Phase Transitions, Backbones, Measurement Precision and Phase-inspired Approximation

Weixiong Zhang

Computer Science Department
Washington University, St. Louis

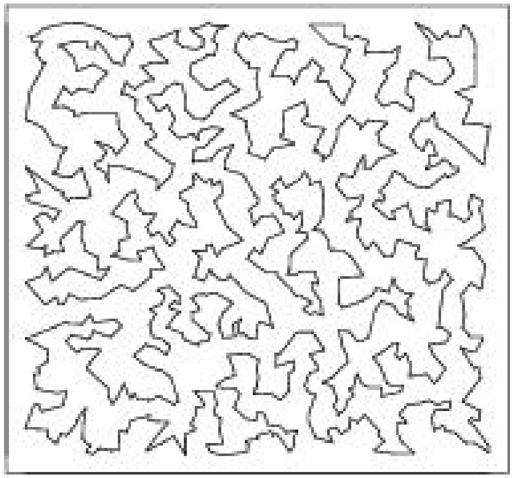
http://www.cs.wustl.edu/~zhang
email: zhang@cs.wustl.edu

Road Map

- Background and insight from tree search
  - The Asymmetric TSP (ATSP) and branch-and-bound
  - Incremental tree model
  - Phase transitions in optimization
- Phase transitions of ATSP
  - Complexity
  - Number of optimal solutions
  - Backbone
- A phase-inspired approximation
- Generality of the results
  - Number partitioning
  - MAX-SAT

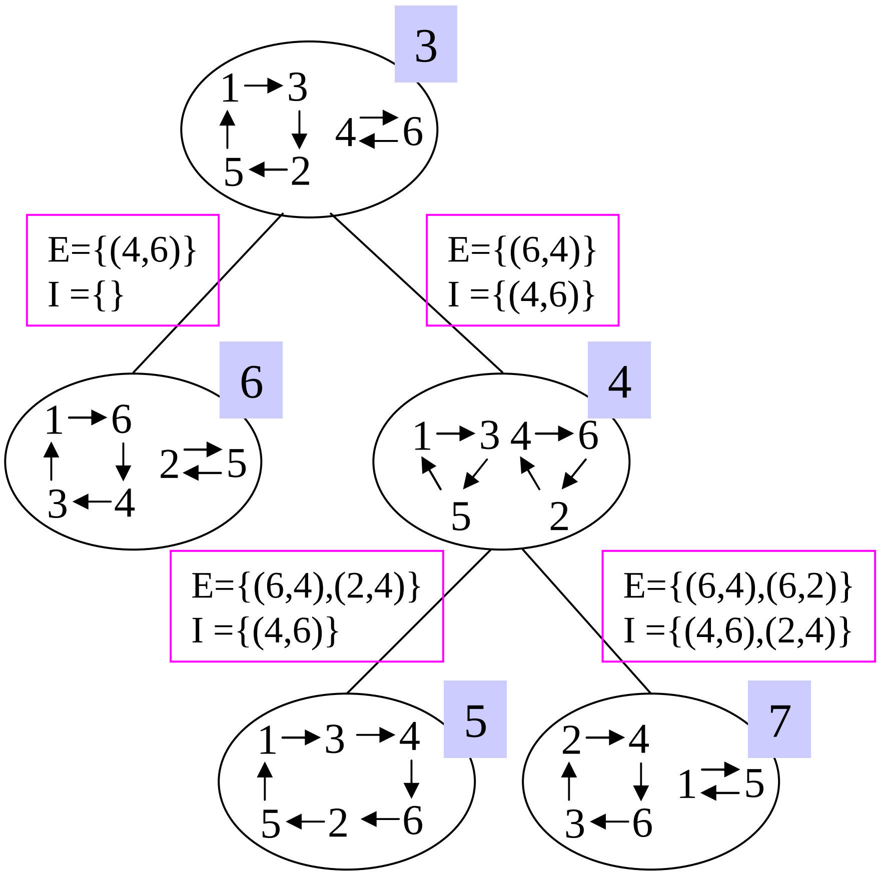
The Traveling Salesman Problem

- TSP: Given n cities and all intercity distances d(u,v), find a shortest complete tour.
- Asymmetric TSP (ATSP): TSP in which not necessarily d(u,v)=d(v,u)
  - More difficult than symmetric TSP
- Many real applications
  - Scheduling, VLSI routing, etc.
- NP-hard
  - NP-completeness is a worst-case measure
  - How does the average-case complexity behave?



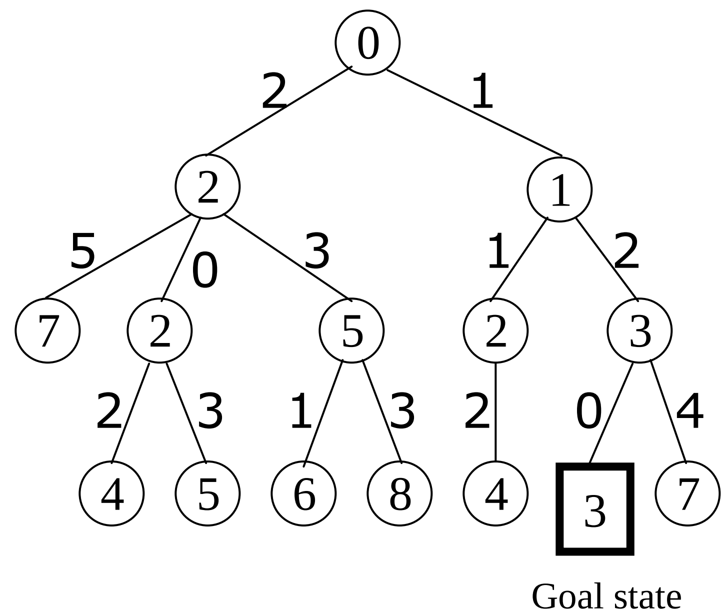
Branch-and-Bound for the ATSP

- Lower bound
  - Assignment problem (AP): ATSP with no complete tour restriction
  - Solvable in O(n^3) time
  - The # of APs solved is the # of node visited in search
- Upper bound
  - Karp's patching algorithm – merge subtours in AP solution
- Branching rules
  - Inclusion and exclusion principle
  - Break the smallest subtour
- Search space is a tree



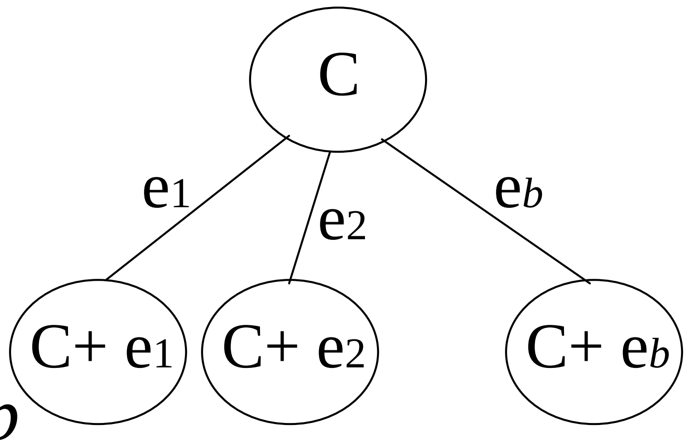
Analytic Model for Tree Search

- Incremental random tree
  - Node = states
  - Branching factor *b*
    - number of child nodes
    - Random variable
  - Edges = state transitions
  - Edge costs (operator costs): i.i.d non-negative random variable (may take 0)
  - Node costs (state quality): sum of edge costs from the root to the nodes
  - Tree depth *d*
  - Goal state: minimal cost node at the tree depth



Complexity Analysis: Order Parameter

- Order parameter:
  - b = the mean branching factor
  - p = probability that an edge has cost 0
  - bp = expected number of children having the same cost as the parent
- The complexity is determined by bp
  - A local property determines a global behavior!



