Interpolation for MaxCut and MinBisection on Sparse Random Graphs

David Wu

October 7, 2024

Abstract

We study the Maximum Cut and Minimum Bisection problems on random d-regular graphs. We will show how to connect these classical graph problems to a family of associated Gibbs measures, and then sketch out how to use the so-called interpolation method to characterize the precise behavior of these problems. In particular, this approach is the one carried out by [DMS17], who show that with high probability, $\frac{1}{n}\mathsf{MaxCut}(G) = \frac{d}{4} + P_* \sqrt{\frac{d}{4}} + o(\sqrt{d})$, where $P_* \approx 0.763$ is the Parisi constant from statistical physics.

1 MaxCut and MinBisection

Let $G = (V, E)$ be a graph with n vertices and m edges. A cut is a partition (S, S^c) of the vertices, and the size of the cut is $E(S, S^c)$, the number of edges that cross the cut from S to S^c.

A fundamental question in theoretical computer science is to determine the maximum cut of a graph, which we call MaxCut. Similarly, MinBis is the problem of determinine the minimum cut subject to the partition being balanced: $|S| = \lfloor n/2 \rfloor$.

Let us formulate these problems more precisely in a form that will be suggestive for our overall approach. Some useful notation: $\Omega_n = \left\{x \in \{\pm 1\}^n \,|\, \sum x_i = 0\right\}$ (we'll assume that n is even throughout):

$$\max_{x \in \{\pm 1\}^n} \frac{1}{4} \sum_{(i,j) \in E} (x_i - x_j)^2 = \frac{|E|}{2} - \frac{1}{2} \min_{x \in \{\pm 1\}^n} \sum_{(i,j) \in E} x_i x_j \qquad \text{(MaxCut)}$$

$$\min_{x \in \Omega_n} \frac{1}{4} \sum_{(i,j) \in E} (x_i - x_j)^2 = \frac{|E|}{2} - \frac{1}{2} \max_{x \in \Omega_n} \sum_{(i,j) \in E} x_i x_j \qquad \text{(MinBis)}$$

Famously, MaxCut and MinBis are both NP-hard for worst-case graphs. If we allow ourselves to consider random instances, the story becomes more interesting.

Remark 1.1. For worst-case instances, the problem of *approximating* the MaxCut is an extremely interesting question, which we won't have time to cover. The famous Goemans-Williamson SDP relaxation gives a 0.878-approximation to MaxCut, with better approximation ratios known for special classes of graphs. Amazingly, this is essentially tight for worst-case instances — it is *unique-games* hard to approximate MaxCut better than the SDP.

For concreteness, we will consider random d-regular graphs on n vertices, where d is a constant. One can transfer over these results to the Erdős-Rényi random graph model $G(n, d/n)$, but directly analyzing $G(n, d/n)$ in this sparse regime is more annoying.

A lot of effort has gone into characterizing MaxCut and MinBis on sparse random graph models. For isntance, consider the random edge model $G(n, \lfloor dn \rfloor)$. With some knowledge of random graphs and the trusty moment method, one can establish that there's a phase transition for MaxCut at $d_* = \frac{1}{2}$ and MinBis at

$d_* = \log 2$. In particular, for both problems, below the critical degree parameter, the (excess) cut size is $O(1)$, whereas above, the cut size is $\Omega(n)$.

Even restricted to random graph settings, this problem has been studied for decades (see, e.g., [Bol88, FO05, CO07, BGT10, GL18], just to name a few). A variety of techniques have been used to bound the cut sizes, such as spectral relaxations / SDPs — this may seem natural given the quadratic programming formulation in (MaxCut) and (MinBis). As a brief demonstration of this approach, one can relax min bisection to a spectral relaxation with the graph Laplacian. From this, we can observe that on random d-regular graphs, combined with Friedman's theorem, that with high probability $\frac{1}{n}\mathsf{MinBis}(G) \geqslant \frac{d}{4} - \sqrt{d-1}$. The upshot of this line of research is that $\mathsf{MinBis}(G) = \frac{nd}{4} - \Theta(\sqrt{d})n$ and $\mathsf{MaxCut}(G) = \frac{nd}{4} + \Theta(\sqrt{d})n$, where Θ is asymptotic in d. Since a d-regular graph has $\frac{dn}{2}$ edges, comparing against (MaxCut) and (MinBis) means that these results are equivalent to showing that

$$\min_{x \in \{\pm 1\}^n} \sum_{(i,j) \in E} x_i x_j = -\Theta(\sqrt{d})n$$

$$\max_{x \in \Omega_n} \sum_{(i,j) \in E} x_i x_j = \Theta(\sqrt{d})n,$$

respectively. The main goal of this talk is to pin down the leading order behavior for this $\Theta(\sqrt{d})$ term; in other words, we will explicitly(!) pin down the constant in front of $\sqrt{d}$.

