Avrim's notes on clique approximation hardness. 10/30/98

As Steven's notes mention, there is a known constant $\epsilon > 0$ such that approximating MAX-3-SAT to better than a $(1+\epsilon)$ ratio is NP-hard. Specifically, several results have shown stronger and stronger versions of the following:

A poly time transformation T from 3CNF formulas to 3CNF formulas with the following property:

- If formula ϕ is satisfiable, then $\phi' = T(\phi)$ is satisfiable.
- If ϕ is not satisfiable, then not only is ϕ' not satisfiable, but also any assignment to the variables of ϕ' satisfies at most a $(1-\varepsilon)$ fraction of the clauses of ϕ'. The quantity ε is a specific constant greater than 0, like 1/113.

This transformation is the [ALMSS92] result.

It turns out you can use this to show that for the Clique problem, not only is it NP-hard to find the (size of the) largest clique, and not only can't you approximate it to, say, a factor of 5, but you can't even approximate the optimal to a ratio of $n^{\epsilon'}$ for some $\epsilon' > 0$. (Actually, it's recently been shown by Hastad that you can't even approximate it to $n^{1-\epsilon'}$).

Here's a neat randomized reduction. We will show that if you could solve the Clique approximation, then you could solve SAT in randomized poly time (so would imply NP $\subseteq$ RP). This can be derandomized using the [Impagliazzo-Zuckerman] "recycling random bits via walks on expanders" result, which then implies hardness based on P $\neq$ NP.

The reduction: Pick a clause at random from ϕ'. Repeat this $\log_2(n)$ times and call that set of at most $\log_2(n)$ clauses S_1. Repeat that n^3 times to create $S_2, \ldots, S_{n^3}$. For each set S_i create one node for each assignment to the variables in that set which satisfy the clauses in that set. So, there are at most $n^3 2^{3\log_2(n)} = n^6$ nodes. Let N be the number of nodes. Create the edges by connecting two vertices if their partial assignments are consistent.

If ϕ' is satisfiable, then there is a clique of size n^3. If ϕ' is not satisfiable, consider some fixed assignment A to all the variables. Let X_i be the random variable that equals 1 if A satisfies S_i and equals 0 if it does not. Over the random choice of the clauses in S_i, $\mathbf{Pr}[X_i = 1] < (1-\varepsilon)^{\log_2(n)} < \frac{1}{2}n^{-\varepsilon}$. Let $Y = X_1 + \ldots + X_{n^3}$, so $\mathbf{E}[Y] < \frac{1}{2}n^{3-\varepsilon}$. Each X_i is an independent $\{0,1\}$ random variable, so by Chernoff bounds, $\mathbf{Pr}[Y > n^{3-\varepsilon}] < e^{-n^{3-\varepsilon}/6} < e^{-n^2}$. Since there are "only" 2^n possible assignments A, with overwhelming probability $(> 1 - 2^n e^{-n^2})$ there is no assignment that satifies more than $n^{3-\varepsilon}$ sets S_i, and so there is no clique of size greater than $n^{3-\varepsilon}$. Thus, we have a gap of n^ε or $N^{\varepsilon/6}$. ■

Side note: if you increase the size of the sets, you increase the gap as a function of n (e.g., if you replace $\log_2(n)$ with $\frac{1}{\varepsilon}\ln(n)$ then the gap becomes roughly n), but you also increase N, so as a function of N it doesn't help.