Threshold Phenomena and Influences

Gil Kalai and Shmuel Safra

1 Introduction: Threshold behavior: voting, random graphs, percolations and complexity

"Threshold phenomena" refer to situations in which the probability for an event to happen changes rapidly as some underlying parameter varies in some interval. Threshold phenomena play an important role in probability theory and statistics, physics and computer science, and are related to issues studied in economics and political science.

A simple yet illuminating example that demonstrates the sharp threshold phenomenon is Condorcet's Jury Theorem (CJT), which can be described as follows: Say one is running an elections process, where the results are determined based on simple majority, between two candidates Alice and Bob. If every voter votes for Alice with probability $p > \frac{1}{2}$ and for Bob with probability $1 - p$, and if the probabilities for each voter to vote either way is independent of the other votes, then as the number of voters tends to infinity the probability that Alice is elected tends to 1 (Condorcet (1875), see also

Young (1988)). The probability that Alice is elected is a monotone function of p, and when there are many voters it rapidly changes from being very close to 0 when $p < \frac{1}{2}$ to being very close to 1 when $p > \frac{1}{2}$.

The reason usually given for the interest of CJT to economics and political science is that it can be interpreted as saying that even if agents receive very poor (yet independent) signals, indicating which of two choices is correct, majority voting nevertheless results in the correct decision being taken with high probability, as long as there are enough agents, and the agents vote according to their signal. This is referred to in economics "asymptotically complete aggregation of information".

Condorcet's Jury theorem is a simple consequence of the weak law of large numbers. The central limit theorem implies that the "threshold interval" is of length proportional to $1/\sqrt{n}$. But some extensions are much more difficult. When we consider general economic or political situations, aggregation of agent's votes may be much more complicated than simple majority, the individual signal (or signals) may be more complicated than a single bit of information, and the distribution of signals among agents can be much more general and, in particular, agents' signals may depend on each other. On top of that, voters may vote strategically taking into account the entire situation in addition to their signal, and different voters may have different goals and interests, not only different information. Another complication is that the number of candidates may be larger than two, which results in a whole set of new complications.

Let us now mention briefly three other areas in which understanding threshold behavior emerges. The study of random graphs as a separate area of research was initiated in the seminal paper of Erdős and Renyi [?] from 19??. Considered a random graph $G(n, p)$ on n vertices where every edge among the $\binom{[n]}{2}$ possible edges appears with probability p. Erdős and Renyi proved a sharp threshold property for various graph properties: For example, if $p = \log n/n(1 + \epsilon)$ the graph is connected with probability tending to 1 (as n tends to infinity) while for $p = \log n/n(1 - \epsilon)$ the probability that the graph will be connected tends to zero. In recent years, general results concerning threshold behavior of random graphs were proved. What would be the appropriate value if we replace "connectivity" with another property of graphs"? What would be that value if we consider random subgraphs of an arbitrary graph to start with? It turns out that *symmetry* plays a crucial role in this study.

Next, we would like to mention complexity theory. Threshold phenomena play a role in various aspects of computational complexity theory, both conceptual and technical. One of the major developments in complexity theory in the last two decades is the emerging understanding of the complexity of approximation algorithms. Here is an important example: for a graph G let $m(G)$ be the maximum number of edges between two disjoint sets of vertices of G. MAX-CUT, the problem of dividing the vertices of a graph into two parts with maximum number of edges between them is known to be NP-hard. However, to find a partition such that the number of edges between the two parts is at most $m(G)/2$ is easy. The emerging picture is that if we wish to find a partition of the vertices with at least $cm(G)$ edges between the parts then there is a critical value of c (c_0 =???) such that the problem is easy (there is a randomized polynomial time algorithm) for $c < c_0$

and hard (NP-hard) for $c > c_0$. (In fact, in the range that the problem is easy it is very easy: a very simple algorithm works.) The study of threshold phenomena is also an important technical tool in understanding the hardness of approximation. Another connection with complexity theory occurs in the area of circuit complexity. It turns out that very "low" complexity classes necessarily exhibit coarse threshold behavior. ???????

Finally, threshold phenomena are, of course, of great importance in statistical physics. The notions of sharp threshold and phase transition came primarily from physics and many of the mathematical ideas for their study came from mathematical physics.

