Relevance Estimation for Evidence Marshalling

23 May 2006

Robert E. Wray
Soar Technology, Inc.

Ron Chong
Cognitive Composites, LLC

Outline

- Goal:
  - Support evidence marshalling (EM) with automated reasoning
- Problem:
  - Infinite search spaces
  - Difficult to know if/when progress is being made
    - One step from "Eureka!"
- Proposed Solution: Relevance Estimation
  - Algorithms that guide EM searches and get better with experience
- Methods:
  - Test bed for evaluating relevance estimation and evidence marshalling

- Warning:
  - Not a lot of Soar in this story; mostly laying groundwork…

Soar Technology
Thinking *inside* the box.

Evidence Marshalling

- Process of “connecting the dots” to form a coherent narrative about a collection of evidences
  - Pop-culture Examples:
    - Sherlock Holmes
    - Procedural TV shows (Perry Mason, Law & Order)
  - Specialized human techniques for EM
    - Wigmorean analysis
- Many potential applications for automated methods:
  - Police investigations
  - Insurance fraud
  - Intelligence analysis

Soar Technology
Thinking *inside* the box.

Automating EM

- Automated reasoning techniques have been demonstrated as effective tools for EM in very restricted problem domains
  - cause-of-death conclusions from evidence
  - insurance fraud detection
- Limits to automated EM in open-ended domains
  - Hypothesis generation
  - Knowledge sources
  - Efficient control of search process

Automating EM: Near-term solutions

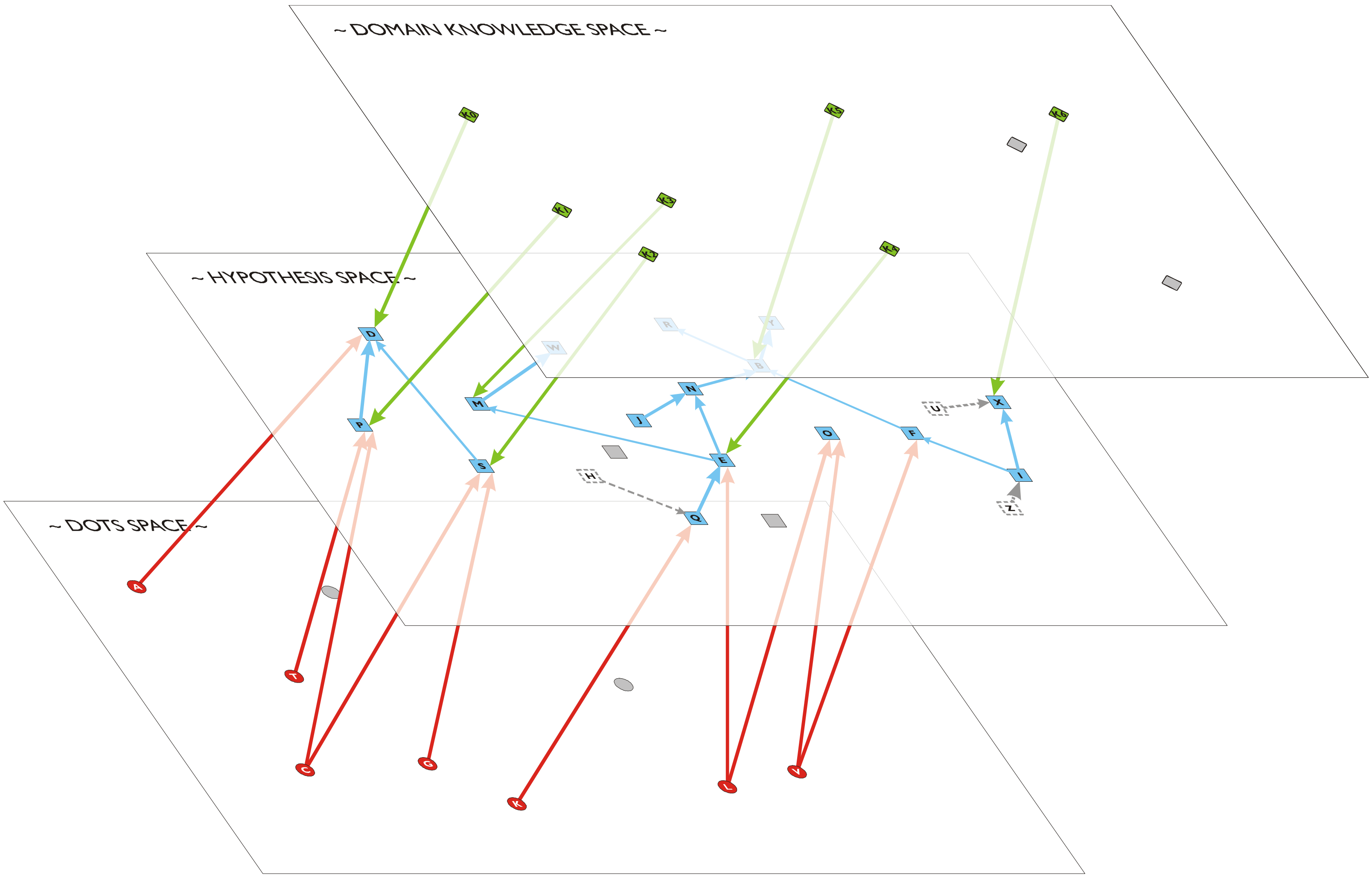
- Hypothesis generation:
  - In the near- to mid-term, no automated approach is likely to challenge human creativity
  - Our approach leverages user insight and experience by capturing their notions and "hunches" and then searching for confirming or disconfirming evidence
- Sources of knowledge
  - Trifles
    - Many existing automated sources: data extraction tools
  - Domain knowledge
    - Facts and inferences from applicable domain(s)
    - Increasing stores of general knowledge resources: Cyc, Semantic Web
- Efficient search
  - Must find intermediate hypotheses/assertions that can connect trifles to user hypotheses