Complexity Analysis: The Main Results

- **With a high probability, the # of nodes generated as *d* $\rightarrow\infty$**

algorithm	*bp*<1	*bp*=1	*bp*>1
Best-first search	$\Theta(\beta^d)$ optimal	$\Theta(d^2)$ optimal	$\Theta(d)$ optimal
Depth-first branch and bound	$\Theta(\beta^d)$ asymptotic optimal	$O(d^3)$	$O(d^2)$

Karp+Pearl, 1983
McDiarmid, 1992
Zhang+Korf, 1993

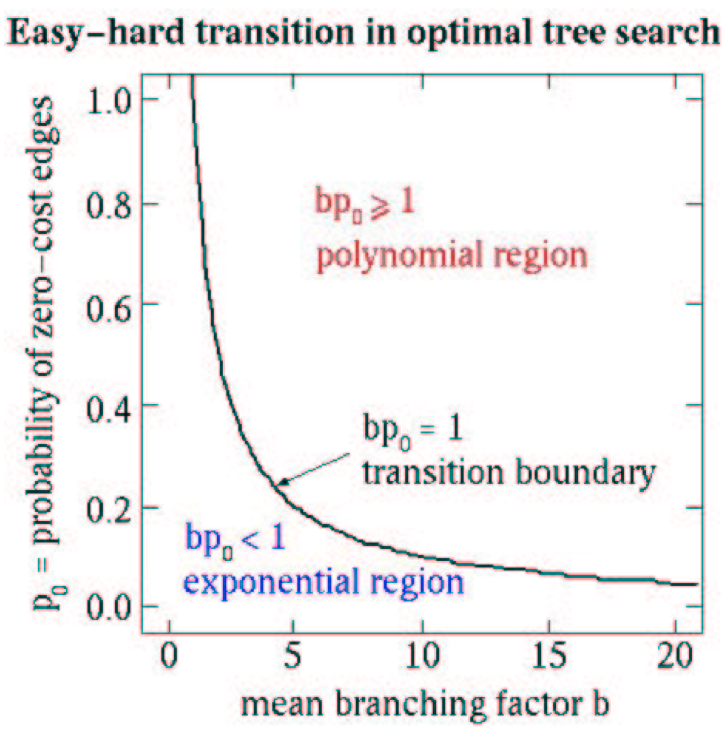
- **With a high probability, the optimal solution cost C* as *d* $\rightarrow\infty$**

bp<1	*bp*=1	*bp*>1
$C^*/d \rightarrow \alpha$	$C^*/\log\log d \rightarrow 1$	C* bounded by a constant

β is a constant > 1

Phase Transitions and Phase Diagram

- Order parameter *bp* = expected number of children having the same cost as the parent
- Phase transition: exponential vs. polynomial



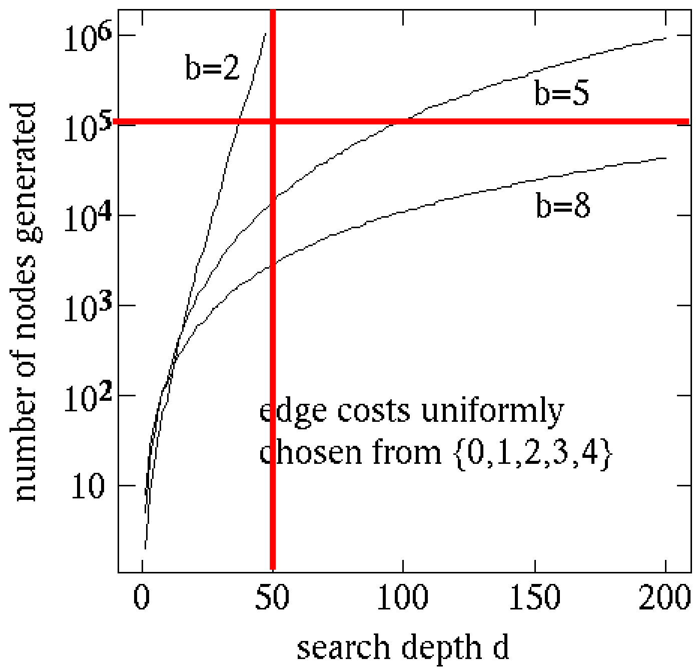
Zhang+Korf, 1995

Complexity Anomaly ?

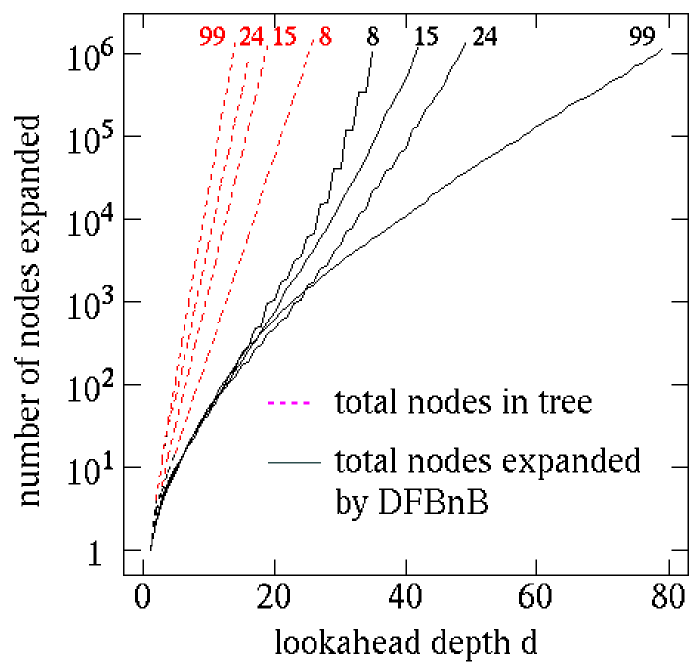
Zhang+Korf, 1995

- A tree with a smaller branching factor is more difficult to search than one with a larger branching factor

