not correlated with response time or error
    - contrary to some previous work
- search set edge-edge crossings had small effect over previous work:
  - response time: additional 1.8% variance
  - error: additional 4.2% variance

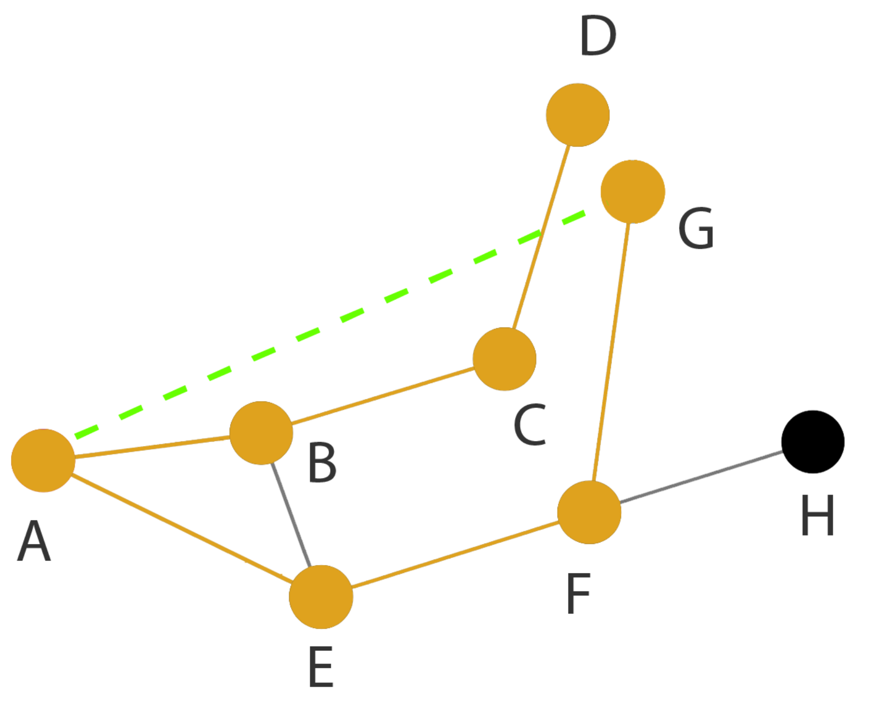
    ... on top of what all solution path metrics explained
- search set edge-edge crossings improved efficiency
  - fewer total variables needed to account for same variance

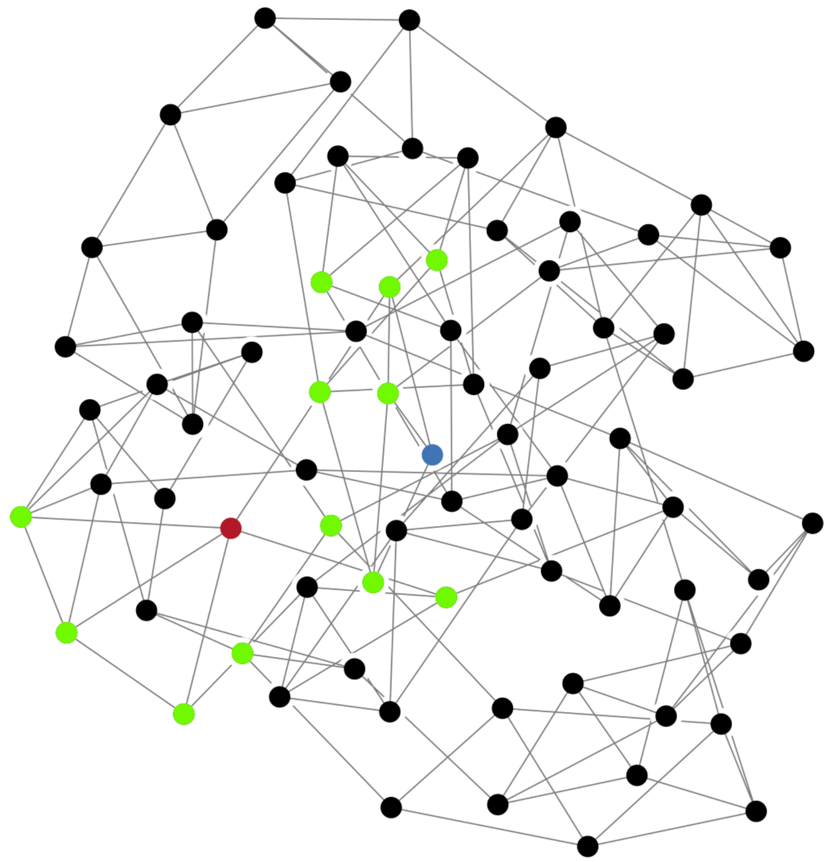
Key results

- final regression models
  - 79% of variance in response time explained by
    - solution path length
    - solution path continuity
    - search set edge-edge crossings
  - 60% of variance in error explained by
    - search set edge-edge crossings
    - solution path continuity

