Advanced Algorithms	20 November 2018

Lecture 10: Graph Sparsification I

Lecturer: Mohsen Ghaffari *Scribe: Davin Choo*

In this lecture and the next, we will look at graph sparsification techniques. Given a simple, unweighted, undirected graph G with n vertices and m edges, can we *sparsify* G by ignoring some edges such that certain desirable properties still hold? In this lecture, we will look at *preserving distances*.

1 Preserving distances

We will consider simple, unweighted and undirected graphs G. For any pair of vertices $u, v \in G$, denote the shortest path between them by $P_{u,v}$. Then, the distance between u and v in graph G, denoted by $d_G(u, v)$, is simply the length of shortest path $P_{u,v}$ between them.

Definition 1 (((α, β)-spanners). *Consider a graph $G = (V, E)$ with $|V| = n$ vertices and $|E| = m$ edges. For given $\alpha \geq 1$ and $\beta \geq 0$, an (α, β)-spanner is a subgraph $G' = (V, E')$ of G, where $E' \subseteq E$, such that*

$$d_G(u, v) \leq d_{G'}(u, v) \leq \alpha \cdot d_G(u, v) + \beta$$

Remark The first inequality is because G' has less edges than G. The second inequality upper bounds how much the distances "blow up" in the sparser graph G'.

For an (α, β)-spanner, α is called the *multiplicative stretch* of the spanner and β is called the *additive stretch* of the spanner. One would then like to construct spanners with small $|E'|$ and stretch factors. An $(\alpha, 0)$-spanner is called a *α-multiplicative spanner*, and a $(1, \beta)$-spanner is called a *β-additive spanner*. We shall first look at α-multiplicative spanners, then β-additive spanners in a systematic fashion:

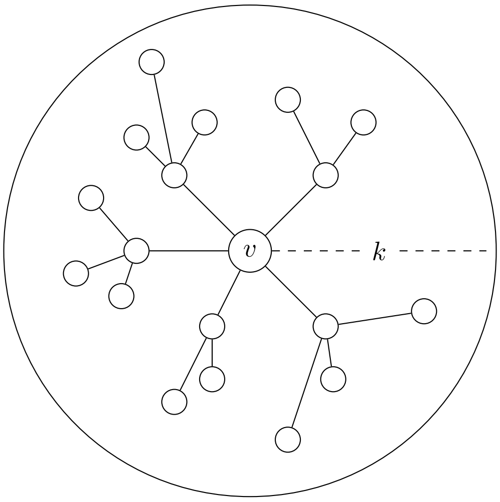
1. State the result with respect to the number of edges and the stretch factor
2. Give the construction
3. Bound the total number of edges $|E'|$
4. Prove that the stretch factor holds

Remark One way to prove the existence of an (α, β)-spanner is to use the *probabilistic method*: Instead of giving an explicit construction, one designs a random process and argues that the probability that the spanner existing is *strictly larger than 0*. However, this may be somewhat unsatisfying as such proofs do not usually yield a usable construction. On the other hand, the randomized constructions shown later are explicit and will yield a spanner *with high probability*[1].

1.1 α-multiplicative spanners

Let us first state a fact regarding the girth of a graph G. The *girth* of a graph G, denoted $g(G)$, is defined as the length of the shortest cycle in G. Suppose $g(G) > 2k$, then for any vertex v, the subgraph formed by the k-hop neighbourhood of v is a tree with distinct vertices. This is because the k-hop neighbourhood of v cannot have a cycle since $g(G) > 2k$.

[1] This is shown by invoking concentration bounds such as Chernoff.



Theorem 2. [ADD+93] *For a fixed $k \geq 1$, every graph G on n vertices has a $(2k-1)$-multiplicative spanner with $\mathcal{O}(n^{1+1/k})$ edges.*

Proof.
Construction

1. Initialize $E' = \emptyset$
2. For $e = (u, v) \in E$ (in arbitrary order):
   If $d_{G'}(u, v) \geq 2k$ currently, add (u, v) into E'.
   Otherwise, ignore it.

