When are Hard Problems *Genuinely* Hard? Phase-transition Phenomena and the Implications for Efficient Algorithms

Paul E. Dunne, {Alan Gibbons, Michele Zito}
Department of Computer Science
University of Liverpool
Liverpool L69 7ZF
Great Britain

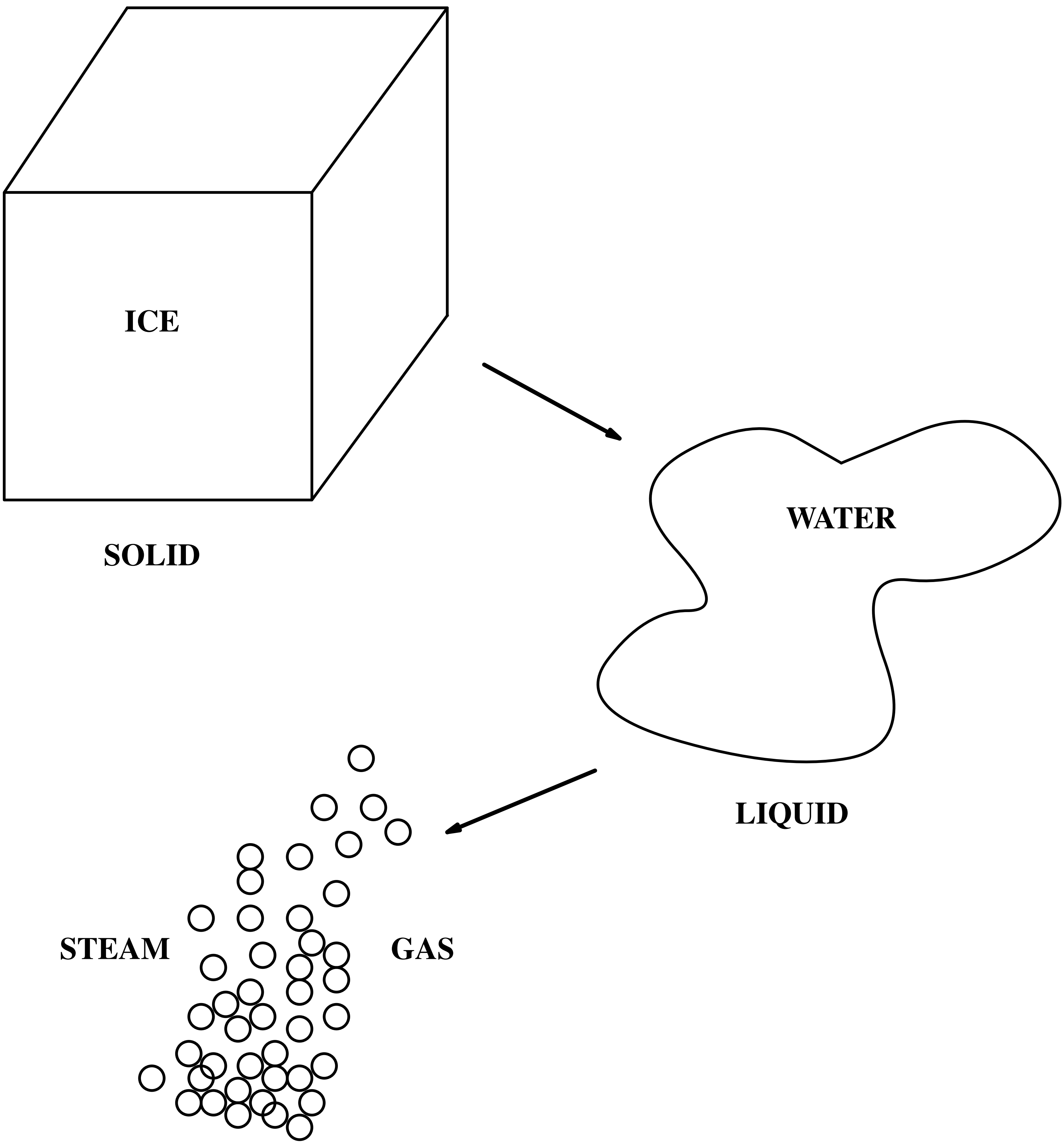
(EPSRC Grant: GR/L/77089)