A basic model that we consider is that of a Boolean function, a function $f(x_1, x_2, , \dots, , x_n)$ where each variable x_i is a Boolean variable, namely it can accept the value '0' and '1' and the value of f is also '0' or '1'. A Boolean function is monotone if $f(y_1, y_2, , \dots, , y_n) \geq f(x_1, x_2, , \dots, , x_n)$ when $y_i \geq x_i$ for every i. Monotone Boolean functions describes "natural election rule" and we use this description to name some important classes of Boolean functions. The function $f(x_1, x_2, , \dots, , x_n) = x_i$ is called "dictatorship". "Juntas" are Boolean functions which depend on a bounded number of variables (independent of the number of variables).

Before we describe this papers' sections it is worth noting that sharp threshold is an asymptotic property and therefore it applies to a sequence of Boolean functions when the number of variables becomes large. Giving explicit, realistic and useful estimates is an important goal. (In the example above concerning elections between Alice and Bob the central limit theorem provides explicit, realistic and useful estimates, however, in more involved settings this may turn out to be difficult.)

The main massages of this paper can be summarized as follows:

- *The threshold behavior of a system is intimately related to combinatorial notions of "influence" and "pivotality". (Section* **??***)*

- *Sharp threshold is common. We can expect sharp threshold unless there are good reasons not to. (Section* **??***,* **??***)*

- *Higher symmetry leads (in a subtle way) to sharper threshold behavior (Section* **??***).*

- *Sharp threshold occurs unless the property can be described "locally". (Section* **??***).*

- *Systems required for their description a very low complexity classes have rather coarse (not sharp) threshold behavior. (Section* **??***).*

- *In various optimization problems when we seek approximate solutions there is a sharp transition between goals which are algorithmically easy and those which are computational intractable. (Section* **??***).*

- *Basic mathematical tools in understanding threshold behavior are high dimensional Fourier analysis and discrete isoperimetry. (Section* **??***).*

We list here a few main topice for further research:

- Explain the emergence of "power laws". (The threshold interval behaves like $n^{-\beta}$, $\beta > 0$ a real number. (Section **??**).

- Relate the threshold behavior with the threshold's location, find methods to exclude the possibility of an oscilating critical probabilities. (Sections **??**).
- Study other models, especially non-product probability distributions (Section **??**).
- Explore other applications (Section **??**).

In Section 2 we will introduce the notions of pivotality and influence and discuss "Russo's Lemma" which relates these notions to threshold behavior. In Section 3 we will describe basic results concerning threshold behavior of Boolean functions. In Section **??** we will discuss the connection to random graphs and hypergraphs and to the k-SAT problem. In Section 5 we will discuss the connections to computational complexity. Section 6 is devoted to the related phenomena of noise sensitivity. Section **??** discuss exensions and possible extensions to various other models. Section 9 describes the mathematical infra-structure behind the basic results that we described and especially the connection to Fourier analysis.

We are very thankful to Itai Benjamini, Ehud Friedgut, Nati Linial, Elchanan Mossel and Yuval Peres for inspiring discussions.

2 Pivotality, influence, power and the threshold interval

In this section we will describe the n-dimensional hypercube, and define the *notions of "pivotal" variables and influence for Boolean function. We state Russo's fundamental lemma that connect influences and thresholds.*

2.1 The discrete cube

Let $\Omega_n = \{0,1\}^n$ denote the discrete n-dimensional cube, namely the set of vectors of length n with 0-1 coordinates. A Boolean function is a map from Ω_n to $\{0,1\}$. Boolean functions on Ω_n are of course in 1-1 correspondence with subsets of Ω_n. Elements in Ω_n are themselves in 1-1 correspondence with subsets of $[n](=: \{1,2,,\ldots,,n\})$. Boolean functions under different names appears in many areas of science. We will equip Ω_n with a metric (namely a distance function) and a probability measure. For $x, y \in \Omega_n$ the Hamming distance $d(x,y)$ is defined by $d(x,y) = |\{i : x_i \neq y_i\}|$.

Denote by $\Omega_n(p)$ the discrete cube endowed with the product probability measure $\mathbf{P}_p$, where $\mathbf{P}_p\{x : x_j = 1\} = p$.

2.2 Pivotality and influence of variables

Consider a Boolean function $f(x_1, x_2,,\ldots,,x_n)$ and the associated event $A \subset \Omega_n(p)$, such that $f = \chi_A$, namely f is the characteristic function of A. .

