The Quality of IDA* Heuristics

R. Clausecker, F. Schintke

ZUSE INSTITUTE BERLIN

Introduction

IDA* (iterative deepening A*) is an algorithm to solve the *single source shortest path problem* in large graphs. In contrast to uninformed search algorithms like Dijkstra's algorithm, a problem-specific *heuristic function* $h(v)$ is used to make informed decisions about what nodes to expand.
Finding good heuristics for search problems is a challenging problem often made even harder by difficulties in telling which heuristic of a set of candidate heuristics is best. Using our novel measure of *heuristic quality* η, the performance of heuristic functions can be analysed effectively and the best heuristic for a given problem space chosen.

The Heuristic Quality η

Based on the Korf et al. (2001) formula for the number of expanded vertices $E(v_0, d, P)$ for an IDA* search starting from v_0 to depth d

$$E(v_0, d, P) = \sum_{i=0}^{d} N_i P(d-i) \qquad (1)$$

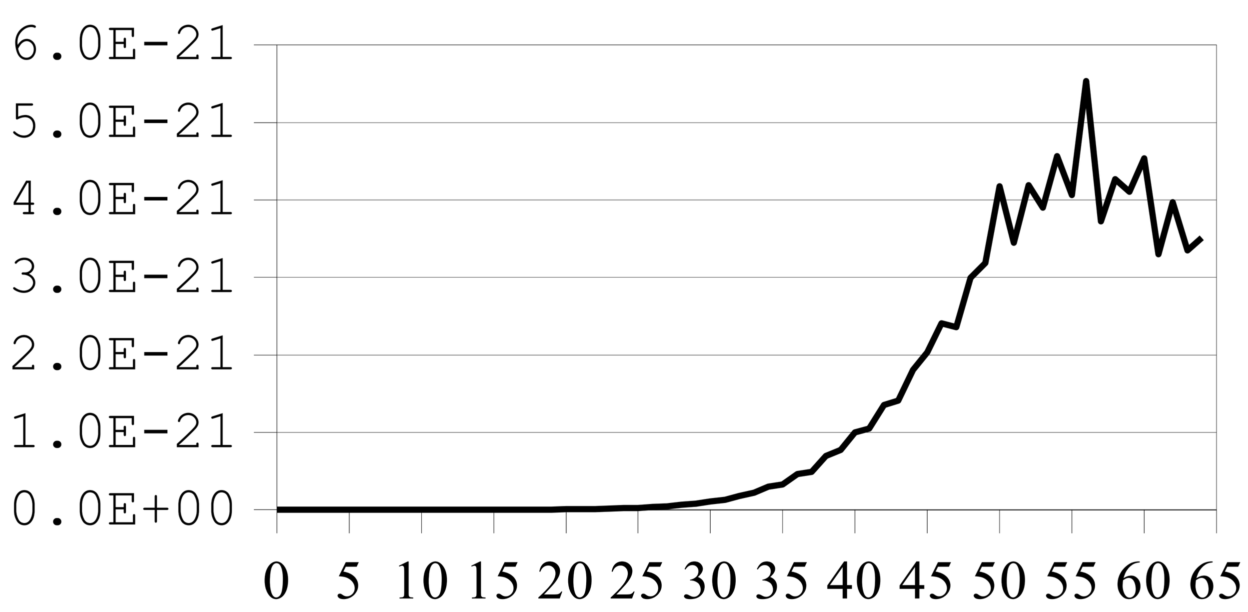
where N_i is the number of vertices at distance i from v_0 and $P(d-i)$ is the probability that $h(v) \leq d-i$ for some random vertex v, we define the *heuristic quality*

$$\eta = \sum_{i=0}^{\infty} \frac{p(i)}{b^i} \qquad (2)$$

With b being the *branching factor* and $p(i)$ being the probability that $h(v) = i$ for some random vertex v. Using η, we can show that the number of expanded nodes is proportional to b^d with η telling us how much the heuristic speeds up the search.