Soar Technology
Thinking *inside* the box.

Challenge: Search Space is Infinitely Large

- Domain knowledge and trifles combine in infinitely many possible ways
  - Bill Clinton was POTUS in 1996
  - Hillary Clinton is Bill Clinton's spouse
  - …
- Difficult to judge if any individual intermediate hypothesis is getting closer to the goal
  - Hillary Clinton lived in the White House in 1996
- Multiple, simultaneous searches
  - Trifles: Which "trifles" are relevant
  - Domain knowledge: Which facts about the world are relevant?
  - Hypotheses: Which intermediate hypotheses are on the right track?
  - Confidence: Which assertions are more likely/probable?

Simultaneous searches



Hypothesis:
Efficient search for EM is a <u>solvable</u> challenge

- Evidence:
  - Although the search space is infinitely large, people perform evidence marshalling adequately
  - How?
    - Training & experience, tools (Wigmorean analysis)
    - “Natural” pruning of the search space
      - Humans generate heuristic choices (“hunches”) that can guide them thru very large search spaces
- Goal: explore computational methods for:
  - Generating and assessing “hunches” (relevance estimation)
  - Learning from experience to improve the relevance estimation process

Proposal: Relevance Estimation

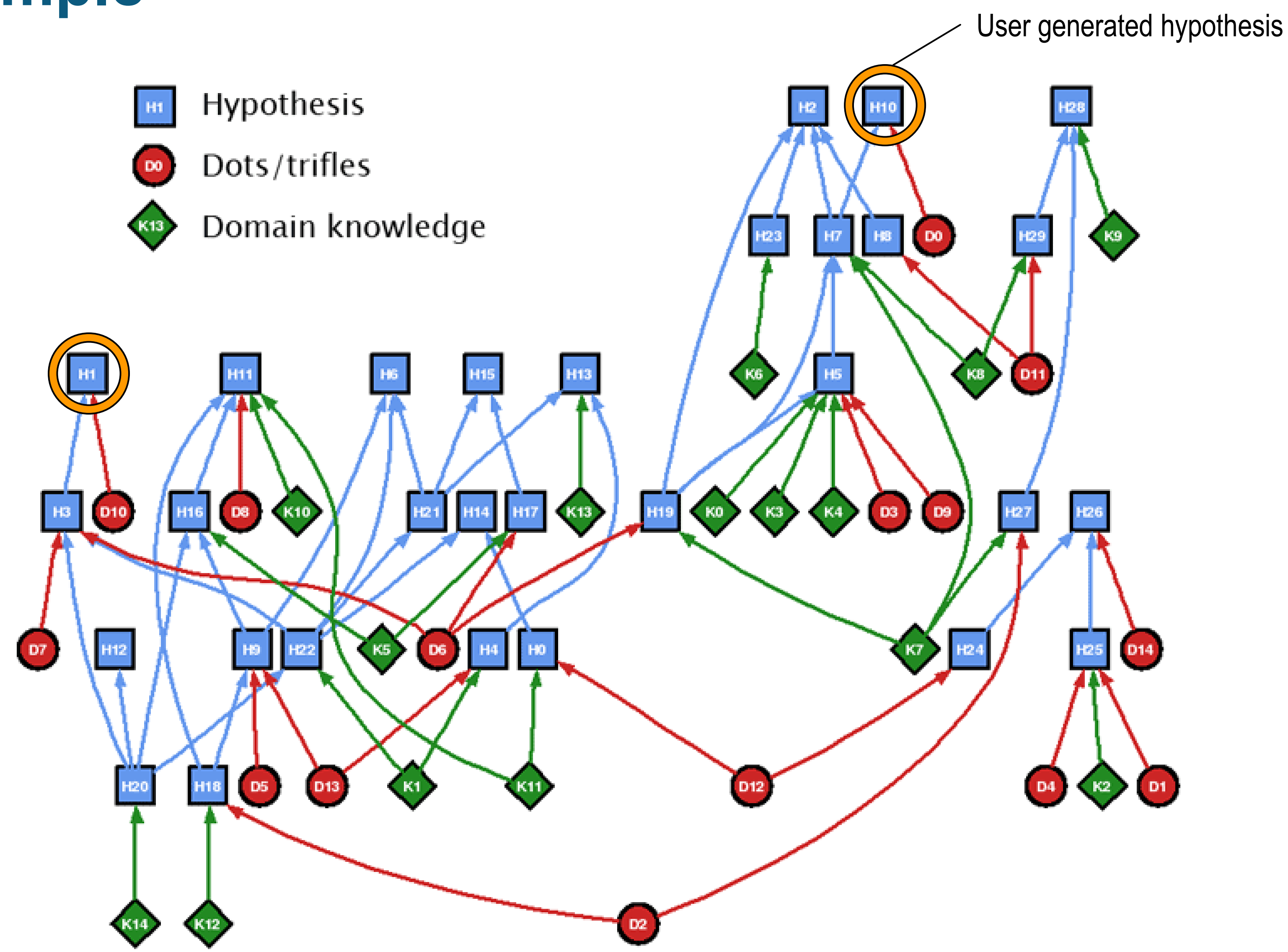
- Goal: Evaluate the relevance/importance of each new assertion
  - Cannot assume system already "knows" the answer or knows that some new inference is an important "dot"
    - Many potential assertions can be generated
    - May not be obvious which assertions will actually be valuable or "relevant" to the problem (Avoid "wishful thinking")
  - Relevance estimation is the "strategy" for prioritizing direction(s) of reasoning and establishing "sufficiency" of a potential direction of reasoning
  - Context is important: Not evaluating in isolation
- Contributing methods:
  - Domain neutral: analogy, deduction, classification, etc.
  - Domain specific: social network analysis, model of user's interest
- Metrics for relevance determination:
  - complexity-based (Occam's razor)
  - graph-based (path length)
  - knowledge-based
    - information scent
    - ontological relationships

Soar Technology
Thinking *inside* the box.

Test bed for RE/EM Research

- Problem:
  - Need large bodies of encoded knowledge to evaluate approaches to RE/EM
  - Access to “real” data is difficult for research purposes
    - security, sensitivity, size, etc.
- Proposed Approach:
  - Test bed for RE/EM
    - Automatically generate potential search spaces
    - Capture size and complexity of the space
    - Systematically vary parameters that represent the “connectedness” of possible hypotheses in the example
    - Simulate relevance across a number of orthogonal and over-lapping dimensions