For $x = (x_1, x_2,,\ldots,,x_n) \in \Omega_n$ we say that the kth variable is *pivotal* if flipping the value of x_k will change the value of f; formally, if

$$f(x_1,,\ldots,,x_{k-1},x_k,x_{k+1},,\ldots,,x_n) \neq f(x_1,,\ldots,,x_{k-1},1{-}x_k,x_{k+1},,\ldots,,x_n)).$$

The *influence* of the kth variable on a Boolean function f, denoted by $I_k^p(f)$ (and also by $I_k^p(A)$), is the probability that the kth variable is pivotal. The total influence $I^p(f)$ equals $\sum I^p_k(f)$. (We omit the superscript p for $p = \frac{1}{2}$.) For a monotone Boolean function thought of as an election method $I_k(f)$ ($= I_k^{\frac{1}{2}}(f)$) is refered to as the Banzhaf power index of voter k. The quantity $\phi_k(f) = \int_0^1 I^p_k(f)$ is called the Shapley-Shubik power index of f.

Mathematical study (under different names) of pivotal agents and influences is quite basic in percolation theory and statistical physics and also in probability theory and statistics, reliability theory, distributed computing, complexity theory, game theory, mechanism design and auction theory, other areas of theoretical economics, and political science.

2.3 Russo's lemma and threshold intervals

A Boolean function f is monotone if its value does not decrease when we flip the value of any variable from 0 to 1. For a monotone Boolean function $f \subset \Omega_n$ (i.e., χ_A is a monotone function), let $\mu_p(f)$ be the probability that $f(x_1, , \dots, , x_n) = 1$ with respect to the product measure μ_p. Note that $\mu_p(f)$ is a monotone function of p. Russo's fundamental lemma (see [**?**]) asserts that

$$\frac{d\mu(A)}{dp} = I^p(f).$$

Given a small real number $\epsilon > 0$, Let p_1 be the unique real number in $[0, 1]$ such that $\mu_{p_1}(f) = \epsilon$ and let p_2 be the unique real number such that $\mu_{p_2}(f) = 1-\epsilon$. The interval $[p_1, p_2]$ is called a *threshold interval* and its length $p_2 - p_1$ is denoted by $T_\epsilon(f)$. Denote by p_c the value so that $\mu_{p_c}(f) = \frac{1}{2}$, and call it the critical probability for the event A.

By Russo's lemma, large total influence around the critical probability implies a short threshold intervals.

3 Basic results concerning influences and threshold behavior of Boolean functions

The length threshold interval is at most proportional to the critical probability. Dictatorship and Juntas have coarse thrshold and when the critical probability is bounded away from zero and one coarse threshold implies that the function "looks" like a junta.

Some basic facts on influences and the corresponding results concerning threshold intervals are given by:

- The sum of influences cannot be overly small.

Theorem 3.1 (Loomis-Whitney, [?], Hart, [?]) *For a Boolean function f,*

$$\sum I_k(f) \geq \mu(f) \log_2(1/\mu(f)).$$

(In particular if $\mu_{\frac{1}{2}}(f) = \frac{1}{2}$ then $I(G) \geq 1$ and equality holds if and only if f is a "dictatorship" namely $f(x_1, , \ldots, , x_n) = x_i$ for some i or "antidictatorship" $f(x_1, , \ldots, , x_n) = 1 - x_i$ for some i). This inequality has its origins in the works of Whitney and Loomis, Harper, Bernstein, Hart and others and has great importance in many mathematical contexts.

An upper bound for the length of the threshold interval can be derived from the bounds on the sum of influences combined with Russo's lemma.

Theorem 3.2 (Bollobas and Thomason) *For every monotone Boolean function,*

$$T_\epsilon(f) = O(\min(p_c 1 - p_c)$$

The Theorem of Bollobas and Thomason is the basis for the following definition: We say that a sequence (f_n) of Boolean functions have a *sharp threshold* if for every $\epsilon > 0$,

$$T_\epsilon(f_n) = o(\min(p_c 1 - p_c).$$

Otherwise we say that the sequence demonstrate a coarse threshold behavior.

- Simple majority maximizes total influence.

By Russo's lemma it follows that:

Proposition 3.3 *Let f be a monotone Boolean function with n variables with $p_c(f) = \frac{1}{2}$, and let M be a simple majority function with n variables. Then, for every $p > \frac{1}{2}$, $\mu_p(M) \geq \mu_p(f)$. For odd n equality holds if and only if f is a simple majority function.*

- Not all influences can be small.