Discussion: Search set

- utility of search set concept
  - analysis of graph subset most relevant to the task can be very informative
  - example: might explain inconsistent findings on global edge-edge crossings
    - most previous studies used small graphs, where search set and global don't differ much
    - in large graphs, less overlap between them
- future work could explore use of search set for other applications:
  - design of new interaction techniques
  - new automatic graph layouts that make subtle changes to preserve consistency





Discussion: Methods

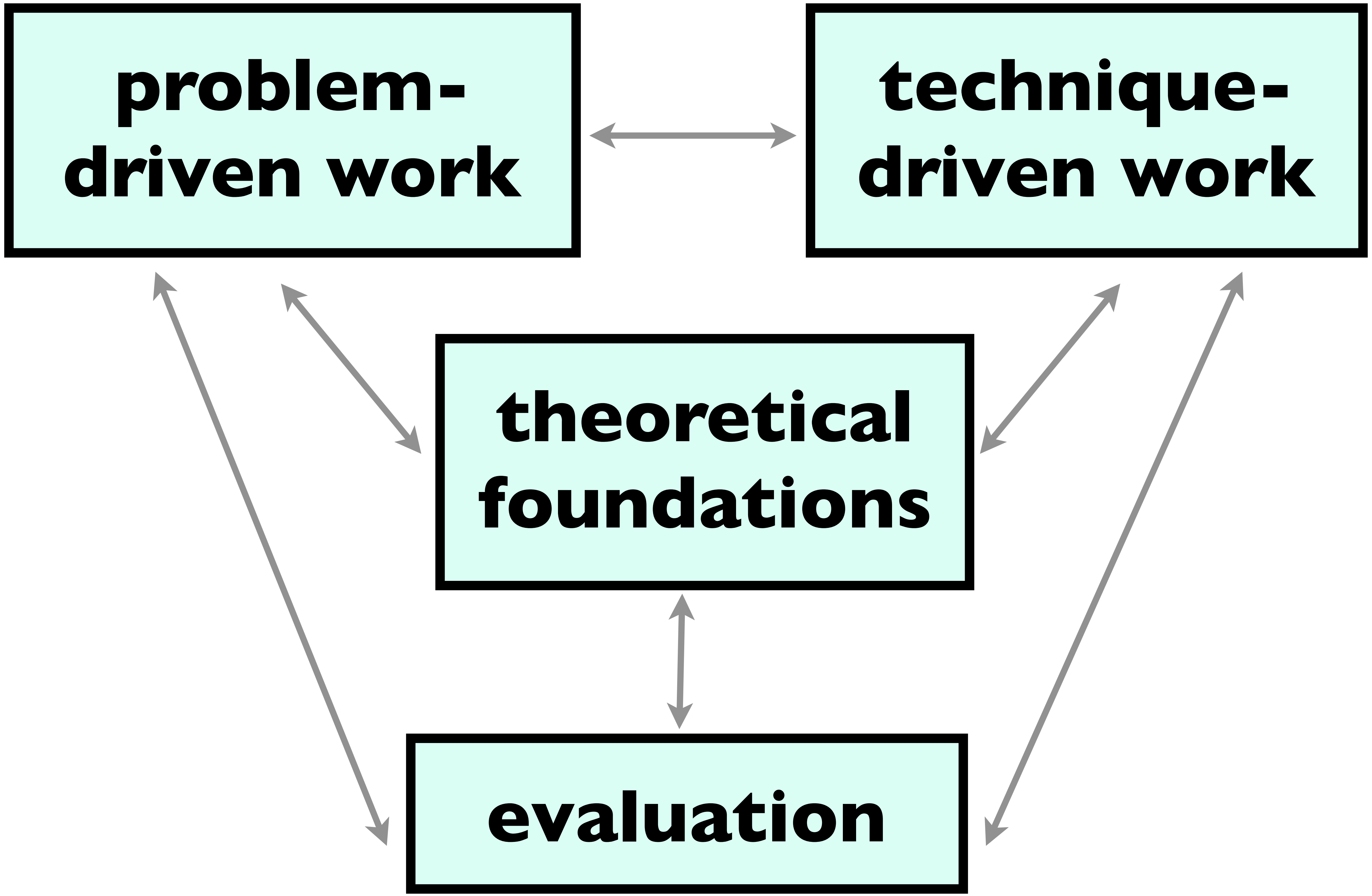
- hope to see more use of multiple regression in quantitative evaluation of visualization
  - vs current dominance of significance testing
  - esp. for quantitative metrics in contexts beyond graph drawing
- building up from qualitative analysis to quantitative metrics
  - deeply interested in both!