2 Ising models

To get to the correct behavior, we will study a Gibbs measure! This might be surprising if you haven't seen these types of arguments before. Physicists were able to give predictions for the behavior of MaxCut and MinBis in these settings, but rigorous proofs were lacking. Let us first recall the definition of Gibbs measures on the hypercube.

Definition 2.1 (Gibbs measure and free energy)**.** A Gibbs measure on $\{\pm 1\}^n$ is specified by a Hamiltonian $H : \{\pm 1\}^n \to \mathbb{R}$ and an inverse temperature parameter $\beta \in \mathbb{R}$ with $\mu_{\beta,H}(x) \propto \exp(\beta H(x))$. When β, H are clear from context, we will omit these subscripts. The partition function is defined by

$$Z_H(\beta) \triangleq \sum_{x \in \{\pm 1\}^n} \exp(\beta H(x)),$$

and the normalized free energy is given by

$$\phi_{n,H}(\beta) = \frac{1}{n} \log Z_H(\beta).$$

If H is deterministic, then we will often drop this subscript. If H is random, then we instead define the normalized free energy by $\phi_n(\beta) = \frac{1}{n}\mathbf{E} \log Z(\beta)$, where the expectation is taken over the disorder (randomness) of H.

What is the connection between these Gibbs measures and MaxCut and MinBis? We actually have the following generic fact about the free energy of a Gibbs measure:

Fact 2.2. *Fix a Hamiltonian $H : \{\pm 1\}^n \to \mathbb{R}$. Let $e_{\max} = \frac{1}{n}\max_{x \in \{\pm 1\}^n} H(x)$ and $e_{\min} = \frac{1}{n}\min_{x \in \{\pm 1\}^n} H(x)$. For all $\beta > 0$, we have*

$$e_{\max} \leqslant \frac{1}{\beta}\phi_n(\beta) \leqslant e_{\max} + \frac{\log 2}{\beta},$$

and for $\beta < 0$, we have

$$-e_{\min} \geqslant \frac{1}{\beta}\phi_n(\beta) \geqslant -e_{\min} + \frac{\log 2}{|\beta|}.$$

Proof. By definition,

$$\phi_n(\beta) = \frac{1}{n}\log Z(\beta) = \frac{1}{n}\log \sum_{x\in\{\pm 1\}^n} \exp(\beta H(x)).$$

Thus, we can crudely estimate

$$\begin{aligned}\beta e_{\max} \leqslant \phi_n(\beta) &\leqslant \frac{1}{n}\log\left(2^n \cdot \exp(n\beta e_{\max})\right) \\ &\leqslant \beta e_{\max} + \log 2.\end{aligned}$$

The proof for the $\beta \to -\infty$ case is the same by replacing H with $-H$. □

Remark 2.3. The above lemma clearly holds if H is random, by redefining the quantities in terms of their quenched averages (taking $\mathbf{E}$ over the randomness in H everywhere).

In view of the above fact, together with (MaxCut), (MinBis), we will study the free energy of a certain Gibbs measure associated to the sparse graph called a *diluted* Ising model.

Definition 2.4 (Ising model)**.** Let $\beta \in \mathbb{R}$, and $G = (V, E)$ be a graph on n vertices, and let A be its adjacency matrix. The diluted Ising model on a sparse graph G is the Gibbs measure with Hamiltonian $H^{\mathsf{D}}(x) = \sum_{(i,j)\in E} x_i x_j = x^\top A x$. In other words,

$$\mu_\beta^{\mathsf{D}}(x) \propto \exp\left(\beta \sum_{i\sim j} x_i x_j\right) = \exp\left(\beta x^\top A x\right).$$

We denote the free energy by $\phi_n^{\mathsf{D}}(\beta)$, and if G is a sparse random graph then we take expectation over the randomness in G, as usual.