Theorem 3.4 (Kahn, Kalai &Linial, [?]) *If $mi_p(f) = \frac{1}{2}$ then*

$$\max_k I_k(f) \geq K\mu(A) \log n/n.$$

Note that this theorem implies that when all individual influences are the same (e.g., when A is invariant under the induced action from a transitive permutation group on $[n]$), then the total influence is larger than $K \log n$. An extension for arbitrary product probability spaces was found by Bourgain, Kahn, Kalai, Katznelson and Linial. Talagrand extended the result of Kahn, Kalai and Linial in various directions and applied these results for studying threshold behavior. Talagrand also presented a very useful extension for arbitrary real functions on the discrete cube. Talagrand's extension for the product maesure μ_p is stated as follows:

Theorem 3.5 (Talagrand, 1994)

$$\sum_{k=1}^{n} I_k^p(f)/(\log(I_k^p(f)) \geq Kp(1-p)\mu_p(f)(1-\mu_p(f)).$$

Theorem 3.6 (Friedgut) *For every real numbers $z > 0, A > 1$ and $\gamma > 0$, there is $C = C(\gamma, A, z)$ such that if $z \leq p \leq 1 - z$ the following assertion holds: For a monotone Boolean function f, if $I^p(f) \leq A$ then there exists a collection S of at most C variables and a monotone Boolean function g that depends only on the variables in S such that*

$$\mu_p\{x \in \Omega_n : f(x) \neq g(x)\} < \gamma. \tag{1}$$

Theorem 3.7 (Russo-Talagrand-Friedgut-Kalai) *For every $\epsilon > 0, \frac{1}{2} \geq t > 0$ and $\gamma > 0$ there are $\delta_i = \delta_i(\gamma, \epsilon, t) > 0$, $i = 1, 2, 3$ such that the following assertion holds:*

Let f be a Boolean function and suppose that $t \leq p_c(f) \leq 1 - t$. Then any of the following conditions implies that

$$T_\epsilon(f) < \gamma.$$

(1) For every value of $p, 0 < p < 1$ *and every* k, $1 \leq k \leq n$ $I_k^p(f) \leq \delta_1$.

(2) For some value of p *for which* $\epsilon < \mu_p(f) < 1 - \epsilon$ *(e.g. for* $p = p_c(f)$*),*

$$I_k^p(f) < \delta \text{ for every } k.$$

(3) The Shapley-Shubik power index of each variable in f *is at most* δ.

The first part of the theorem was proved by Russo [?]. A sharp version was proved by Talagrand [?] and Friedgut and Kalai [?] based on KKL's theorem and its extensions. Theorem 3.6 asserts that if the critical probability is bounded away from 0 and 1 and the threshold is coarse then for most values of p in the threshold probability, f can be approximate by a Junta with respect to the probability measure μ_p. Parts (2) and (3) are derived (based on Friedgut's result and some additional observations) in [?] but the values of δ_2, δ_3 are rather weak.

Theorem 3.8 *For a sequence* (f_n) *of monotone Boolean functions,* $\lim n \rightarrow inftyT_\epsilon(f_n) = 0$*, for every* $\epsilon > 0$ *if and only if the maximal Shapley-Shubik power index for* f_n *tends to zero.*

4 From Erdös and Renyi to Friedgut: random graphs and the k-SAT problem

4.1 Graph properties and Boolean functions

We first tell how to represent a graph property by a Boolean function.

Another origin for the study of threshold phenomena in mathematics is random graph theory and especially the seminal works by Erdös and Renyi.

Consider a graph $G =< V, E >$, where V is a set of vertices and E is the set of edges. Thus, $E \subset \binom{V}{2}$. Let $x_1, x_2, , \ldots, , x_e$ be Boolean variables which correspond to the edges of G. An assignement of the values 0 and 1 to the variables x_i corresponds to a subgraph H of G. This basic Boolean representation of subgraphs (or substructures for other structures) is very important. A graph property P is a property of graphs which does not depend on the labeling of the vertices. In other words P only depends on the isomorphism type of G. Examples include: "the graph is connected", "the graph is planar (can be drawn in the plane without crossing)" "the graph contains a triangle", "the graph contains a Hamiltonial cycle". Understanding the threshold behavior of random graphs was the main motivation behind the theorem of Bollobas and Thomason. Their result applies to arbitrary monotone Boolean functions so it does not use the symmetry that we have for graph properties.