- Random trees
  - edge costs uniformly from {0,1,2,3,4}

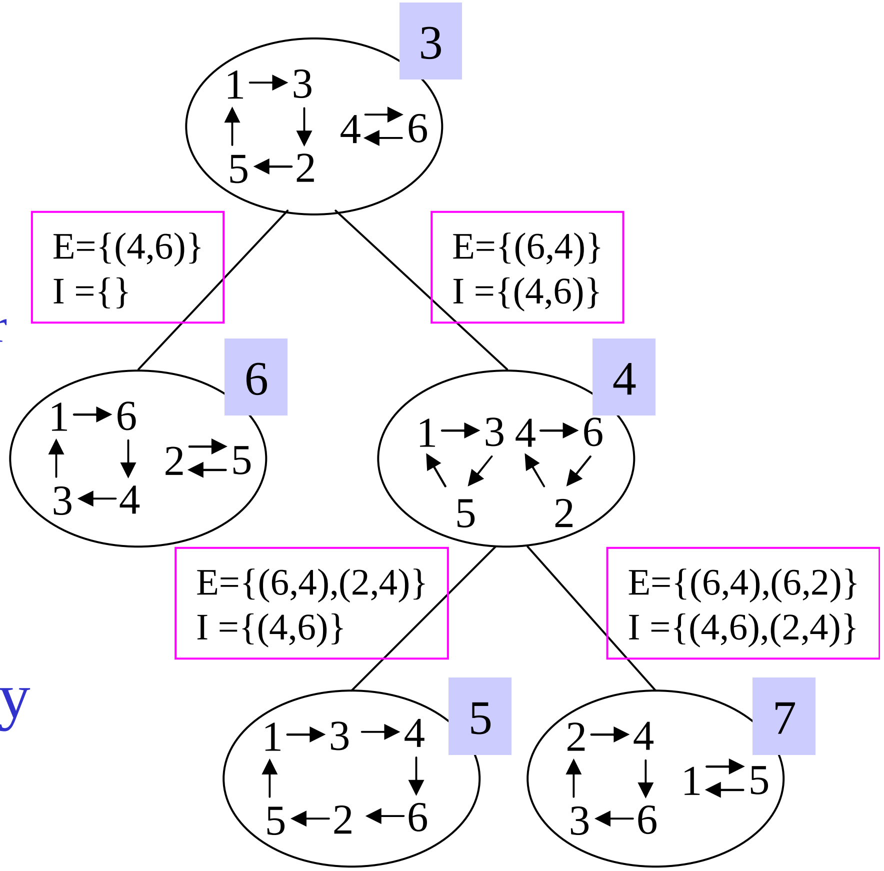


- Sliding-tile puzzles
  - Backup values at a fixed depth (e.g., playing chess)



Implication of Theoretical Results on ATSP

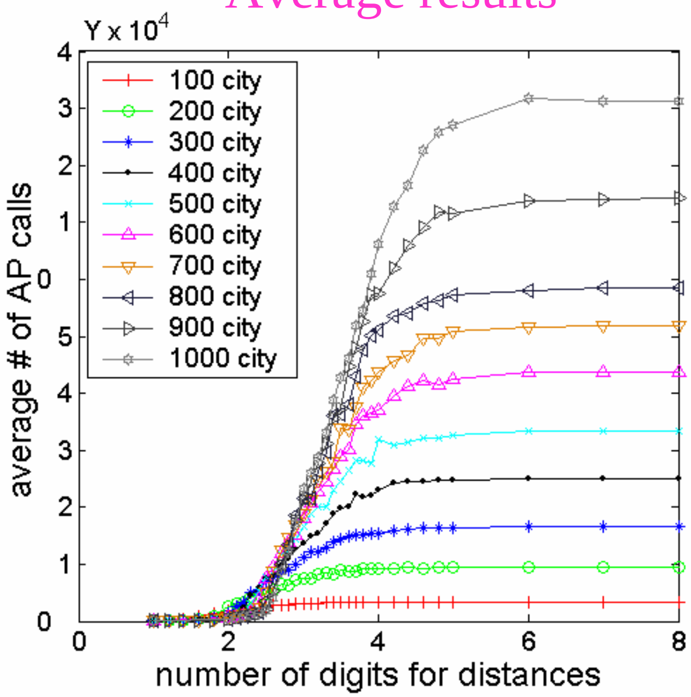
- Implications
  - Edge costs in search space
    - Less the number of distinct values of edge costs, easier the ATSP
  - AP solutions
    - More nodes have the same costs as their parents, less the number of unique edge costs in search tree
  - Precision of distance measurement
    - More nodes have the same AP costs as their parents as the measurement precision is reduced
- Distance precision affects complexity
- Phase transitions in the ATSP?
  - Distance precision may be a control parameter
  - But, the assumptions in the analytic model (e.g., i.i.d.) will be violated



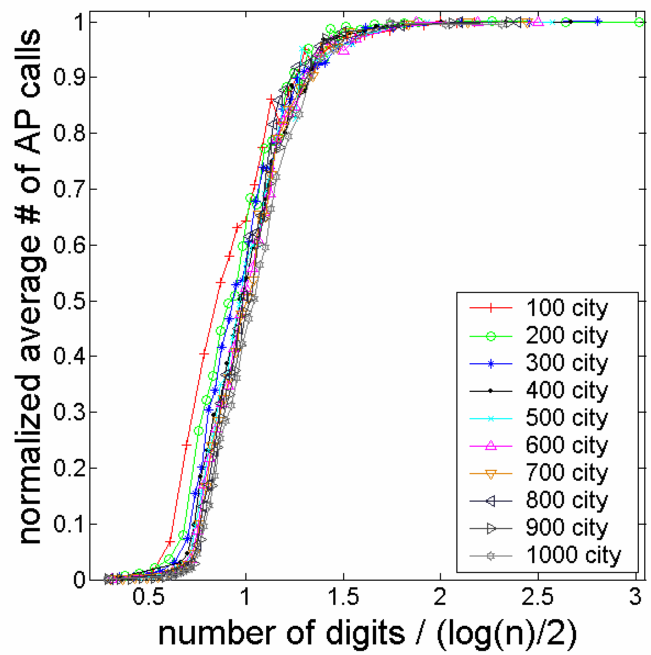
Phase Transitions – Complexity

- Random problem instances
  - 1000 instances with intercity distances being uniform, random variables
- Control parameter:
  - # of digits for intercity distances