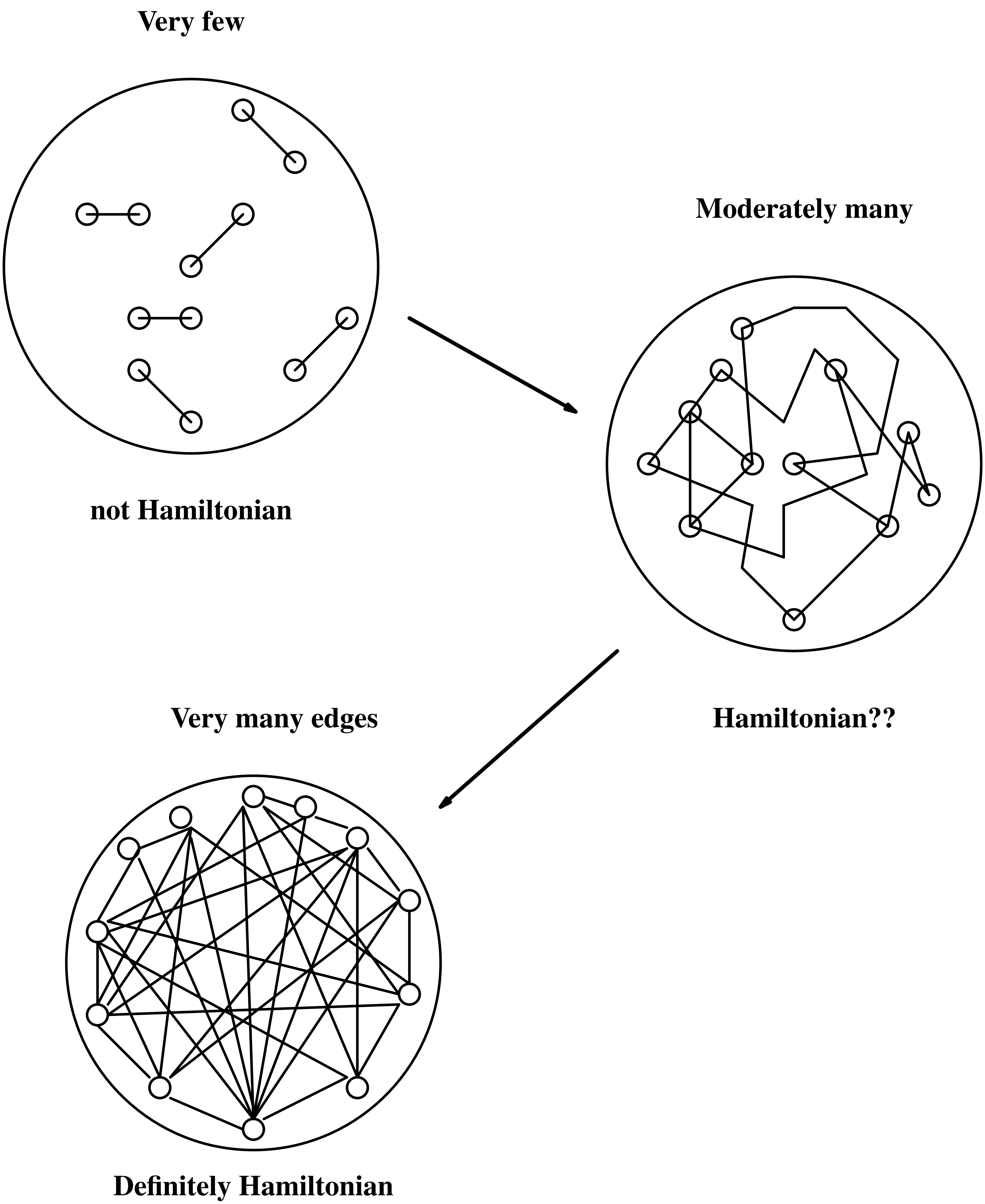
Forsan et haec olim meminisse iuvabit

Organisation of Talk

- Phase Transitions - what are they?
- Algorithmic and Combinatorial Issues
- Investigative Methods
- Open issues and future work.
- Conclusions.

-- --


ICE
SOLID
WATER
LIQUID
STEAM
GAS


Very few
not Hamiltonian
Moderately many
Hamiltonian??
Very many edges
Definitely Hamiltonian

A transition between different states occurs as some ‘measure’ increases.

In the physical example:

- The ‘measure’ is ‘temperature’.
- The transition occurs at sharply defined thresholds:

  Melting point ($0^{\circ} C$)

  Boiling point ($100^{\circ} C$)

What about the ‘Hamiltonian cycle’ example?

- The ‘measure’ is ‘number of edges’.
- But, can this be related (as precisely) to Hamiltonicity?

More generally,

C, is some set of *combinatorial structures*,

Undirected Graphs

Boolean Formulae

Binary Strings

A *property* is a subset of such structures,

An undirected graph is Hamiltonian

A Boolean Formula is Satisfiable

A binary string has even parity

Suppose we define for each structure, S,

size - $\sigma(S)$; *weight* - $\tau(S)$

$\sigma(G) = |V|$; $\tau(G)$ number of edges;
$\sigma(F)$ number of variables; $\tau(F)$ 'length';
$\sigma(x)$ length of string; $\tau(x)$ number of 1s

Hamiltonicity and Graph Density

Consider choosing an n-vertex, m-edge graph, G, uniformly at random from the set, $G_{n,m}$, of all such graphs.

If $m(n)/0.5n \log_e n < 1$ then

$$\lim_{n \to \infty} P\,[\, G \in G_{m(n),m} \ \textit{is Hamiltonian} \,] = 0$$