Theorem 4.1 (Kalai and Friedgut) *For every property P of graphs,*

$$T_\epsilon(P) \leq C \log(1/\epsilon)/\log n.$$

This theorem is a simple consequence of KKL's theorem and its extensions combined with Russo's lemma. All influences of variables are equal for Boolean properties defined by graph properties. As a matter of fact this continues to be true for Boolean functions f that described random subgraphs of an arbitrary edge transitive graph. All influences being equal imply that the total influence I^p is at least as large as $K \min(/P_p(f), 1 - /P_p(f)) \log n$ and by Russo lemma this gives the required result.

In [?] Friedgut and Kalai raised several questions which were adressed in later works:

- What is the relation between the group of symmetries of a Boolean function and its threshold behavior?
- What would guarantee a sharp threshold when p tends to zero with n?
- What is the relation between influences, the threshold behavior and other isoperimetric properties of f?

The third questions was adressed by several papers of Talagrand [?] and also [?] and we will not elaborate on it here. We will describe in some details the work of Bourgain and Kalai on the first question and the works of Friedgut

4.2 Threshold under symmetry

Here we describe a measure of symmetry which is related to the threshold behavior. The more symmetry the sharper the threshold. The measure of symmetry is based on the the size of orbits.

Bourgain and Kalai studied the effect of symmetry on the threshold intervals and their work have led to the following result:

Theorem 4.2 (Bourgain and Kalai) *For every property P of graphs with m vertices,*

$$T_\epsilon(P) \leq C(\tau) \cdot \log(1/\epsilon)/(\log m)^{2-\tau}.$$

Let Γ be a group of permutations of $[n]$. Γ acts on Ω_n by $g((x_1, x_2, , \ldots, , x_n) = (x_{\pi(1)}, x_{\pi(2)}, , \ldots, , x_{\pi(n)})$, for $g \in \Gamma$.. A boolean function is Γ invariant if $f(g(x)) = f(x)$ for every $x \in \Omega_n$. For boolean function f describing A graph property for graphs with m vertices is described by a Boolean function on $n = \binom{m}{2}$ variables. Such Boolean functions are invariant under the induced action from the symmetric group S_m on the vertices (the group of all permutations of the vertices) acting on the all edges. (Recall that the variables of f correspond to the edges of the complete graph on m vertices.)

The discrete hypercume Ω_n is divided to layers: Write Ω_n^r for the vectors in Ω_n with r ones altogether. For a group Γ of permutations of $[n]$ let let $T(r)$ denotes the number of orbits in the induced action on Ω_n^r and let $B(r)$ be the smallest size of an orbit of Γ acting on $Omega_n^r$. In our case, for graph properties $T(r)$ is the number of isomorphism types of graphs with m vertices and r edges and $B(r)$ is the minimum number of (labelled) graphs with m vertices and r edges that are isomorphic to a specific graph H. Recall that the number of graphs isomorphic to H is $m!/|Aut(H)|$, where $Aut(H)$ denotes the automorphism graph of H.

The parameter $\kappa(\Gamma)$ that "measure" how large is the group of symmetries is defined as follows:

$$\kappa(\Gamma) = \min\{r : B(r) < 2^r\},$$

When we consider graph properties for graphs with m vertices $B(r)$ behaves like $\binom{m}{\sqrt{r}}$. To see this note that when $r = \binom{s}{2}$ graphs H with a fewest isomorphic copies (hence largest automorphism groups) are complete graphs.

Define also:

$$\kappa_\tau(\Gamma) = \min\{r : B(r) < 2^{r^\tau}\},$$

Bourgain and Kalai showed that for every $\tau > 0$ the total influence $I^p(f)$ of a Γ-invariant Boolean function is always at least

$$K(\tau) kappa_\tau(\Gamma)\mathbf{P}_p(f)$$

where $K(\tau)$ is a positive function of τ. This gives Theorem **??** when we specialize to graph properties. They gave examples of a *Gamma*-invariant function f_n such that $\mathbf{P}(f_n)$ is bounded away from 0 and 1 and $I(f_n) = \Theta(\kappa(f_n))$.

4.3 Threshold behavior for small critical probabilities

Here we describe theorems by Friedgut and by Bourgain which show that when the critical probabilities are small coarse threshold imply that the function has "local" behavior

In this section we state theorems by Friedgut and by Bourgain that study sharp threshold when the critical probability p_c is small:

Theorem 4.3 (Friedgut) *For every $\epsilon > 0, t > 0 and A > 0$, there is $C = C(t, A, \epsilon)$ such that if f represebt a graph property, $t \le \mu_p(f) \le 1 - t$ and $I^p(f) < A$ then there a family of graphs $\mathcal{G}$ such that $e(G) \le A$ for every $A \in \mathcal{G}$ such that*

$$\mu_p x : \|f(x) \neq g_{\mathcal{G}}(x)\} \le \epsilon$$

.

