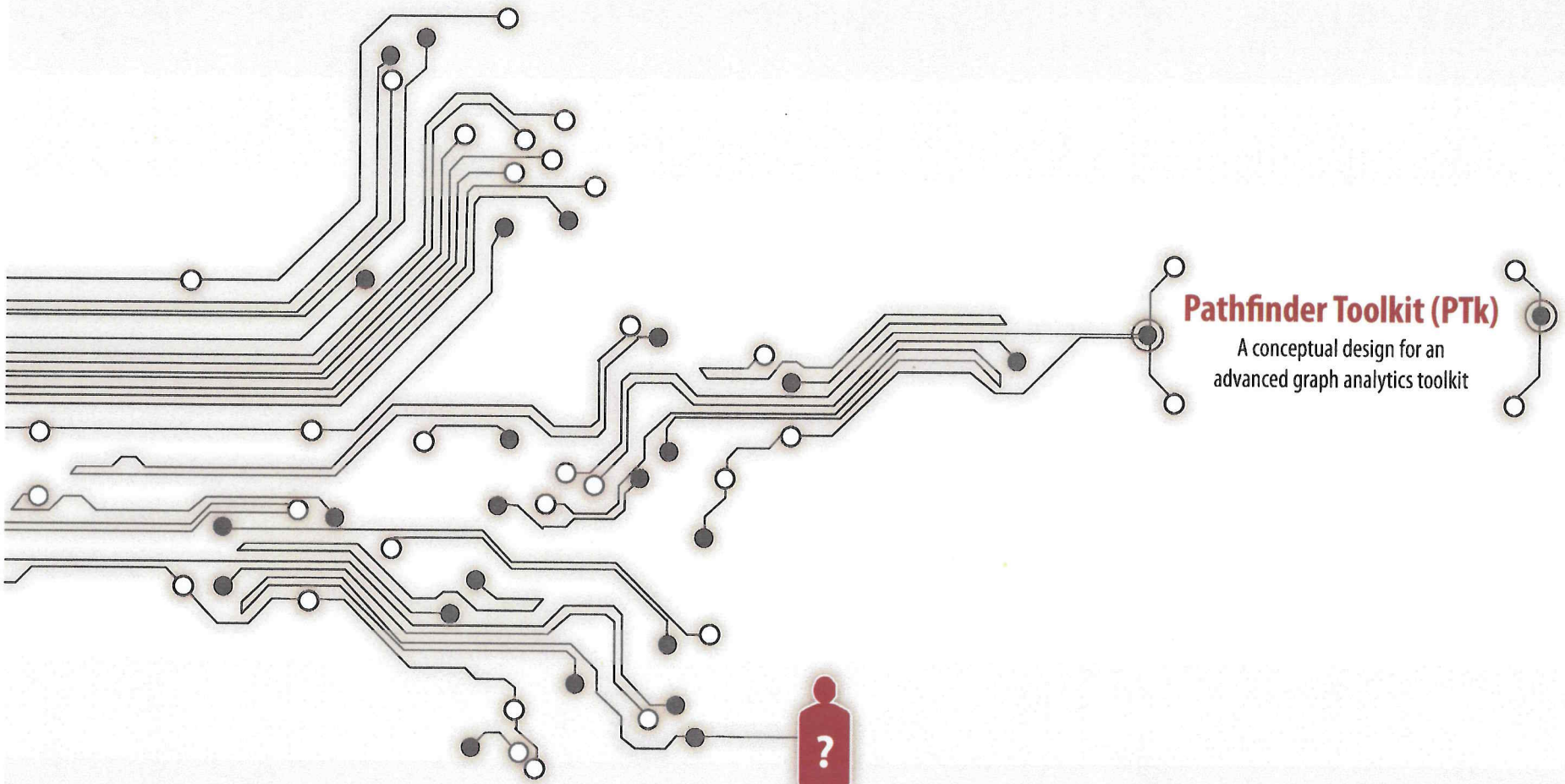


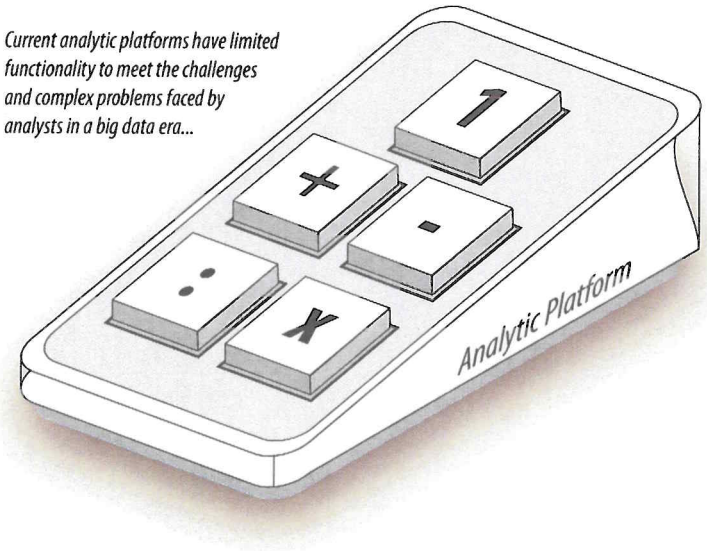
Pathfinder Toolkit (PTk)