Example


User generated hypothesis
Hypothesis
Dots/trifles
Domain knowledge

Simulating relevance in the test bed

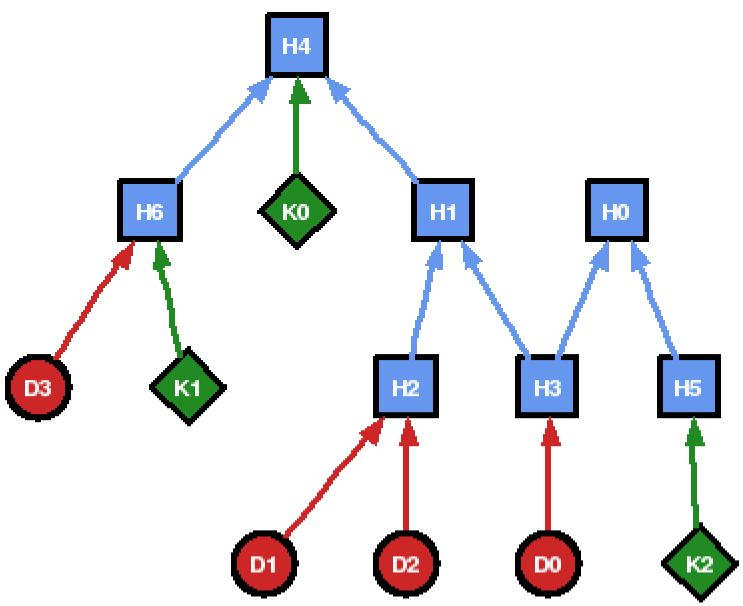
- Relevance:
  - Predicts the likelihood of a relationship between m and n
  - Estimated by aggregating multiple knowledge sources
  - RE $= f(re_0, re_1, re_2, \ldots re_n)$, where re_i is a source of knowledge about the relationship between m and n
- Contributing knowledge sources, re_i, can include:
  - ontological relationships between m and n
  - semantic and syntactic similarity between m and n
  - completeness of analogical mapping between m and n
  - "cognitive" distance between m and n
  - expectation of success of getting to m and n
  - recognition of m in the context of n

Soar Technology
Thinking *inside* the box.

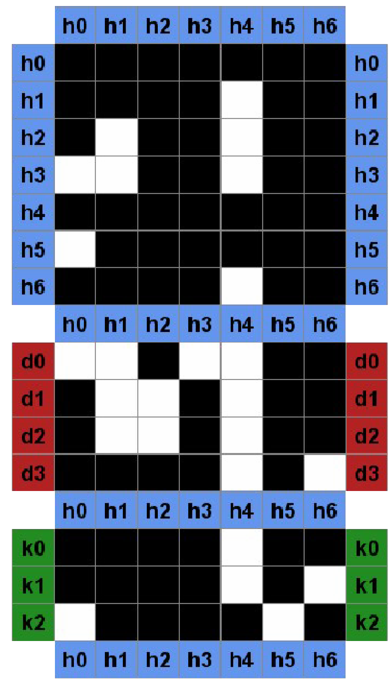
Simulating relevance in the test bed, cont'd

- Some sources can be simulated using graph theoretic operations, e.g., transitive closure in graph implies perfect knowledge of relations between dots, domain facts, and hypotheses.