Friedgut's proof relies on symmetry and the statement extends to hypergraphs and similar structures. The crucial property seems to be that the number of orbits of sets of size t is bounded uniformly for all n. (For graphs this reads: there is a bounded number of isomorphism types of graphs with n vertices and t edges when t is fixed.)

Friedgut conjectured that the theorem applies for arbitrary Boolean functions. A theorem of Bourgain comes close to this conjecture:

Both Friedgut and Bourgain theorems are very useful to prove sharp threshold behavior in many cases. We refer the reader to Friedgut's recent survey article [**?**],

4.4 Margulis' theorem

Here we describe a theorem of Margulis with another general method to prove a sharp threshold behavior

Margulis [**?**] found in 1974 a remarkable condition that guarantees a sharp threshold for Boolean function and applied it for random subgraphs of highly connected graphs. (This paper contains also an earlier proof of Russo's lemma.) An improvement of his theorem was found by Talagrand.

Theorem 4.4 (Margulis)

Theorem 4.5 (Talagrand)

Let $f = \chi_A$ be a Boolean function and suppose that for every $x \in A$, if $h(x) > 0$ then $h(x) \geq k$. (Alternatively we can make the same assumption for the complement of A.) It follows that $I(A) \geq k \cdot \partial_v(A)$ and from relation **??** we conclude that $I(A) \geq C\sqrt{k}$. By Margulis-Russo's lemma the length of the threshold interval is at most $1/\sqrt{k}$.

Here is Margulis' original application: Let G be a k-connected graph and consider a randon spanning subgraph H where an edge of G is taken to be an edge of H with probability p. We assume that H has n edges and let f be the Boolean function that represent the property: "H is connected." Margulis proved that the threshold interval for f is of length $O(1/\sqrt{k})$. The reason is that if H is a spanning subgraph of G, H is not connected and it is possible to make H connected by adding a single edge of G, then H must have precisely two connected components and since G is k-connected, there are at least k edges in $G \backslash H$ such that adding them to H yield a connected graph.

5 Threshold behavior and complexity

In this Section we will discuss two areas where threshold phenomena and complexity theory are related. First we will described results concerning bounded depth circuits which is a very basic notion in computational complexity. Second we will describe the connection to the area of "hardness of approximation".

5.1 Bounded depth Boolean circuit

Boolean functions wich belongs to AC0- a very low complexity class (and very exciting nevertheless)-must have a pretty coarse threshold behavior

Theorem 5.1 (Linial, Mansour, Nisan; Hastad)

Conjecture 1 (Reverse Hastad)

Remark: A large number of papers appeared in recent years that suggest a bold and far-reaching statistical physics approach to fundamental questions in complexity. These papers suggest a new approach in which one regards classical optimization problems as zero-temperature cases of statistical physics systems. This approach further proposes that the complexity of problems is related to the type of phase transition of the physical system. In addition statistical physics suggests both a way of thinking and heuristic mathematical machineries to deal with these problems. This approach was accepted with a great deal of skepticism by the complexity theory community. And indeed hard evidence for the usefulness of this approach is still missing. There is a result by Hastad, Linial-Mansour-Nisan and Boppana which can be interpreted as going in the direction suggested by physicists.

5.2 Hardness of approximation and PCP

Can we approximate?? Sometimes approximation is intractible sometimes it is easy. There is often a sharp transition between the two behaviors.

Sharp threshold phenomena are technically important to prove that some approximation problems are difficult.

Threshold behavior has recently turned out to play a fundamental role in complexity theory and especially in PCP . Proving tight hardness of approximation results for approximation problems, many times requires analysis of a code-word of the long-code, which can in fact be thought of as simply a Boolean function (where the legal code-words are dictatorships). Such an analysis appears first in the work of Hastad [**?**, **?**], where the linearity-test and long-code test are analyzed. This type of analysis still proved inadequate for the showing hardness of approximation for the Vertex-Cover problem – where the bound of 7/6 of Hastad remained the best know for several years. A recent paper of Dinur and Safra utilized influences of variables on Boolean function, and in particular Friedgut's Theorem, to show an improved hardness result for this problem, namely, showing it is NP-hard to approximate to within a factor larger than 4/3. This is still far from the best known approximation algorithm, which yields an approximation ratio slightly smaller than 2.