Notice that we allow β to be negative, which might look strange, but this gives us flexibility in studying both MaxCut and MinBis on equal footing.

We will study the free energy of the Ising model on G for different values of β. Clearly, $\mathbf{E}\,\mathsf{MaxCut}(G) = e^{\mathsf{D}}_{\min} = \lim_{\beta\to-\infty} \frac{1}{|\beta|}\phi_n^{\mathsf{D}}(\beta)$, and $\mathbf{E}\,\mathsf{MinBis}(G)$ has an analogous formula, except now the configuration space is restricted to Ω_n instead of $\{\pm 1\}^n$.

Hence, we have reduced the problem of studying MinBis to studying the large β behavior of the free energy of a certain Ising model. How will we get a handle on this free energy? The key idea is to relate it to the free energy of a model that we do understand. This will look a little strange at first, but it turns out to be a very powerful idea which is fruitful in other settings at well.

To be concrete, we will relate the free energy of our Ising model to the free energy of the *Sherrington-Kirkpatrick* model, which is the canonical mean-field spin glass.

Definition 2.5 (SK Model)**.** Let $W \sim \mathsf{GOE}(n)$ be a GOE matrix. In particular, W is a symmetric matrix with $W_{ij} \sim N(0, \frac{1}{n})$ for $i < j$ and $W_{ii} \sim N(0, \frac{2}{n})$. The SK model is the Gibbs measure with $H(x) = x^\top W x$, so for $\beta \geqslant 0$ we have

$$\mu_\beta^{\mathsf{SK}}(x) \propto \exp\left(\beta x^\top W x\right).$$

The partition function is given by $Z^{\mathsf{SK}}(\beta) = \sum_{x\in\{\pm 1\}^n} \exp(\beta x^\top W x)$, the free energy is $\phi_n^{\mathsf{SK}}(\beta) = \frac{1}{n}\mathbf{E}\log Z^{\mathsf{SK}}(\beta)$, and the (expected) ground state energies are $e^{\mathsf{SK}}_{\max} = \frac{1}{n}\mathbf{E}\max_{x\in\{\pm 1\}^n} H(x)$ and $e^{\mathsf{SK}}_{\min} = \frac{1}{n}\mathbf{E}\min_{x\in\{\pm 1\}^n} H(x)$.

As alluded to, the free energy of the SK model is quite well understood. In particular, its behavior is governed by the celebrated Parisi formula [Tal06], which we will *not* attempt to prove here.

Theorem 2.6 (Parisi's formula)**.** *For the SK model, we have*

$$-\lim_{n\to\infty} e^{\mathsf{SK}}_{\min} = \lim_{n\to\infty} e^{\mathsf{SK}}_{\max} = \mathsf{P}_*,$$

where $\mathsf{P}_* \approx 0.763$ *is the Parisi constant.*

The main theorem that we will sketch the proof of is the following:

Theorem 2.7 ([DMS17])**.** *Let G be a random d-regular graph. Then with probability* $1 - o_n(1)$*, we have*

$$\frac{1}{n}\mathsf{MaxCut}(G) = \frac{d}{4} + \mathsf{P}_*\sqrt{\frac{d}{4}} + o_d(\sqrt{d})$$
$$\frac{1}{n}\mathsf{MinBis}(G) = \frac{d}{4} - \mathsf{P}_*\sqrt{\frac{d}{4}} + o_d(\sqrt{d}).$$

First, by a standard second moment argument, we will argue that it suffices to instead study the quenched quantities $\mathbf{E}\,\frac{1}{n}\mathsf{MaxCut}(G) = e^{\mathsf{D}}_{\min}$ and $\mathbf{E}\,\frac{1}{n}\mathsf{MinBis}(G) = e^{\mathsf{D}}_{\max}$. Roughly, the argument goes as follows. By constructing a coupling of the graph distributions, it suffices to study the problem when the graph model is an $G(n, 2dn)$, or Poissonized version. Then, one can use the fact that the number of edges concentrates tightly around $2dn$ and Efron-Stein to conclude that the variance of the MaxCut is $O(n)$, in comparison to the expectation, which is $\Omega(n)$. So we have reduced the problem to studying the expected max cut or min bisection on your favorite random graph model of fixed average degree d.