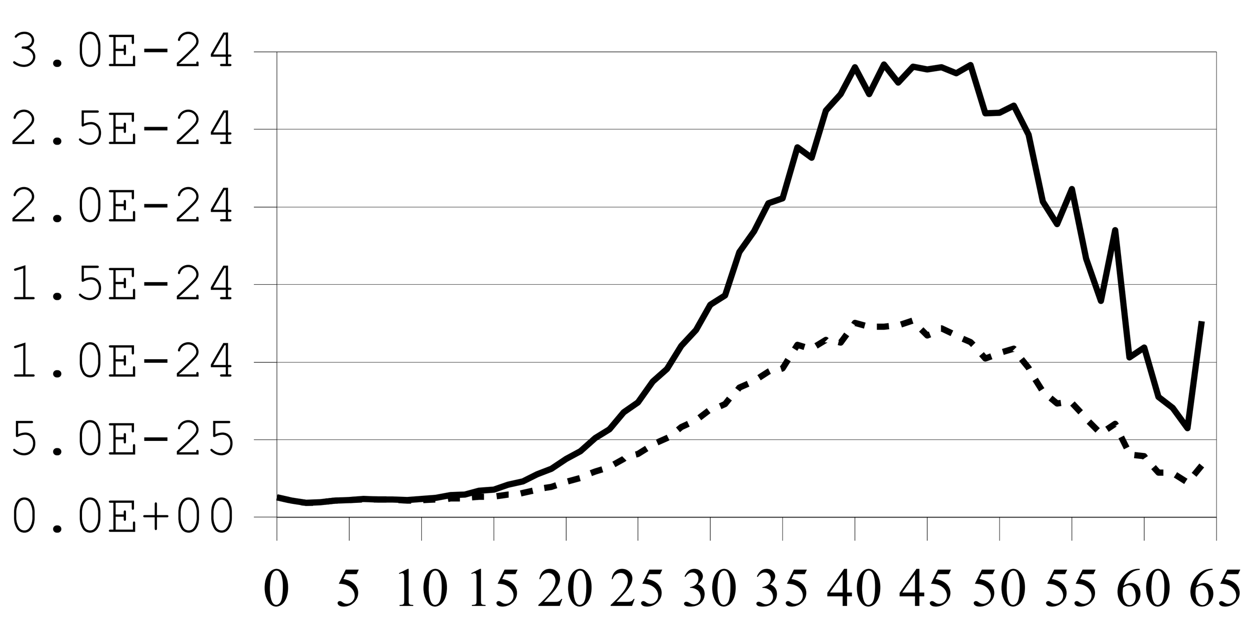
$$E(b, d, P) = \frac{b}{b-1} b^d \eta \qquad (3)$$

Quality Histograms



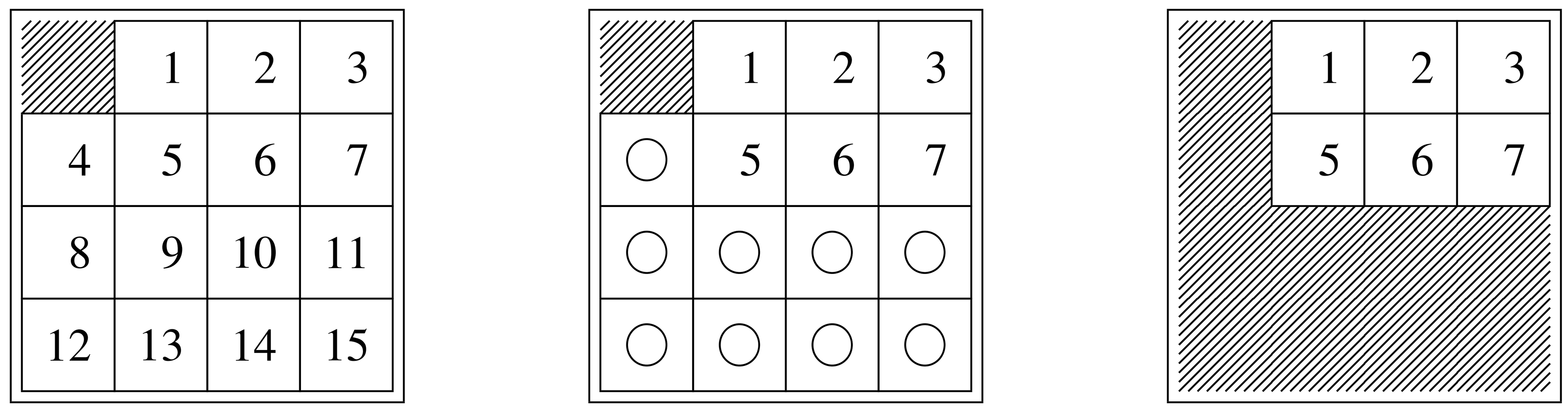
histogram of a poor (Manhattan) heuristic

- plot η by the contribution of each sphere to its value
- large contribution $\rightarrow$ much time spent at this distance
- good heuristics spend more time closer to the goal
- answers long standing question: h values of what part of the search space are most critical to performance?
- many heuristics have tweakable parameters
- so tweak them according to the histogramm