A conceptual design for an
advanced graph analytics toolkit

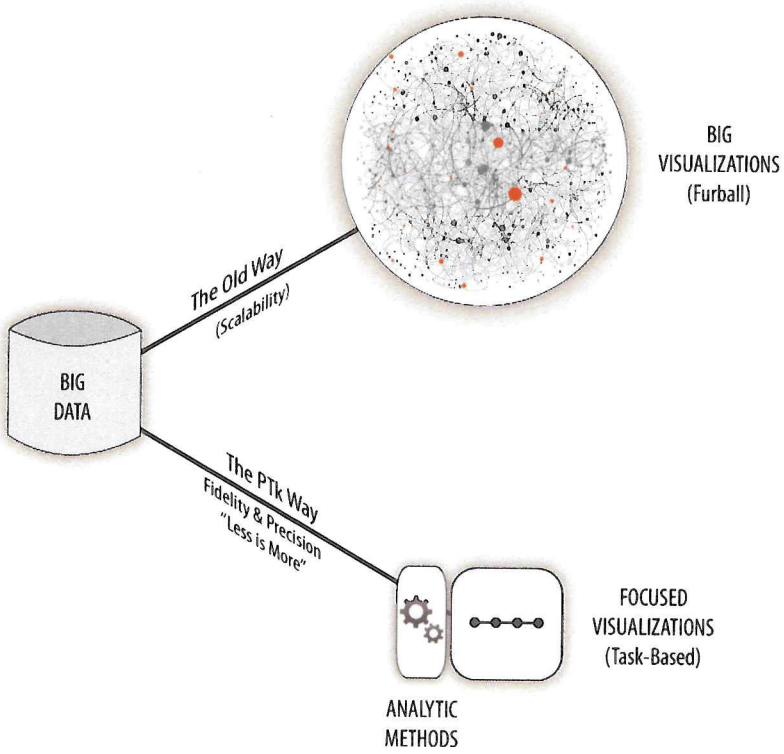
Identity Discovery
Challenge

Challenge to Innovate

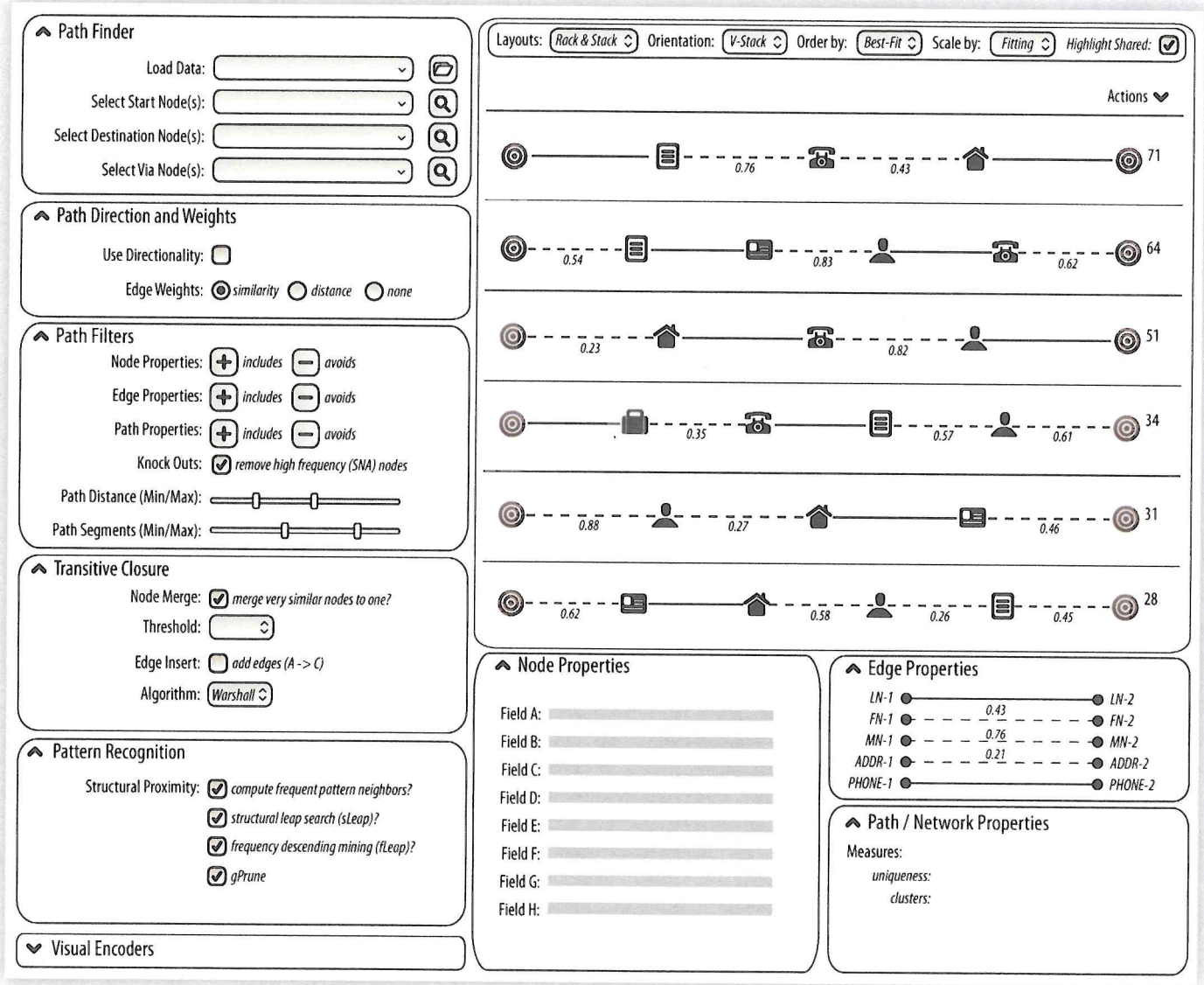
Current analytic platforms have limited functionality to meet the challenges and complex problems faced by analysts in a big data era...