3 Interpolation

The main idea of the proof is to interpolate between the free energies of the Ising model and the SK model. Using it, we will prove the following crucial lemma:

Lemma 3.1 ([DMS17, Proposition 2.2])**.** *There exists universal constants* $C_1, C_2 > 0$ *such that*

$$\left|\phi^{\mathsf{D}}_n\left(\tfrac{1}{\sqrt{d}}\beta\right) - \phi^{\mathsf{SK}}_n(\beta)\right| \leqslant C_1\frac{|\beta|^3}{\sqrt{d}} + C_2\frac{\beta^4}{d}.$$

In particular, setting $|\beta| = d^{1/6}$*, we get*

$$\left|\frac{1}{\beta}\phi^{\mathsf{D}}_n\left(\tfrac{1}{\sqrt{d}}\beta\right) - \frac{1}{\beta}\phi^{\mathsf{SK}}_n(\beta)\right| \leqslant O_d(d^{-1/6}).$$

Once we establish the above lemma, it is not hard to see how to obtain the desired result.

Proof sketch for Theorem 2.7. By triangle inequality, we have

$$\left|\frac{1}{\sqrt{d}}e^{\mathsf{D}}_{\max} - e^{\mathsf{SK}}_{\max}\right| \leqslant \left|\frac{1}{\sqrt{d}}e^{\mathsf{D}}_{\max} - \frac{1}{\beta}\phi^{\mathsf{D}}_n(\tfrac{1}{\sqrt{d}}\beta)\right| + \left|\frac{1}{\beta}\phi^{\mathsf{D}}_n(\tfrac{1}{\sqrt{d}}\beta) - \frac{1}{\beta}\phi^{\mathsf{SK}}_n(\beta)\right| + \left|\frac{1}{\beta}\phi^{\mathsf{SK}}_n(\beta) - e^{\mathsf{SK}}_{\max}\right|.$$

By Fact 2.2, we can upper bound the first and third terms by $\frac{\log 2}{\beta}$, and setting $\beta = d^{1/6}$ we can apply Lemma 3.1 to get an overall bound of $O(d^{-1/6})$. Hence,

$$\left|\frac{1}{\sqrt{d}}e^{\mathsf{D}}_{\max} - e^{\mathsf{SK}}_{\max}\right| = O(d^{-1/6}).$$

By Theorem 2.6, as $n \to \infty$, we have $-e^{\mathsf{SK}}_{\min} = e^{\mathsf{SK}}_{\max} \to \mathsf{P}_*$. It follows that $e^{\mathsf{D}}_{\max} = \mathsf{P}_*\sqrt{d} + O(d^{1/3})$, so that $\frac{1}{n}\,\mathbf{E}\,\mathsf{MinBis}(G) = \frac{nd}{4} - \mathsf{P}_*\sqrt{\frac{d}{4}} + o(\sqrt{d})$. □

So, what is interpolation? The idea is that we interpolate between the Hamiltonians of the two models. For $d \in \mathbb{R}$ (not necessarily an integer), let $H_d^{\mathsf{D}}(x) = \sum_{(i,j)\in E} x_i x_j$ where $G \sim G^{\mathsf{Poi}}(n,d)$. Hereafter, we will suppress the dependence on n and β for simplicity.

In particular, for $t \in [0,1]$, define the interpolating Hamiltonian $H_t(x) = \frac{1}{\sqrt{d}} H^{\mathsf{D}}_{d(1-t)}(x) + \sqrt{t} H^{\mathsf{SK}}(x)$.[1] The corresponding free energy is just $\phi(t) = \frac{1}{n}\,\mathbf{E}\log Z_t$, where $Z_t = \sum_{x\in\{\pm 1\}^n} \exp(\beta H_t(x))$, and here $\mathbf{E}$ is over the disorder in both the SK hamiltonian (iid Gaussians) and the Poissonized graph $G \sim G^{\mathsf{Poi}}(n, d(1-t))$. Observe that $H_0 = H_d^{\mathsf{D}}$ and $H_1 = H^{\mathsf{SK}}$, from which we conclude that

$$\left|\phi^{\mathsf{SK}} - \phi_d^{\mathsf{D}}\right| = \left|\phi(1) - \phi(0)\right| = \left|\int_0^1 \frac{\partial \phi(t)}{\partial t}\,\mathrm{d}t\right| \leqslant \int_0^1 \left|\frac{\partial \phi(t)}{\partial t}\right| \mathrm{d}t\,.$$