pEM graph



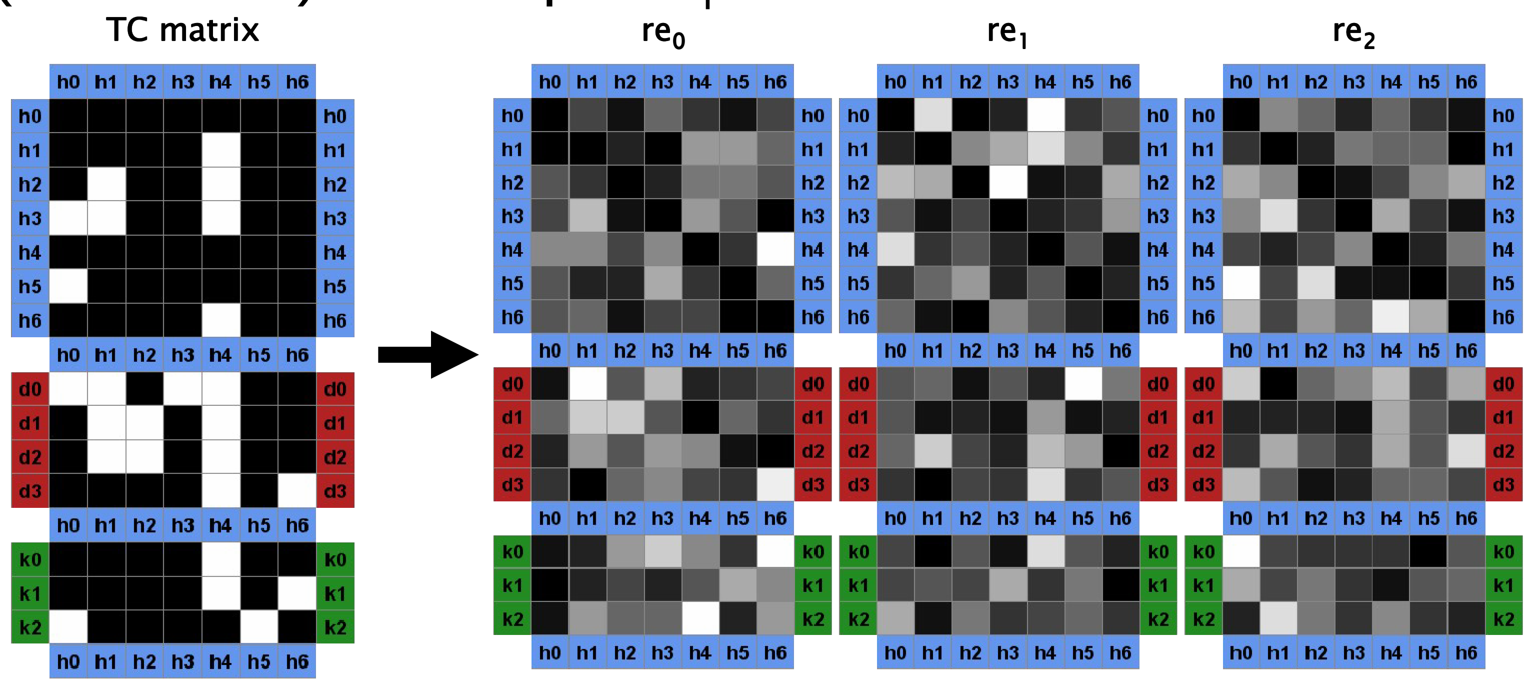
TC matrix



Soar Technology
Thinking *inside* the box.

Simulating relevance in the test bed, cont'd

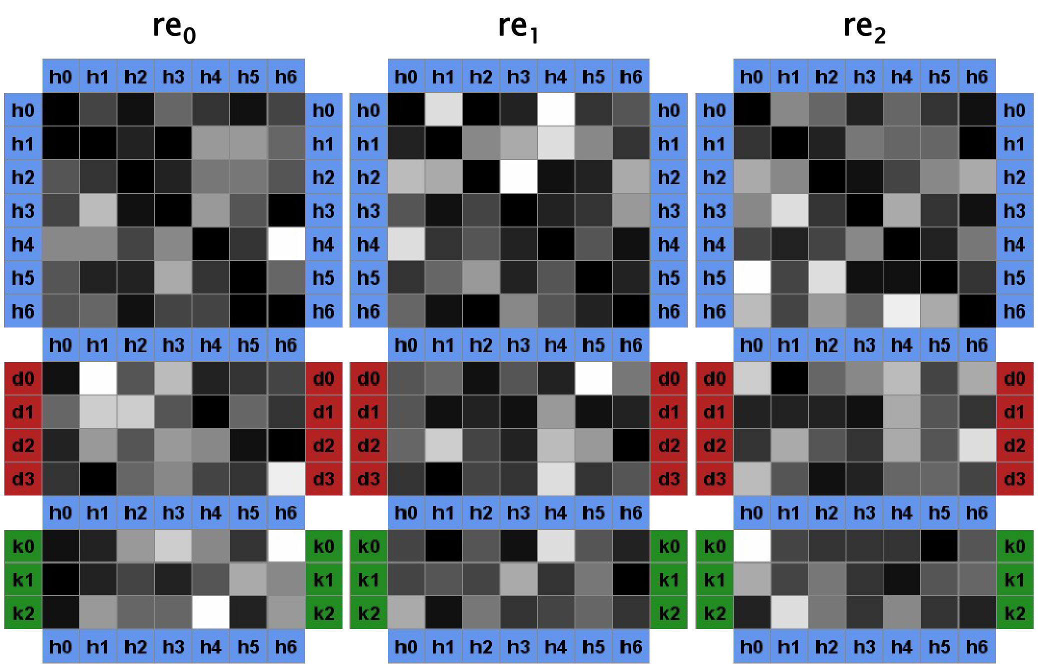
- Real knowledge is imperfect and from multiple, potentially contradicting, sources
- In the testbed, the perfect TC relationships are distributed (with noise) to multiple re_i matrices