Average results



Normalized average results

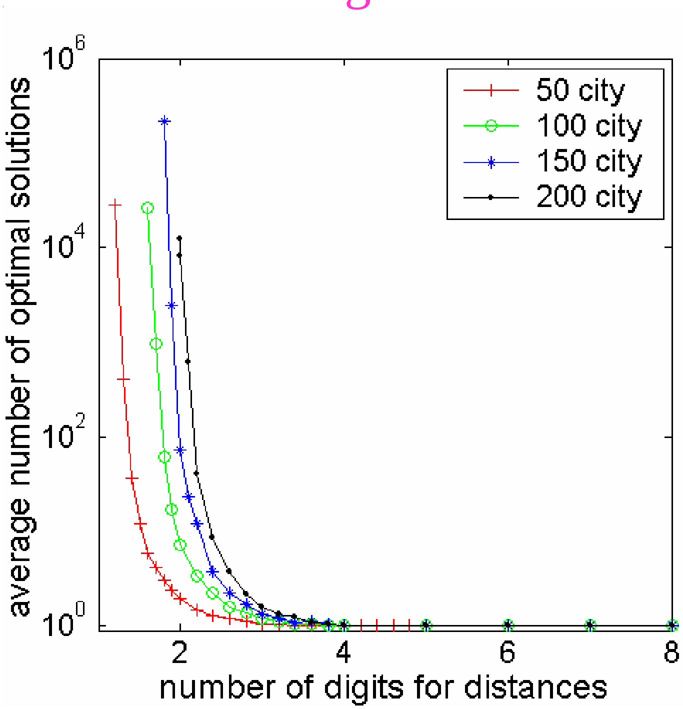


Phase Transitions – Number of Solutions

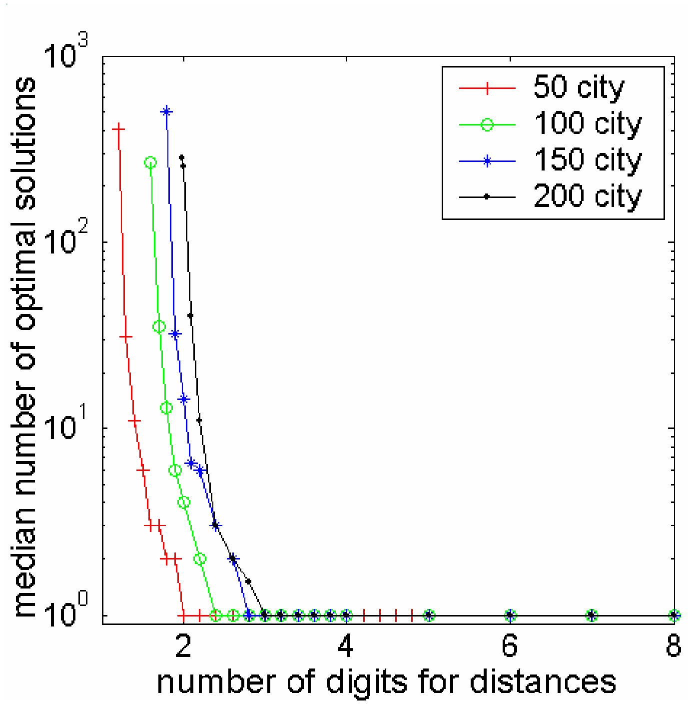
used a method in Climer+Zhang, 2002

- The # of optimal solutions decreases exponentially with the control parameter (# of digits for intercity distances)

Average results



Median results

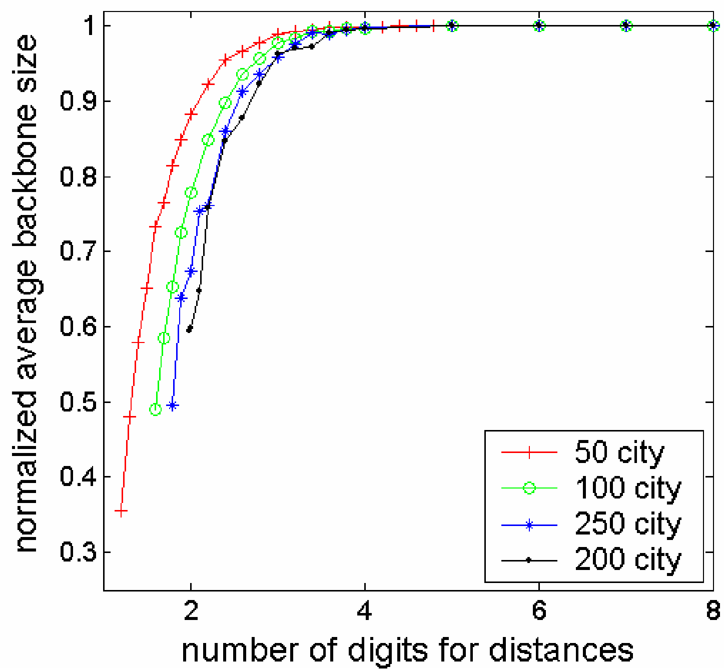


Phase Transitions – Backbones

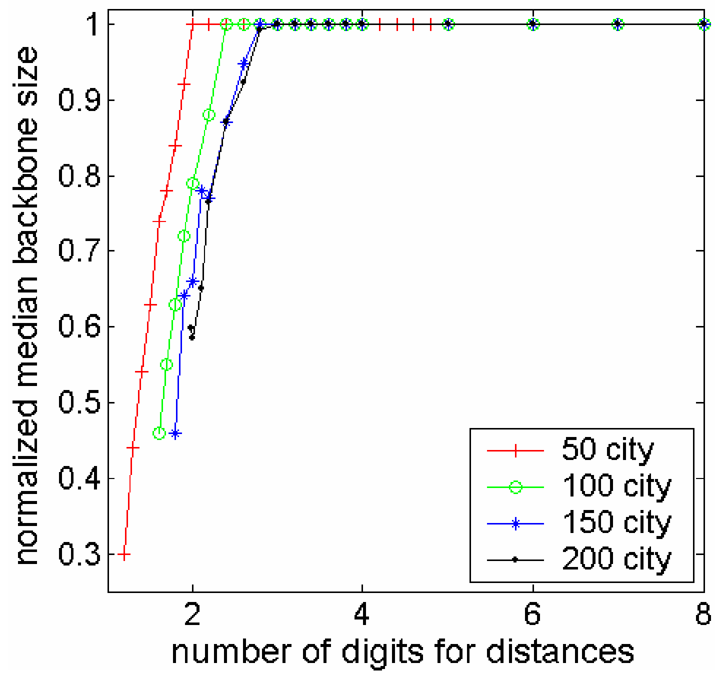
used a method in Climer+Zhang, 2002

- Backbones emerge quickly as the control parameter increases
  - Backbone variable: an edge appears in all optimal solutions

Average results



Median results



Phase Transitions and NP-Completeness

- The ATSP is NP-complete in a strong sense
  - Hamiltonian cycle is a special case
  - The TSP with distances 1 and 2 is still NP-complete
- Contradiction ?