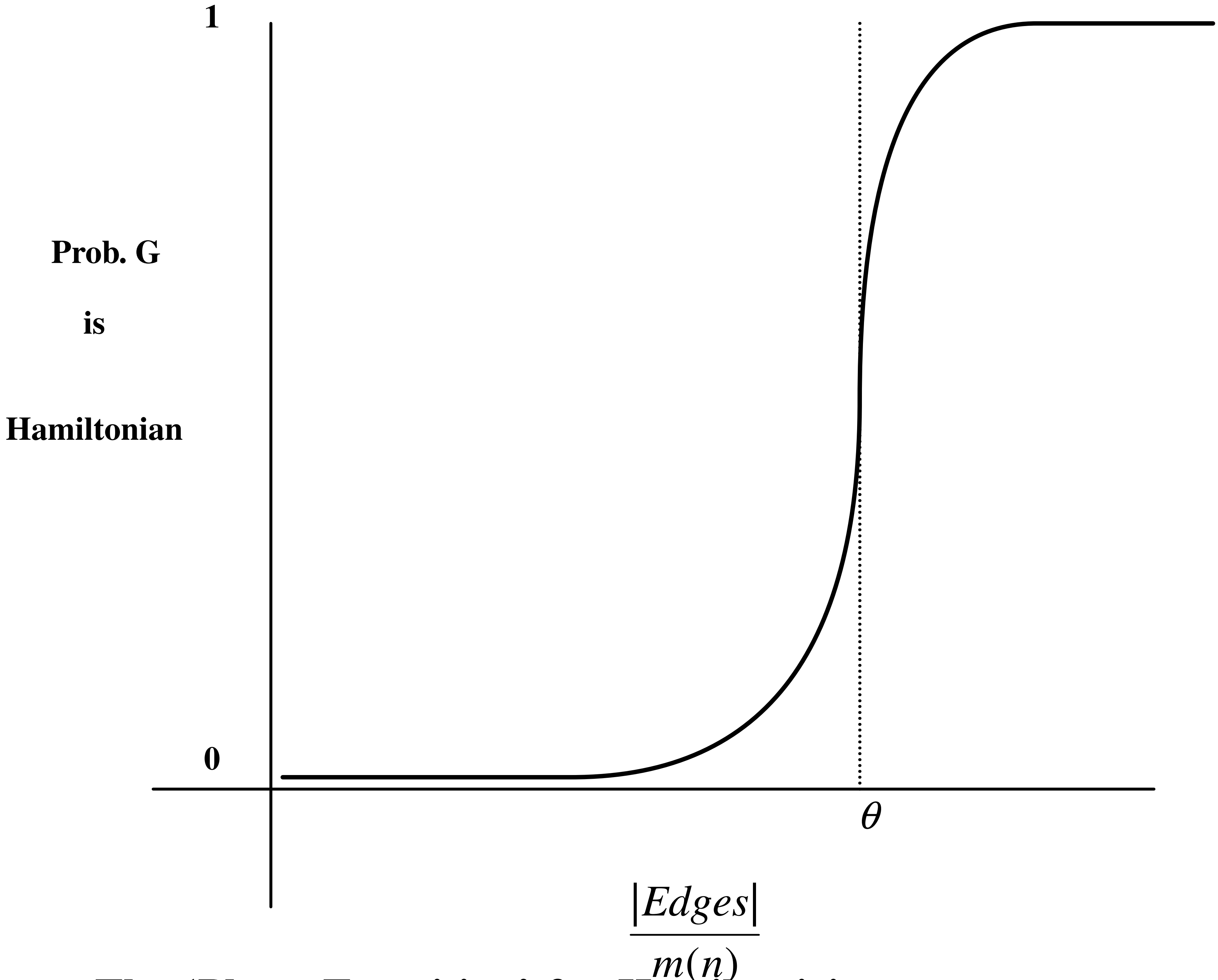
(Erdős and Rényi, 1956)

If $m(n)/0.5n \log_e n > 1$ then

$$\lim_{n \to \infty} P\,[\, G \in G_{m(n),m} \ \textit{is Hamiltonian} \,] = 1$$

(Koršunov, 1976)

Thus the likelihood of a random m-edge graph being Hamiltonian changes sharply from ‘close to’ 0 to ‘close to’ 1 in a ‘very small’ range of choices for m.



The ‘Phase-Transition’ for Hamiltonicity.

Similar phenomena have been observed for many graph properties:

3-Colourability, Connectivity, k-Clique, Planarity.

and for other combinatorial problems:

k-Satisfiability, prime implicate length.

So, for many ‘computationally interesting’ properties, there are close links between

The ‘size’ of a structure
Its ‘weight’
It having the property of interest

In addition, there ‘appear to be’ similar links with the run-time of search algorithms for determining if a given structure has a particular property.

Much of the ‘algorithmic’ study of phase-transitions suffers from a number of weaknesses:

- Little in the way of proven links between the algorithmic and combinatorial properties.
- Studies tend to be problem-specific, few attempts to build general theory.
- Terminology is ill-defined: many studies treat the notion of ‘phase-transition’ as intuitive.

Phase-Transitions in Probability

(f, σ, τ) is a decision problem over a set of *instances*, I; σ associates a *size* and τ a *weight* with each instance.

Suppose that $I_{n,m}$ is the set of instances of size n and weight m.