histogram of two good (PDB collection) heuristics

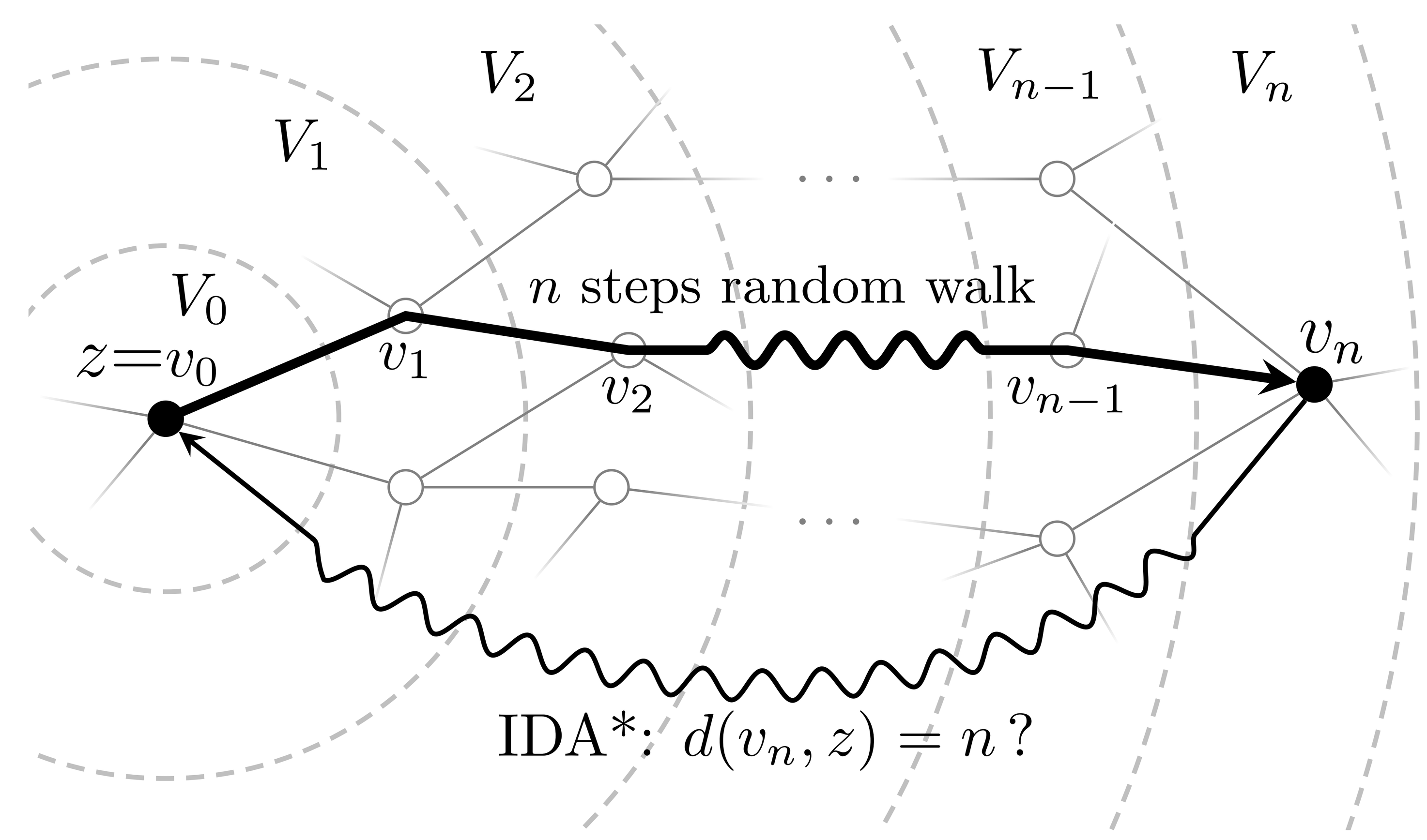
Heuristic Functions: Pattern Databases



15 puzzle | non-additive PDB | additive PDB

Pattern Databases: powerful IDA* heuristics obtained by taking a subset of the problem configuration and tabulating the solution lengths for the entire problem space. If designed carefully, h values of disjoint pieces of problem state can be added up to a powerful *additive PDB heuristic*.

Sphere-Stratified Sampling



The quality of arbitrary heuristics is computed by treating it as an expected value

$$\eta = \mathrm{E}[b^{-h(v)}] \qquad (4)$$

over all vertices in the search space. For this purpose, we developed *sphere stratified sampling*, a novel sampling scheme were the search space is stratified into *spheres* of vertices with a given distance to the goal vertex v_0. Samples are taken according to the scheme

1. perform a random walk of n steps from v_0 to obtain a candidate vertex v_n
2. enumerate all shortest paths from v_0 to v_n; if there is a path of less than n steps, discard v_n
3. otherwise, take v_n as an observation and compensate for the sampling bias by computing the probability with which v_n has been reached by the random walk from v_0

Although the sample is biased, the bias can be compensated for effectively, leading to low margins of error on η with reasonable sample sizes.

Publication

To appear as: R. Clausecker, F. Schintke. 2021. The Quality of Heuristic Functions for IDA*. In *Proceedings of the Fourteenth International Symposium on Combinatorial Search (SoCS 2021)*.