Thus, it suffices to bound the derivative of the interpolating free energy. The crucial formula is the following, which we will not prove in full detail on the board (but I will record it here for reference):

Lemma 3.2 ([DMS17, Equations 2.17, 2.18])**.** *We have*

$$\frac{\partial \phi(t)}{\partial t} = \left(\frac{\partial \phi(t)}{\partial t}\right)_{\mathsf{SK}} + \left(\frac{\partial \phi(t)}{\partial t}\right)_{\mathsf{D}}$$
$$\left(\frac{\partial \phi(t)}{\partial t}\right)_{\mathsf{SK}} = \frac{\beta^2}{4}(1 - \langle R_2^2\rangle_t)$$
$$\left(\frac{\partial \phi(t)}{\partial t}\right)_{\mathsf{D}} = -d \log\cosh\left(\frac{\beta}{\sqrt{d}}\right) + d\sum_{\ell\geqslant 1} \frac{(-1)^\ell}{\ell}\tanh\left(\frac{\beta}{\sqrt{d}}\right)^\ell \mathbf{E}[\langle R_\ell^2\rangle_t],$$

where $R_\ell \triangleq \frac{1}{n}\sum_{i\in[n]}\prod_{k\in[\ell]}\sigma_k[\ell]$, where $(\sigma_k)_{k\geqslant 1}$ are iid samples (replicas) from μ_{H_t}.

The proof is somewhat technical, but we will sketch out the main ideas.

Proof sketch. Below, recall that G is the randomness of the diluted graph with average degree $d(1-t)$, and $W \sim \mathsf{GOE}(n)$ is the randomness in the SK Hamiltonian. Recall now that $H_t(\sigma) = \frac{1}{\sqrt{d}} H^{\mathsf{D}}_{d(1-t)}(\sigma) + \sqrt{t}H^{\mathsf{SK}}(\sigma)$. Hence, if we condition on the randomness in G, then the only t dependence in H_t is through the SK term. By writing out the definition of $\phi(t) = \frac{1}{n}\mathbf{E}_{G,W}\log Z_t(G,W)$, we see that

$$\begin{aligned}\partial_t \phi(t) = \frac{1}{n}\mathop{\mathbf{E}}_{G,W}\frac{\partial_t Z_t}{Z_t} &= \underbrace{\frac{1}{n}\mathop{\mathbf{E}}_{W}\frac{\partial_t \sum_{\sigma\in\{\pm1\}^n}\exp(\beta H_t(\sigma))}{Z_t}}_{(\partial_t\phi(t))_{\mathsf{SK}}} + \underbrace{\frac{1}{n}\mathop{\mathbf{E}}_{G}\mathrm{stuff(G)}}_{(\partial_t\phi(t))_{\mathsf{D}}},\\ &= \frac{1}{n}\mathbf{E}\frac{\sum_{\sigma\in\{\pm1\}^n}\beta\partial_t(H_t(\sigma))\exp(\beta H_t(\sigma))}{Z_t} + (\partial_t\phi(t))_{\mathsf{D}}\\ &= \frac{\beta}{2\sqrt{t}n}\left\langle H^{\mathsf{SK}}(\sigma)\right\rangle_t + (\partial_t\phi(t))_{\mathsf{D}}.\end{aligned}$$

Here, stuff(G) is what we get by taking the time derivative of the t dependent measure for the random graph G. This is where we will use the Poissonized graph model, to make this measure product.

Let us focus on the SK part, since this is fairly standard and not too difficult to compute. The main idea is that we can use Gaussian integration by parts, which states the following: If $\boldsymbol{G} = (G(\sigma))_{\sigma\in\Sigma}$ is a centered Gaussian process over a finite index set Σ, and f is a sufficiently nice function, then

$$\mathbf{E}[G(\sigma)f(\boldsymbol{G})] = \sum_{\tau\in\Sigma}\mathbf{E}[G(\sigma)G(\tau)]\,\mathbf{E}[\partial_{G(\tau)}f(\boldsymbol{G})].$$

[1] The reason for putting the $1-t$ in the graph parameter is to preserve the first and second moments of the Hamiltonian process.