Let $\mu(n, m)$ denote

$$\frac{|\{ x \in I_{n,m} : f(x) = 1 \}|}{|I_{n,m}|}$$

(f, σ, τ) has a

phase-transition in probability

if there is a function $\theta : \mathbf{N} \rightarrow \mathbf{N}$ and $c \in \mathbf{R}$ such that $\forall \varepsilon > 0$

$$\lim_{n \rightarrow \infty} \mu(n, (c \pm \varepsilon)\theta(n)) = 0(1)$$

Fact 1: (Bollobas and Thomasen, 1987) If $f: \{0, 1\}^* \rightarrow \{0, 1\}$ is *monotone* then for $\sigma(x) = |x|$ and $\tau(x)$ the number of 1s in x, *exis*$\theta: \mathbf{N} \rightarrow \mathbf{R}^+$ such that

$$\lim_{|x| \rightarrow \infty} \mu(|x|, \theta^*(|x|)) = 0(1)$$

when $\theta^*(n) = o(\theta(n))(\omega(\theta(n)))$

Fact 2: (Friedgut and Kalai, 1996) Every symmetric monotone graph-property has a phase transition in probability with $\sigma(G) = |V(G)|$ and $\tau(G) = |E(G)|$.

a) The form of Fact 1, is called a *weak threshold effect*:

a) Both results are *existential* (do not help identifying appropriate θ, c)

b) Existence depends *solely* on a combinatorial property (monotonicity): no indication of any consequent complexity-theoretic or algorithmic property.

In fact, f need not even be decidable:

Theorem: Post's Correspondence Problem (over a binary alphabet), $PCP(0,1)$, may be encoded as a monotone property for which $\sqrt{n}/\log n$ is a weak threshold.

The proof of this result, however, easily yields:

Corollary: There is a decision procedure that (under the uniform distribution) 'almost surely' correctly decides random instances of $PCP(0,1)$ in linear time.

This decision procedure has no practical interest whatsoever.

[A much weaker result can also be shown for Matrix Mortality problems**]**

Methods for studying Phase-Transitions

The Experimental Approach

Given a decision problem, F on graphs, say. and a method, R, for solving the search variant, F_S, of F:

a. Randomly generate a number of graphs of different sizes, n, and numbers of edges, $m(n)$.

b. Evaluate $F(G_{n,m(n)})$ for each sample graph (monitoring the run-time of R on G).

c. Attempt to find correlations between differing values of n, $m(n)$, average search-time, number of positive results, etc.

Advantages

1. Little mathematical sophistication.
2. Conceptually simple.
3. May provide accurate thresholds if $m(n)$ is 'correct'.

Disadvantages

1. Computationally very intensive.
2. Reliable *uniform* random generation.
3. Element of guesswork involved.
4. Are results based **solely** on experiment 'reliable'? e.g. absence of 'very large scale' combinatorial effects.

A potentially much more serious weakness is

5. *LCG* methods for random bit generation, e.g. *drand*48, cannot produce random graphs (since these would have low Kolmogorov complexity in comparison with 'genuinely' random graphs),

This might invalidate experimental claims about 'average' run-time, e.g. via the results of Li and Vitanyi (1992).

(Surprisingly?) none of the standard experimental studies discusses random bit generation, despite using graphs with 1000+ vertices

Analytic approaches

Use techniques from combinatorial mathematics to prove properties (in the limit) of the defining probabilities.

The 'first-moment method': e.g.

estimate the average number of Hamiltonian cycles in $G_{n,m}$;

choose $\theta\, m(n)$ so that this average tends to 0.

$\theta\, m(n)$ is then an *upper bound* on the sharp threshold.

Lower bounds: analysis of specific algorithms to estimate when solutions are found 'almost always'.

Advantages

1. Results are *proved*.
2. Vast technical machinery available:
   a. Combinatorial Analysis.
   b. Probability Theory.
   c. Evolutionary Graph Theory.
   d. Percolation Theory.

Disadvantage

1. Example of proven *sharp* thresholds are rare.

k-Satisfiability

Intensive study of 3-*SAT* originated in the A.I. community with experimental work.

Motivated by the Davis-Putnam algorithm and relevance to KBS.