Conceptually, it turns out that when we try to approximate the solution for an optimization problem, there are various problems for which there is a sharp threshold between cases that are very easy to solve and cases in which the problem is NP-hard. This insight and the methodology for observing such phenomena are fairly recent, and a deeper understanding of the issues involved may lead both to improved approximation algorithms as well as tighter hardness results. Harmonic analysis on Boolean functions, already proved to be a powerful tool for such considerations.

Here are some results concerning sharp transition between easy and hard computational problems:

- 1. MAX-3-LIN(2): given a set of linear equations over Z_2, satisfy as much of them as possible by assigning the variables – satisfying half of the equations is easy, by just taking a random assignment (this "algorithm" can be derandomized easily). However, even distinuishing between $\frac{1}{2}+\epsilon$ and $1-\epsilon$ satisfiable instances is NP-hard.

- 2. MAX-3-SAT: similar problem - only instaed of equations you have ORs over three literals each. 7/8 fraction of the constraints are expectedly satisfied by a random assignment, yet distinuising between $7/8+\epsilon$ and 1 is NP-hard.

- 3. SET-COVER: given a collection of subsets of $[n]$, find the least amount of sets from the collection whose union is [n]. A $\ln n$ approximation (namely using at most $\ln n$ times more sets than actually needed) is simple to obtain, and nothing better can be achieved unless NP is in sub-exponential time.

Khot [**?**] made a remarkable conjecture concerning a very strong form of PCP. Khot's conjecture will imply a sharp threshold behavior for the followin problems.

- MIN-2-SAT-DELETION: the instance is a set of ORs over 2 literals each. The goal however is to delete as little of the ORs as possible, such as the remaining instance is completely satisfiable. Khot's conjecture shows that no constant factor approximation is possible if his conjecture is true.

- Vertex Cover: Given an undirected graph, find the minimal number of nodes that touch all edges. Covering the edges by at most twice the number of nodes needed (namely a 2 approximation) is quite easy - for example by taking BOTH ends of each yet-uncovered edge. Khot's conjecture implies that 2 is tight.

Another important algorithmic problem is MAX-CUT: find a 2-partition of the nodes of a given graph such that as many edges as possible are two colored. We will discuss this problem in the next section.

6 Noise sensitivity, threshold circuits and PCP again

Is there a good voting method as far as imunity to random noise in counting the votes?

Motivated by mathematical physics, Benjamini, Kalai and Schramm (1999) studied (in a different language) the sensitivity of an election's outcome to low levels of noise in the signals (or, if you wish, to small errors in the counting of votes). Their assumption is that there is a probability $\epsilon > 0$ for a mistake in counting a vote and these probabilities are independent. Simple majority tends to be quite stable in the presence of noise. Two-level majority like the US electoral system is less stable and multi-tier council democracy is quite sensitive to noise. A basic result concerning noise sensitivity is:

- Simple and weighted majorities are noise-stable.

The main result of [?] asserts that a sequence of Boolean function (f_n) which is not noise sensitive must have bounded-away-from-zero correlation with sime weighted majority function.

7 Percolation and other models from statistical physics

We already mentioned in the introduction that the primary area where threshold behaior was studied is Physics.

Consider the graph G of an n by $n+1$ planar rectangular grid. Thus, the vertices of G are points of the form $(i, j) : 1 \leq i \leq n, 1 \leq j \leq n+1$, and two vertices are adjacent in the graph G if they agree in one coordinate and differ by one in the other coordinate. Questions concerning percolation in the plane (ususally on the infinite grid) are very important. Russo's lemma was proved in the context of percolation and Kesten proved a sharp threshold result on the route for proving his famous result concerning critical probabilities for planar percolation. Choose every edge to be open with probability p what is the probability of an open path from the left side of the rectangle to the right side? Is there a sharp threshold? We can ask and immediately answer the analogous question "on the torus" when we identify the left and right sides of the rectangle and the top and bottom side. (or even just for a cylinder when we identify only the left and right sides.) When we look for a path homotopic to the horizontal path from (0,0) to $(0, n+1)$ a sharp threshold follows from Theorem **??** above.

The total influence of the Boolean function f described by "left-right" percolation on the $m+1$ by m grid is a basic notion in percolation theory. It is conjectured (but recently proved for some variants based on the works of Smirnov, Lawler, Schramm and Werner) that $I(f_n) \approx m^{\gamma}$ ($\approx n^{3/8}$, where n is the number of variables), where $\gamma = 3/4$.