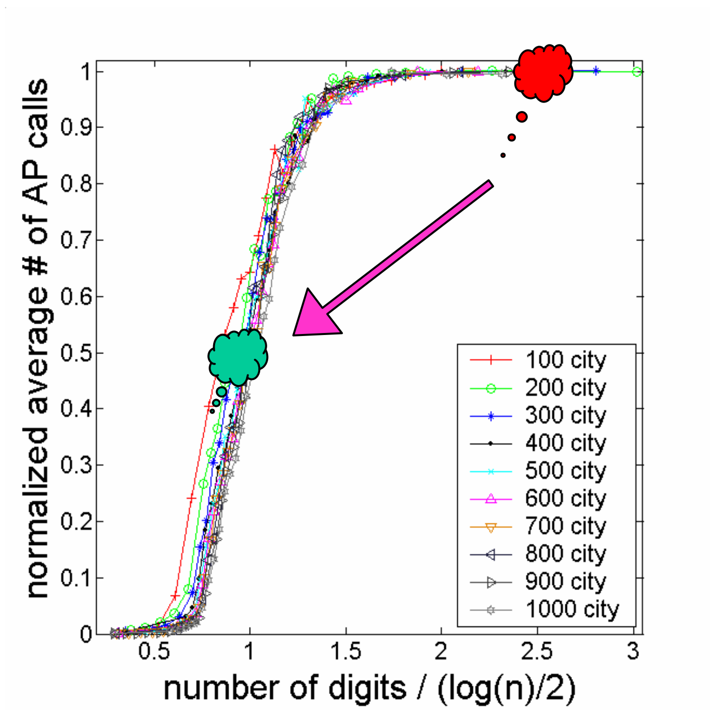
- No
  - NP-completeness is a worst-case measure
  - Phase transition is a typical-case measure
  - A worst-case instance may appear in an easy region

Why Study Phase Transitions?

- characterize difficult problems and their complexity
  - Characterize typical cases
  - Identify and locate difficult instances
- New problem solving methods
  - Important, but seem to be hard
  - Yet to attract sufficient attention !
    - Epsilon-transformation (Zhang+Pemberton, 1994)
    - Heuristic based on backbones for 3-SAT (Dubios, 2001)
    - Phase-inspired approximation (this talk)

Phase-inspired Approximation – The Idea

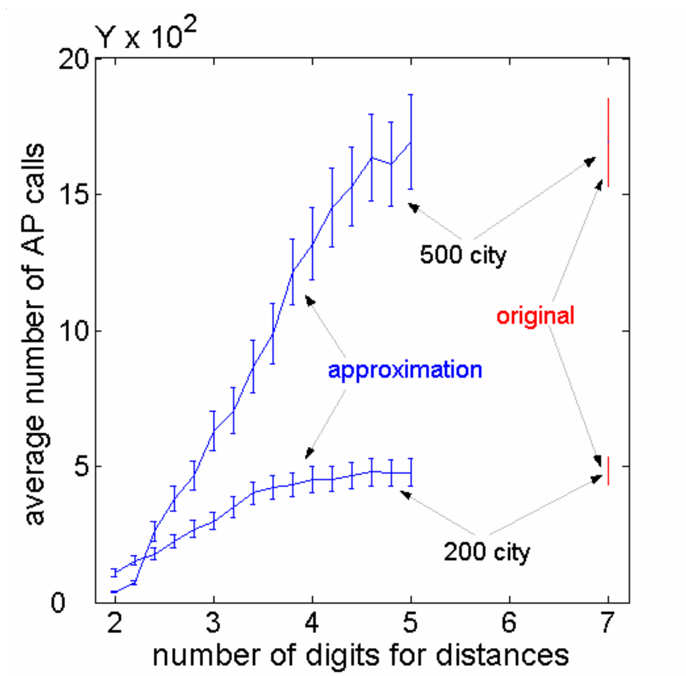
- Question: Can we exploit phase transitions for finding high-quality approximate solutions with reduced computation?
- Main idea:
  - Transform an instance in the difficult region to one in the easy region by educing the precision of distance measure
  - Search in the transformed space
  - Return the optimal solution in the transformed space along with the cost in original distance measure



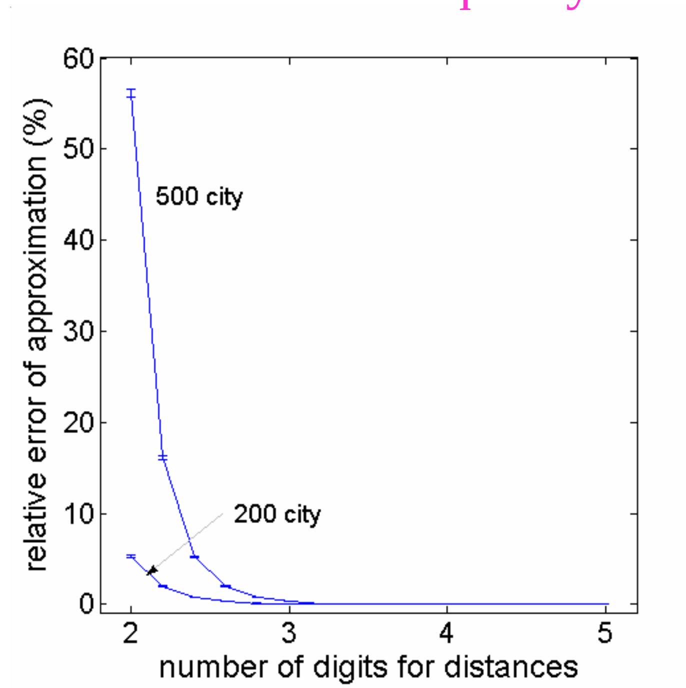
Phase-inspired Approximation – Results (1)

- The method
  - Reducing the number of digits used
  - Search for optimal solution in the transformed cost space
  - Return the optimal solution and its cost in the original cost space
- Results from 1000 random instances with 7 digits for distances

Gain: reduction on search cost



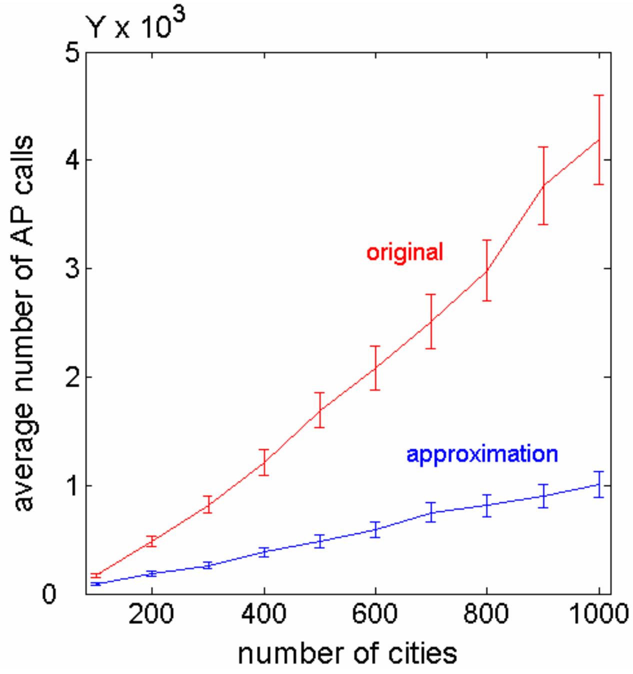
Price: reduced quality



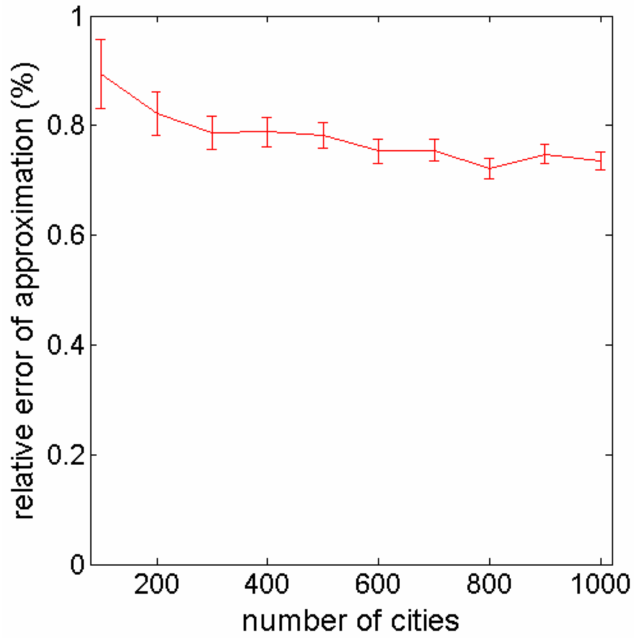
Phase-inspired Approximation – Results (2)