Experiments suggest $\theta \approx 4.24$, i.e.
A 'typical' 3-*CNF* on n variables with more than (less than) $4.24\,n$ clauses is 'almost surely' inconsistent (consistent).

By analytic means,
$3.003 < \theta < 4.579$
(Frieze and Suen, 1992; Zito, Robson, *et al.*, 1998)

A sharp threshold of $\theta = 1$ has been proved for 2-*SAT*.
(Goerdt, 1992; Chvatal and Reed, 1992)

This leads to fast average-case methods for a range of 3-*SAT* instances.

A variant, k-*HSAT*, of importance to Logic Programming is studied in Dunne and Bench-Capon (1997).
The *first sharp* thresholds holding $\forall\ k \geq 2$ are proved.

$$m(n, k) = n^{k-1} \log_e n$$
$$\theta(k) = (k-1)(k+1)/k!$$

Some Questions

1. Does every *NP*-complete (graph) problem have a phase-transition?
2. Does any ‘non-monotone’ problem have a phase-transition?
3. *PSPACE*-complete problems?
4. How does one establish that a problem *does not* have a phase-transition?

Some (Partial) Answers

1) If the weight function can only be defined in terms of $|V|$ and $|E|$ then there are NP-complete graph problems without phase-transitions in probability.

[BUT: the example used is not preserved by isomorphism**]**

2) (Friedgut and Kalai, 1996) show monotonicity is sufficient. It is not necessary.

Example: For constant $k \geq 2$, $G \in Q_k$ if: there is a subset, W of V, having size at least $|V| - k$ such that: the subgraph induced by W is Hamiltonian and $V - W$ is an independent set.

3) Sequential Truth Assignment. For any $f(n) \xrightarrow{n} \infty$, almost all instances, F, with $f(n)$ clauses have $STA(F) = 0$.

Future Work

1) Monotonicity is sufficient (but not necessary) for a weak threshold.

Formulate a definition of ‘nearly monotone’ property in such way that ‘near monotonicity’ is a necessary condition.

Possible Approach (in G_p Model)

For symmetric n-vertex graph properties, Π_n, $\mu_p(\Pi_n)$ is the probability that a random graph is in Π_n. Let $G_{1,2}$ be the set of graphs in which the edge $\{1,2\}$ occurs.

Russo's Formula gives (eventually),

$$\frac{d\mu_p(\Pi_n)}{dp} = \binom{n}{2}\{\mu_p(UP_n^{1,2}) - \mu_p(DN_n^{1,2})\}$$

where

$$UP_n^{1,2} = \{G : G \notin \Pi_n \cup G_{1,2} \ and \ G + \{1,2\} \in \Pi_n\}$$
$$DN_n^{1,2} = \{G : G \in \Pi_n - G_{1,2} \ and \ G + \{1,2\} \notin \Pi_n\}$$

Suppose 'nearly monotone' is defined as: for all 'sufficiently large' n, and for all $p \in (0,1)$, $\mu_p(UP_n^{1,2}) \ne \mu_p(DN_n^{1,2})$.

Conjecture: A symmetric graph property has a (weak) threshold if and only if it is nearly monotone.

[This would imply that monotonicity (in p) of $\mu_p(\Pi)$ is the 'correct' condition**]**

2) Prove the following (or provide evidence that it is not true):

> Let Π be an NP-complete graph property. Let Π_S be the search-variant of Π, i.e. if $G \in \Pi$, $\Pi_S(G)$ returns a witness to this effect.

Conjecture 1: If $(\Pi, |V|, |E|)$ has a phase-transition in probability then Π has an average-case polynomial time *decision* algorithm under the uniform distribution of instances.

Conjecture 2: $(\Pi, |V|, |E|)$ has a phase-transition in probability implies that Π_S is NP-easy.

If both are true, then these give some analytic support to the experimental claims concerning search speed-up. [But ....]