Consider this...



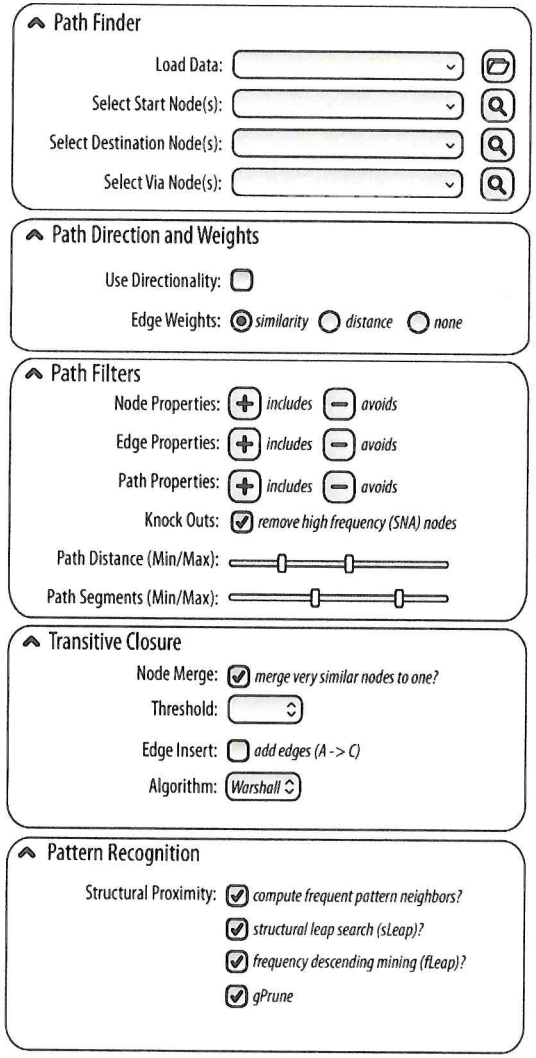
Pathfinder Toolkit: Analytic Methods + Task-Based Visualization



Analytic Methods

What am I looking for?

How can I configure the finder for my specific scenario?



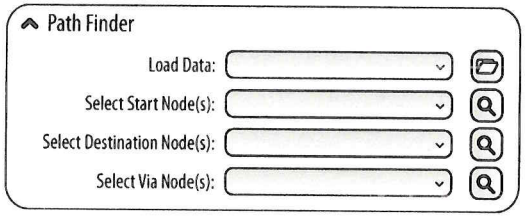
How should path finder interpret the relations as represented in the data?