Thinking *inside* the box.

Simulating relevance in the test bed, cont'd

- Test bed facilitates:
  - Easy use of knowledge sources represented as a matrices
  - Explorations of how to aggregate noisy and contradictory knowledge



Aggregation options:

$\mathrm{RE}(m,n) = \Sigma c_j * re_i(m,n)$

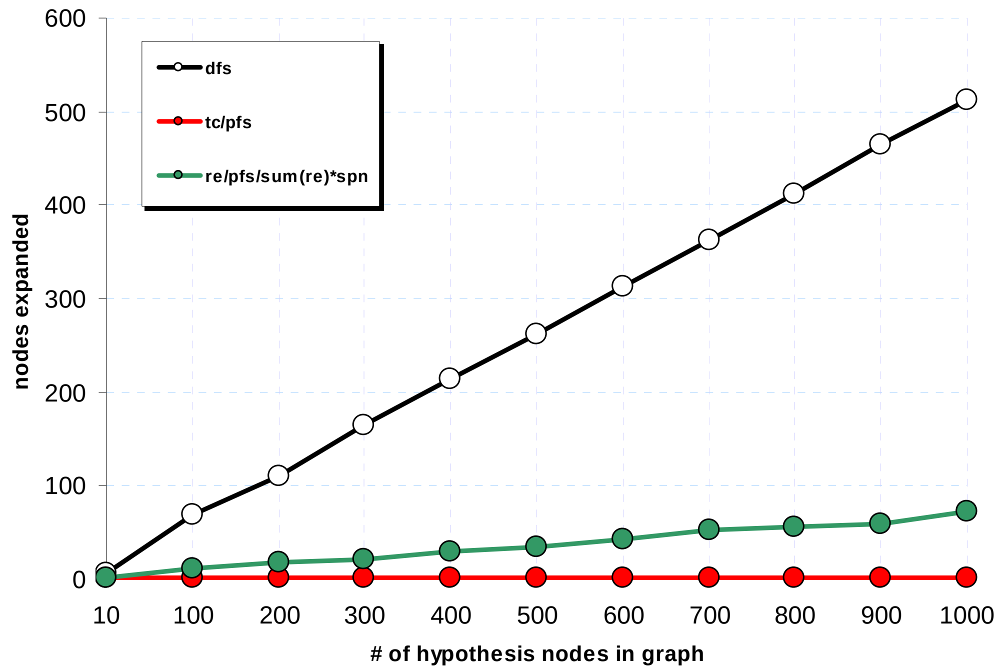
$\mathrm{RE}(m,n) = \max(re_i(m,n))$

$\mathrm{RE}(m,n) = \mathrm{avg}(re_i(m,n))$

$\mathrm{RE}(m,n) = \ldots$

Initial Results (1)

- Search for all nodes in pEM graphs
  - DFS: exhaustive knowledge-free search; upper-bound
  - TC: perfect knowledge give direct path to goal; lower bound
  - RE: RE(m,n) guides priority-first search;
    - 90% fewer expanded nodes in comparison to DFS



Soar Technology
Thinking *inside* the box.

Conclusions

- Humans provide evidence tractable search for EM is feasible and clues for how to do it
- Test beds can be developed for exploring search challenge without large, a priori investments in knowledge encoding
  - Proof-of-concept test bed can generate problems automatically for exploring the consequences of RE design options
  - Plan: Complement with open-source, real-world datasets
- Nuggets:
  - Interesting ideas about how to attack this problem
  - Developed simple testbed for exploring computational approaches
  - Rich literature from which to draw inspiration
- Coal:
  - No feasibility implementation/exploration of solution (“idea ware”)