More on quantification

- Empirical Guidance on Scatterplot and Dimension Reduction Technique Choices. *Sedlmair, Munzner, and Tory. IEEE TVCG (Proc. InfoVis), 19(12):2634-2643, 2013.*
  - alternative to user study with few datasets and many people "data study" with many datasets and few people
    - data characteristics outweigh user differences
    - need for extensive reliable judgements
    - 2 experts quantitatively coded visual separation
      - 816 scatterplots with color-coded clusters: 5460 class judgements, ~80 hrs/coder

- Increasing the Utility of Quantitative Empirical Studies for Meta-analysis. *Lam and Munzner. Proc. BELIV 2008.*
  - how we could improve our reporting of quantitative studies

Research agenda: Angles of attack



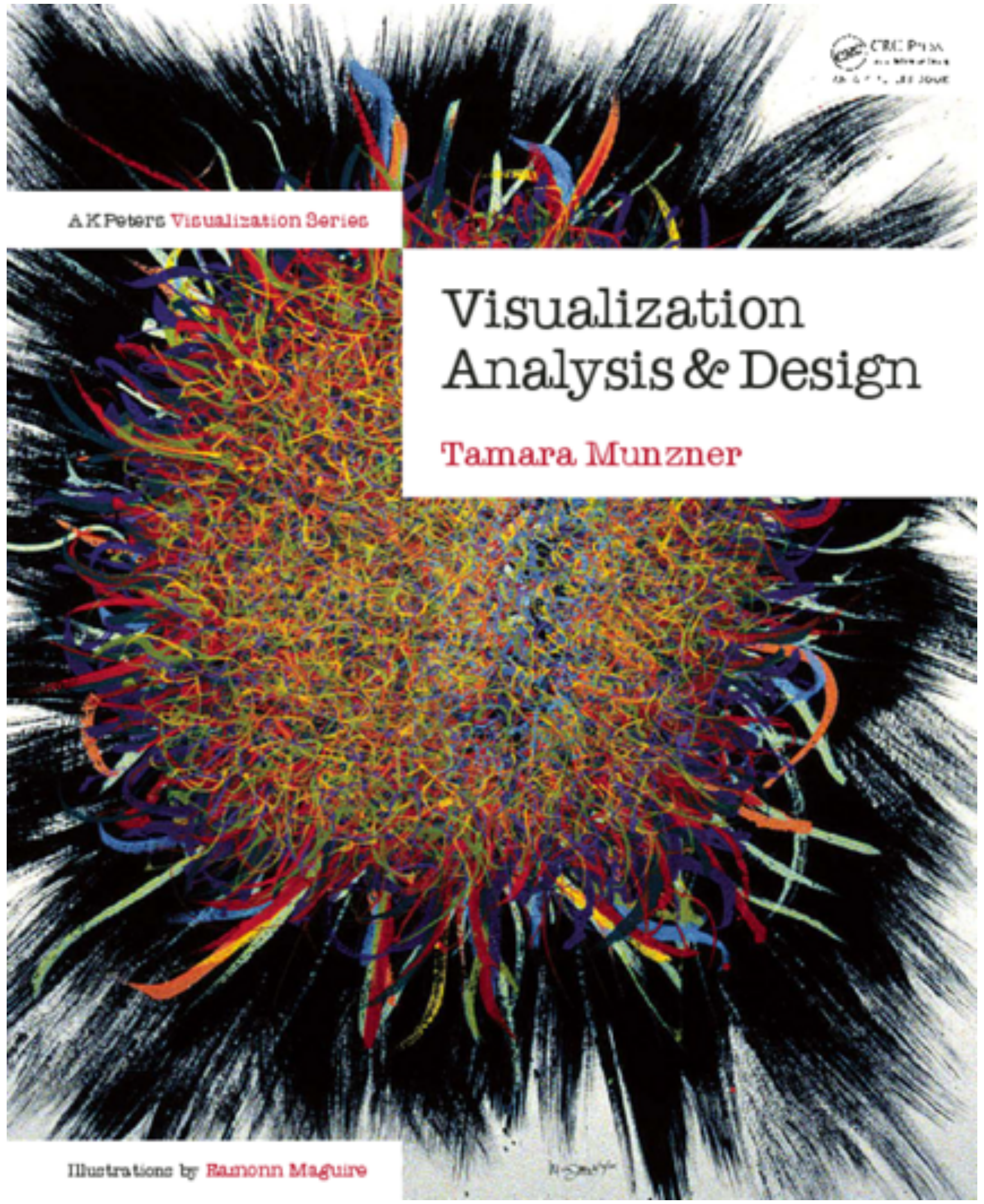
More information

@tamaramunzner

- theoretical foundations: book
  (+ tutorial/course lecture slides)
  http://www.cs.ubc.ca/~tmm/vadbook
  - 20% promo code for book+ebook combo: HVN17
  - http://www.crcpress.com/product/isbn/9781466508910

- this talk
  http://www.cs.ubc.ca/~tmm/talks.html#stuttgart18
- funding: AT&T Research, NSERC

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



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