- Transform to the middle point of a phase transition
- Results from 1000 random instances with distances in 7 digits

Gain: reduction on search cost



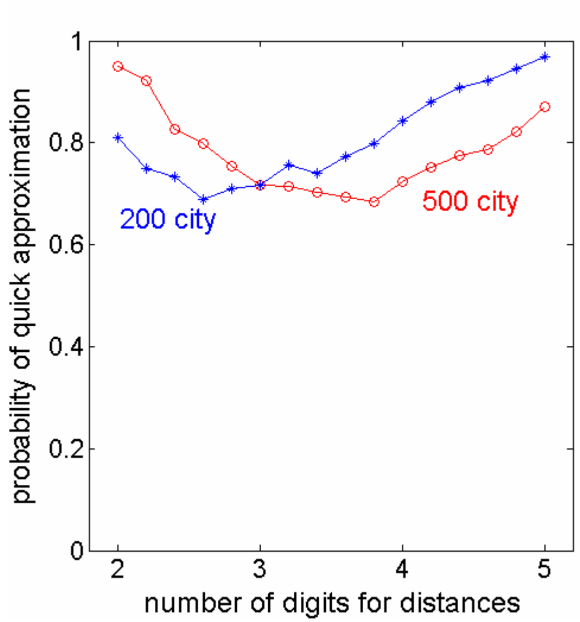
Price: reduced quality



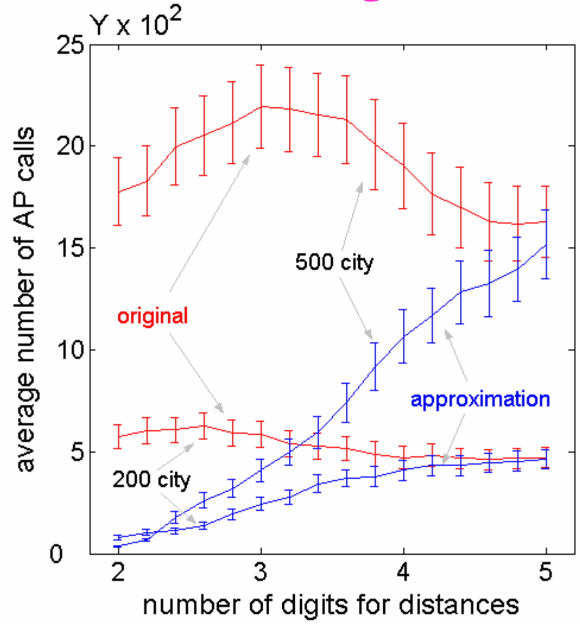
Phase-inspired Approximation – Ouch (1)

- A catch: Not every transformation will reduce complexity
  - You gamble; What is the chance you win?
- A closer look: exam the good and bad transformations separately
  - Winning probability > 0.68
  - Reduction is more than increment on complexity at transition point

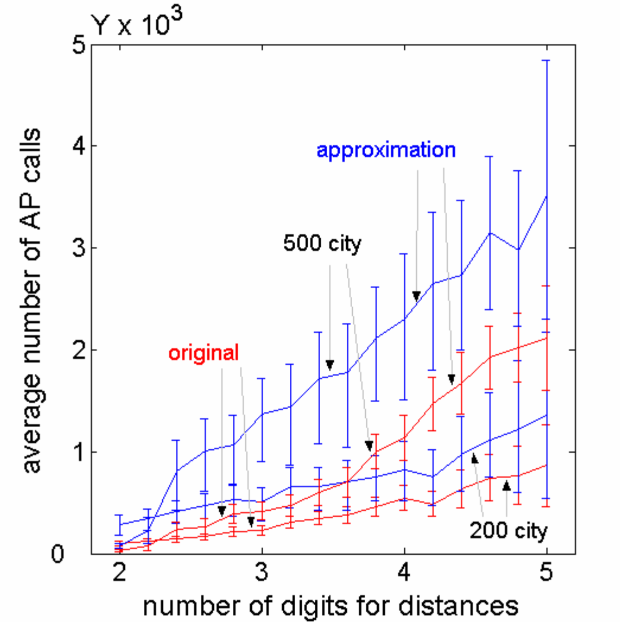
Winning probability



Reduction on good cases



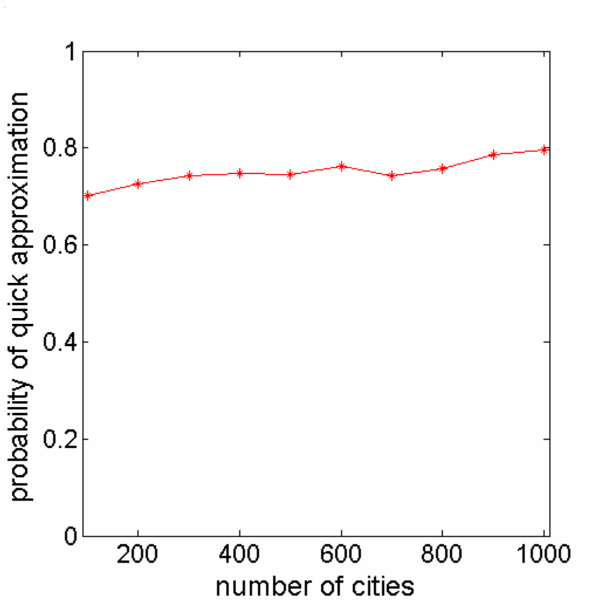
Increment on bad cases



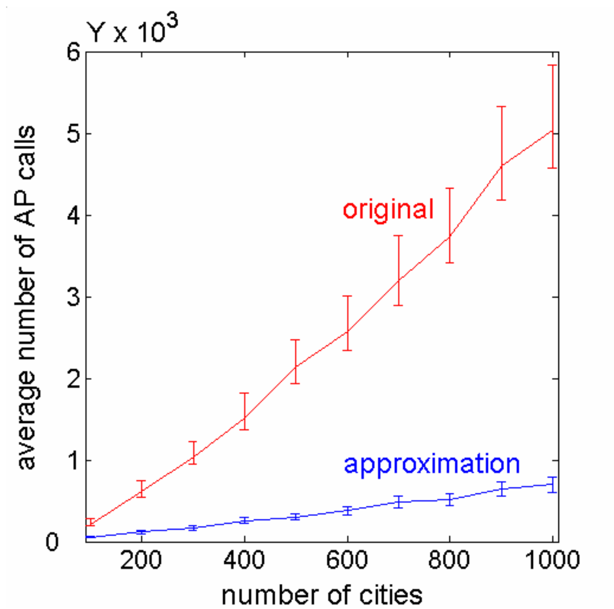
Phase-inspired Approximation – Ouch (2)