Problem 2 (Basic Problem I) *Find sufficient conditions to guarantee that*

(1) for some $\alpha > 0$ $I(f) > n^{\alpha}$*.*

(2) for some $\beta > 0$ $I(f) < n^{\frac{1}{2}-\beta}$*.*

What is the reason that the total influence for percolation behaves like a power of n? We can expect that conceptually the reason lies in some symmetry like the one considered in the Theorem of Bourgain and Kalai. However, there are two facts we should note. The first is that the present formulations of Theorem **??** are not sufficiently strong to yield lower bounds of the form $I(f_n) > n^{\alpha}$. The second is that the Boolean function we described does not admit many symmetries. What it does seem to have is "approximate" symmetries. We expect that as the grid becomes finer there is some "limit object" (The scaling limit) and this reflects some approximate symmetry of our functions under continious maps of the square to itself. (Some approximate symmetry w.r.t continuos maps is expected in any dimension. In dimension two we expect that the limit object is symmetric w.r.t. conformal maps. This was proved by Smirnov (for a different variant of planar percolation).

Remark: Let f be a Boolean function. Consider a real function g defined on the discrete cube. Let $y_1, y_2, , \dots, , y_n$ be identical independent random variables. Define

$$g(x_1, x_2, , \dots, , x_n) = \min\{\sum x_1 y_2 + x_2 y_2 + \cdots + x_n y_n : f(x_1, x_2, , \dots, , x_n) = 1\}. \quad (2)$$

Understanding the behavior of the function g is of interest in percolation theory. In this context f is the Boolean function which describes the existence of a path of open edges between two points on the grid. Quite curiously the same model is related to questions raised in mechanism design in economics theory.

7.1 Influence and threshold behavior without probability independence; the Ising and Potts models

One of the major research challenges is to extend results described in this paper to models where the probability distribution is not a product distribution.

8 Ecomonics, social sciences

8.1 Self organizing criticality: Feddersen and Pesendorfer's work on strategic voting

Why should we care about critical probabilities anyway?

Let us return now to the Condorcet Jury theorem from the Introduction. A key assumption in Condorcet Jury Theorem is that each agent votes according to his signal. There is recent interesting literature on the case that voters vote *strategically* based on their signal. Suppose that every voter wishes to minimize the probability for mistakes. (We can give different weights to mistakes in the two directions.) Feddersen and Pesendorfer considered juries and naturally gave much larger weight to an innocent person being convicted than to a guilty one being acquitted. If, in order to convict you need 2/3 of the votes, and jurors vote according to their signals, then, when $p = 0.51$ and the number of jurors is large, they will hardly ever convict. However, Feddersen and Pesendorfer showed that when the agents vote strategically (and used mixed (randomizes) strategies) the probability of a mistakes tends to zero as the number of jurors grows, even if the signal is weak. They also showed that this is not the case when all jurors are required in order to convict. Feddersen and Pesendorfer result and analysis is based on the notion of Nash equilibrium. Nash equilibria in theis case gives us a nice example of "self-organizing criticality". The behavior at the critical point is significant even when the voting method to start with is biased. When we consider general voting methods it can be shown that "asymptotically complete aggregaion of information" is intimately related to the sharp threshold. In particular, if there is a sharp threshold then there always is a Nash-equilibrium point for which the probability of mistakes tends to zero as the number of jurors grows. Indeed, it is easy to see that the existence of a sharp threshold is equivalent to the existence of a symmetric strategy (every voter has the same strategy) under which the probability for a mistake tends to zero. Since this is a common goal game the vector of individual strategies which minimize the probability for a mistake is a Nash equilibrium.

8.2 Social indetermancy and social chaos

What happend when we replace the discrete cube by other product spaces? What can we say about elections when there are more than two candidates?

9 The Underlying Mathematics

The story we tell in this paper has another important aspect: The mathematicsl toold required for proving the main theorems.

9.1 A proof of Russo's lemma

9.2 Harmonic analysis of Boolean functions

9.3 Discrete Isoperimetric inequalities; concentration of measure

10 Concluding remark

Threshold phenomena and related concepts such as pivotality, influence and noise sensitivity are important in many areas of science and engeneerings. We described some mathematical advances in understanding of threshold behavior and related phenomena and also various applications and connections. One of the interesting aspect of our experience was to see that while the mathematical underlying concepts are very similar in different disciplines the methodology, interpretations and overall approaches can be sharply different.