The crucial observation is that the Hamiltonian itself $H^{\mathsf{SK}}(\sigma)$ is a centered Gaussian process over $\Sigma = \{\pm 1\}^n$ with $\mathbf{E}[H^{\mathsf{SK}}(\sigma)H^{\mathsf{SK}}(\tau)] = \frac{1}{2}R(\sigma,\tau)^2$. Hence, we can apply the above formula with $G(\sigma) = H^{\mathsf{SK}}(\sigma)$ and $f(\boldsymbol{G}) = \mu_{H_t}(\sigma)$. By calculus, we have $\partial_{H^{\mathsf{SK}}(\tau)} \frac{\exp(\beta H_t(\sigma))}{Z_t} = \beta\sqrt{t} \cdot (\delta_{\sigma=\tau}\mu_{H_t}(\sigma) - \mu_{H_t}(\sigma)\mu_{H_t}(\tau))$, so we get

$$\begin{aligned}
\mathbf{E}\left\langle H^{\mathsf{SK}}(\sigma)\right\rangle_t &= \sum_{\sigma} \mathbf{E}[H^{\mathsf{SK}}(\sigma)\mu_{H_t}(\sigma)] \\
&= \sum_{\sigma,\tau} \mathbf{E}[H^{\mathsf{SK}}(\sigma)H^{\mathsf{SK}}(\tau)]\, \mathbf{E}[\partial_{H^{\mathsf{SK}}(\tau)}\mu_{H_t}(\sigma)] && \text{(Gaussian integration by parts)} \\
&= \frac{1}{2}\beta\sqrt{t} \cdot \left(\sum_{\sigma} \mathbf{E}[\mu_{H_t}(\sigma)] - \sum_{\sigma\neq\tau} R(\sigma,\tau)^2\, \mathbf{E}[\mu_{H_t}(\sigma)\mu_{H_t}(\tau)]\right) \\
&= \frac{1}{2}\beta\sqrt{t} \cdot \left(1 - \mathbf{E}\left\langle R(\sigma,\tau)^2\right\rangle_t\right).
\end{aligned}$$

To summarize, we have just proved that

$$(\partial_t \phi(t))_{\mathsf{SK}} = \frac{\beta^2}{4}(1 - \mathbf{E}\left\langle R_2^2\right\rangle_t).$$

We will not go into the calculation for the diluted part, because it is a little technical, but you can check out [DMS17] for the details. The high level idea is to use a "Poisson integration by parts" identity, which yields a tractable analytic formula for the time derivative. The nasty form of the final formula comes from writing down some algebraic identities and Taylor expanding the result. □

Remark 3.3. The actual proof has some additional subtleties that I swept under the rug. Firstly, in order to get the correct asymptotics, we need to actually expand out the above infinite series in a clever way to get higher order Taylor terms. In turn, this requires getting a tight control on $\mathbf{E}\left\langle R_1^2\right\rangle_t$, which is the expected square of the magnetization. A clever way around this issue is to instead study the restricted Gibbs measures on Ω_n, in which case $R_1 = 0$ by definition. One then has to show that the Parisi formula is correct when restricted to Ω_n, which requires another argument.

References

[BGT10] Mohsen Bayati, David Gamarnik, and Prasad Tetali. Combinatorial approach to the interpolation method and scaling limits in sparse random graphs. In *Proceedings of the forty-second ACM symposium on Theory of computing*, pages 105–114, 2010. 2

[Bol88] Béla Bollobás. The isoperimetric number of random regular graphs. *European Journal of combinatorics*, 9(3):241–244, 1988. 2

[CO07] Amin Coja-Oghlan. On the laplacian eigenvalues of gn, p. *Combinatorics, Probability and Computing*, 16(6):923–946, 2007. 2

[DMS17] Amir Dembo, Andrea Montanari, and Subhabrata Sen. Extremal cuts of sparse random graphs. 2017. 1, 4, 5, 6

[FO05] Uriel Feige and Eran Ofek. Spectral techniques applied to sparse random graphs. *Random Structures & Algorithms*, 27(2):251–275, 2005. 2

[GL18] David Gamarnik and Quan Li. On the max-cut of sparse random graphs. *Random Structures & Algorithms*, 52(2):219–262, 2018. 2

[Tal06] Michel Talagrand. The parisi formula. *Annals of mathematics*, pages 221–263, 2006. 3