- Separate good and bad transformations as problem size changes
- Winning probability > 0.7
- Reduction on complexity is more than the increment

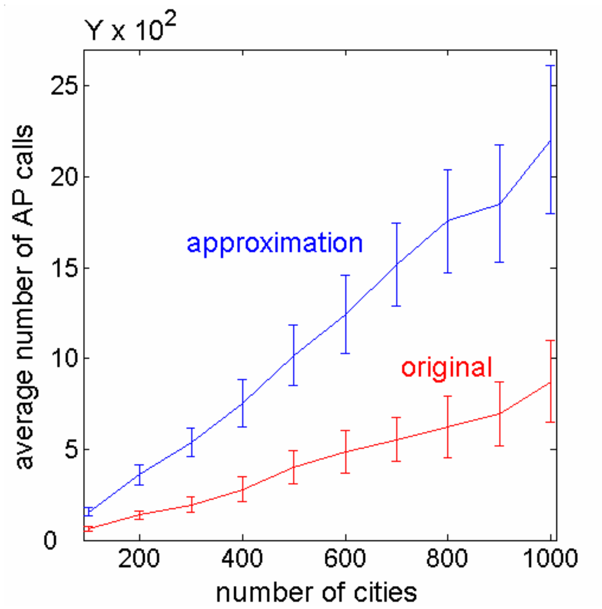
Winning probability



Reduction on good cases



Increment on bad cases



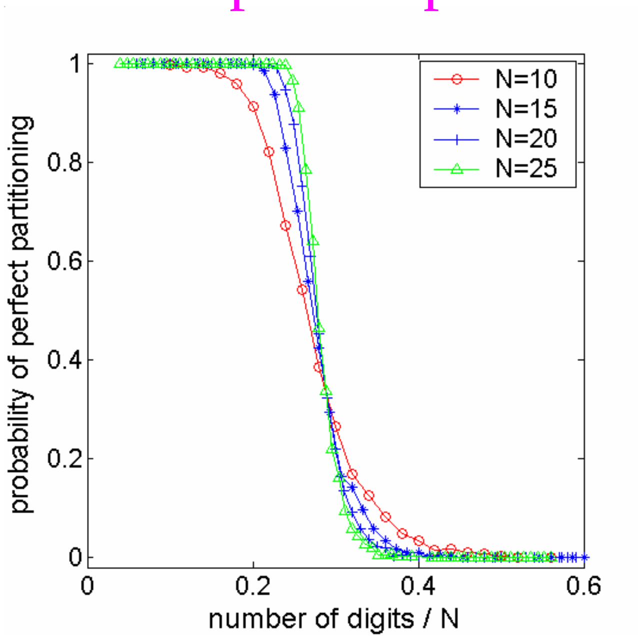
Generality

- Similar phase-transition and backbone results on other combinatorial optimization problems
  - Number partitioning (early by Gent+Walsh, Mertens)
  - MAX-SAT (Zhang, 2001)
    - Note: this is not a number problem

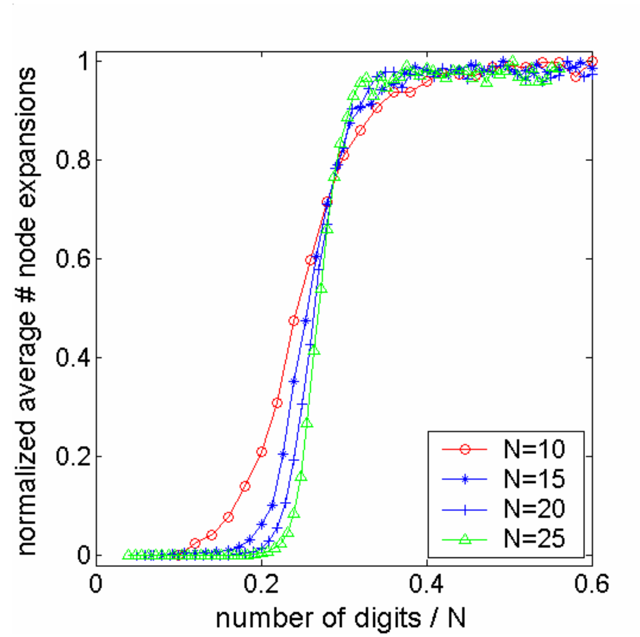
Phase Transitions in Number Partitioning

- Partition *n* integers into two subsets so that the difference of their total sums is minimal possible.
- Perfect partition: a partition with difference 0 or 1.
- Algorithm: Complete Kamakar Karp (CKK) by Korf (Korf, 1996)
- Experiments: 1000 instances with integers independently, uniformly chosen