How can I reduce the noise and enhance the signal with filters and interference detection?

Is there a way to simplify a complex path by merging nodes or inserting edges?

Are there patterns which help to discover meaningful structural or topological characteristics to support the sense-making process?

Show me the paths...

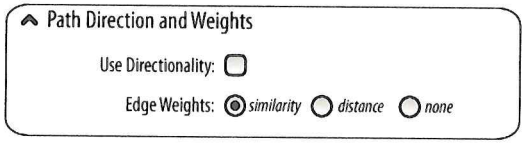


We see this module as similar to a GPS or Directions mapping interface.

Start Node: SEED-1
Destination Node: SEED-2

Via Nodes would be useful for more complex scenarios where intermediate points are known.

We know there are algorithms for shortest path analysis, but how can they be extended to this analytic use case?



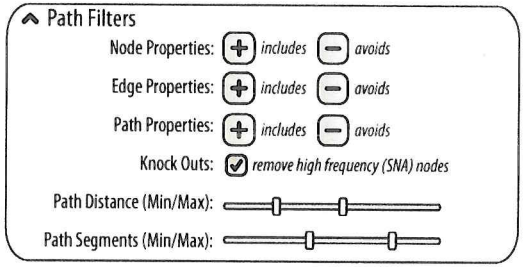
The challenge problem is comprised of edges derived from an entity resolution process.

They are not directional, however they are weighted.

A score of 0 to 1 indicates "similarity" of records, e.g. 1=Exact Match

These controls would be useful to parameterize the algorithm's treatment of weights

Show me only...



Path filters are intended to fine-tune
path features (e.g. avoid highways) and
reduce noise by eliminating possibilities.

Includes and avoids by property reduce graph complexity.

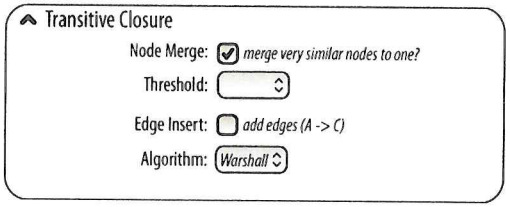
Knock Outs is a conceptual concept to represent
the possible use of SNA Centrality measures to detect
nodes which negatively (and erronerously) affect
path structure.

For example, entity resolution often highlights
the high frequency of common entities whose
linkages are not relevant to the analytic objective.

Path Distance (or Similarity) can be used to filter the strength of the path
(e.g. show only highest scores)

Path Segments can be used to filter the complexity of the path.
(e.g. show only paths of 5 or more segments)

Simplify my paths...



Transitive Closure offers possibilities to
simplify a complex graph structure by
merging nodes (based on criteria)
and inserting edges (shortening paths).

Node merge is based on a decision critieria such as:

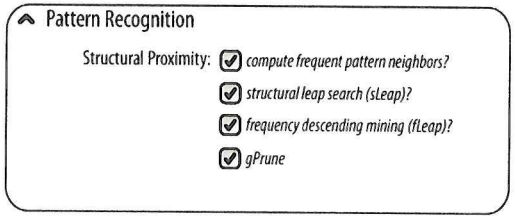
if name A is similar to name B by score of 0.8 or above
AND
address A is similar to address B by a score of 0.71 or above,
then MERGE

Edge inserts may also be employed based
on a logical criteria or use of algorithms
such as Warshall and PageRank (are there others?).

These possible methods are analogous to
cartographic use of simplification methods
such as the Douglas-Peucker algorithm which
"splits and merges" complex vector paths.

However, these methods are very sensitive and
could affect the analytic story (verbosity).

Show me interesting patterns...



We know very little about the possibilities
of graph-based pattern recognition.

However,
research suggests some advancements in
structural / topological pattern discovery approaches.

Are they applicable?

Do methods exist to cluster, classify, associate,
or otherwise detect "interesting" features of a graph structure?

We don't know the answers to these questions. Do you?

Task-Based Visualization: Path Mapper (LoD 0: Overview First)

Based on a path finder configuration, how are the parameters effecting the results — what is the big picture?

Level of Detail (LOD) 0
should provide a bird's eye view, to include:

Path Depth (Segments)
Path Breadth (Candidates)
Edge Sensitivity (Weak Links)
Edge Strength (Strong links)
Node Centrality (Dominance)
Connectedness (Shared nodes/edges)
Distinctness (Unique paths)
Path Distance

Any new ideas for an effective visualization approach?

I'm programmed to be a treasure hunter,
not a caveman

— KITT, Knightrider