Prob. of perfect partition

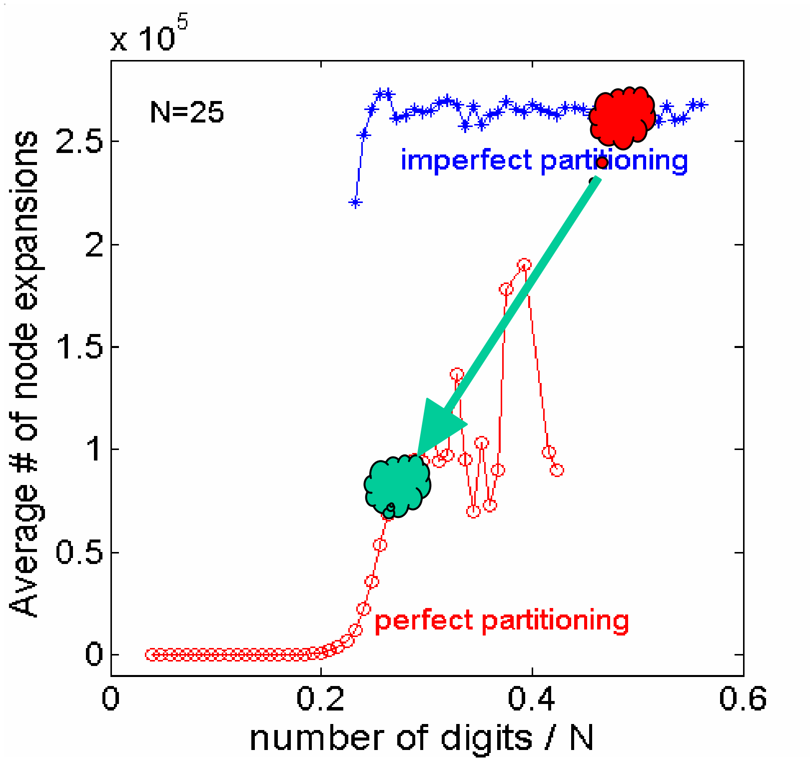


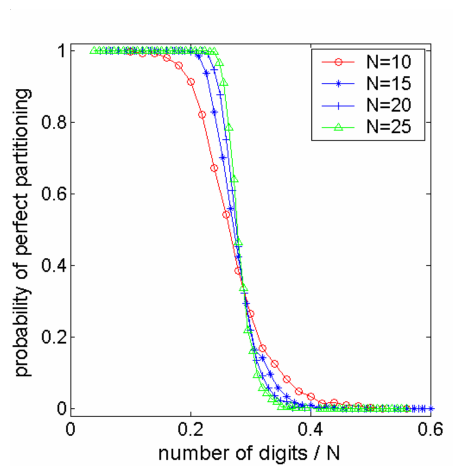
Normalized search cost

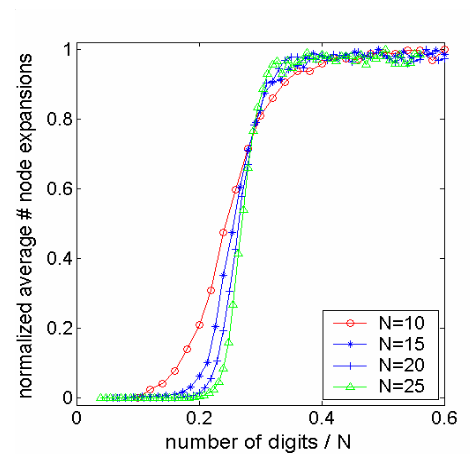


Approximation on Number Partitioning

- A closer look at perfect and imperfect partitions
- Approximation by manipulating precision



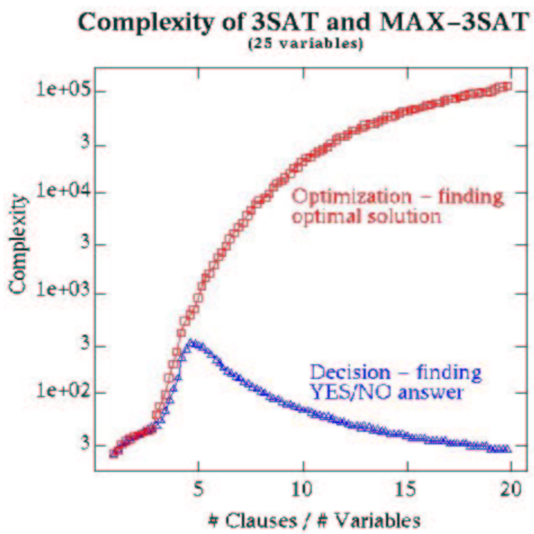




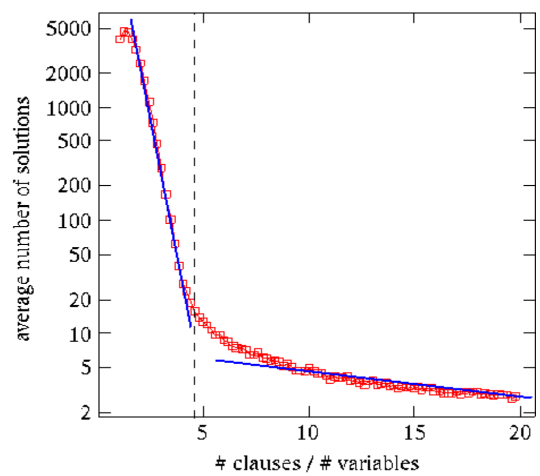
Phase Transitions and Backbone of MAX-SAT

25 variables, 1000 instances

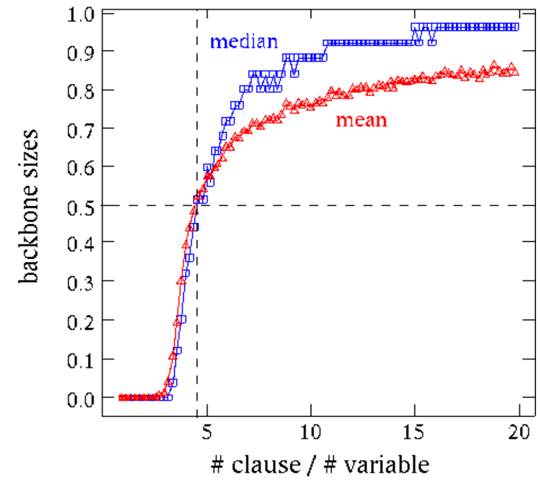
Zhang, 2001



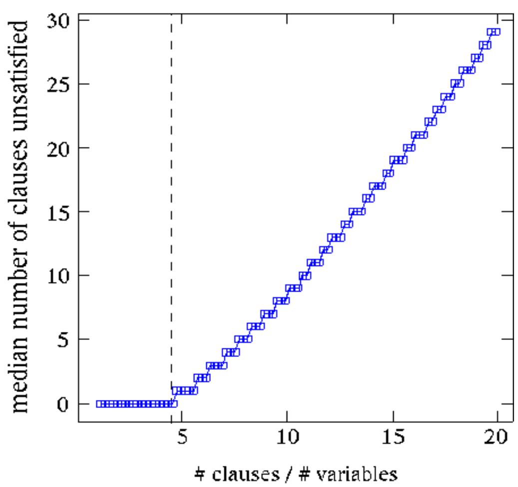
Number of optimal solutions



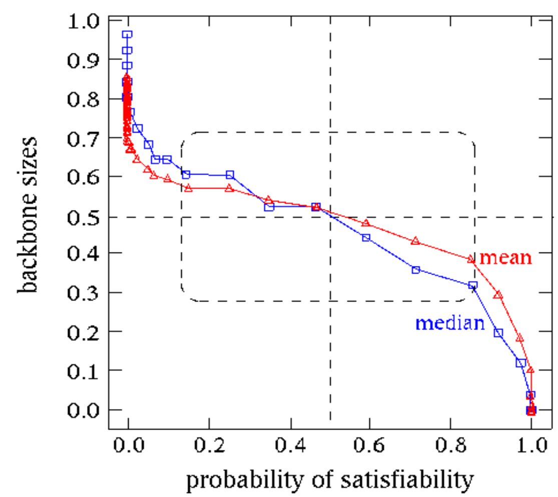
Mean and median backbones



Median cost of optimal solutions



Collocation of two phase transitions ?



Take Home Message (Summary)

- Optimization problems usually exhibit easy-hard phase transitions
  - v.s. the easy-hard-easy phase transitions in decision problems (such as SAT)
- Average-case complexity is usually determined by measurement precision
  - Number of digits used is a good control parameter
- Phase transitions can exploited to develop more efficient search algorithms
    - Phase-inspired approximation