Number of edges We claim that $|E'| \in \mathcal{O}(n^{1+1/k})$. Suppose, for a contradiction, that $|E'| > 2n^{1+1/k}$. Let $G'' = (V'', E'')$ be a graph obtained by iteratively removing vertices with degree $\leq n^{1/k}$ from G'. By construction, $|E''| > n^{1+1/k}$ since at most $n \cdot n^{1/k}$ edges are removed. Observe the following:

- $g(G'') \geq g(G') \geq 2k+1$, since girth does not decrease with fewer edges.
- Every vertex in G'' has degree $\geq n^{1/k} + 1$, by construction.
- Pick an arbitrary vertex $v \in V''$ and look at its k-hop neighbourhood.

$$\begin{aligned}
n &\geq |V''| && \text{By construction} \\
&\geq |\{v\}| + \textstyle\sum_{i=1}^{k} |\{u \in V'' : d_{G''}(u, v) = i\}| && \text{If we only look at } k\text{-hop neighbourhood from } v \\
&\geq 1 + \textstyle\sum_{i=1}^{k} (n^{1/k}+1)(n^{1/k})^{i-1} && \text{Since vertices are distinct and have degree} \geq n^{1/k}+1 \\
&= 1 + (n^{1/k}+1)\frac{(n^{1/k})^k - 1}{n^{1/k}-1} && \text{Sum of geometric series} \\
&> 1 + (n-1) && \text{Since } (n^{1/k}+1) > (n^{1/k}-1) \\
&= n
\end{aligned}$$

This is a contradiction since we showed $n > n$. Hence, $|E'| \leq 2n^{1+1/k} \in \mathcal{O}(n^{1+1/k})$.

Stretch factor For $e = (u, v) \in E$, $d_{G'}(u, v) \leq (2k-1) \cdot d_G(u, v)$ since we only leave e out of E' if the distance is at most the stretch factor at the point of considering e. For any $u, v \in V$, let $P_{u,v}$ be the shortest path between u and v in G. Say, $P_{u,v} = (u, w_1, \ldots, w_k, v)$. Then,

$$\begin{aligned}
d_{G'}(u, v) &\leq d_{G'}(u, w_1) + \cdots + d_{G'}(w_k, v) && \text{Upper bounded by simulating } P_{u,v} \text{ in } G' \\
&\leq (2k-1) \cdot d_G(u, w_1) + \cdots + (2k-1) \cdot d_G(w_k, v) && \text{Apply edge stretch to each edge} \\
&= (2k-1) \cdot (d_G(u, w_1) + \cdots + d_G(w_k, v)) && \text{Rearrange} \\
&= (2k-1) \cdot d_G(u, v) && \text{Definition of } P_{u,v}
\end{aligned}$$

□

Conjecture 1. [Erd64] *For a fixed $k \geq 1$, there exists a family of graphs on n vertices with girth at least $2k+1$ and $\Omega(n^{1+1/k})$ edges.*

If the conjecture is true, then the construction is optimal. Notably, the construction will not remove any edges for any graph that satisfies the conjecture.

1.2 β-additive spanners

In this section, we will use a random process to select a subset of vertices by independently selecting vertices to join the subset. The following claim will be useful for analysis:

Claim 3. *If one picks vertices independently with probability p to be in $S \subseteq V$, where $|V| = n$, then*

1. $\mathbb{E}[|S|] = np$
2. *For any vertex v with degree $d(v)$ and neighbourhood $N(v) = \{u \in V : (u, v) \in E\}$,*
   - $\mathbb{E}[|N(v) \cap S|] = d(v) \cdot p$
   - $\Pr[|N(v) \cap S| = 0] \leq e^{-\frac{d(v) \cdot p}{2}}$

Proof. $\forall v \in V$, let X_v be the indicator whether $v \in S$. By construction, $\mathbb{E}[X_v] = \Pr[X_v = 1] = p$.

1.
$$\begin{aligned} \mathbb{E}[|S|] &= \mathbb{E}[\textstyle\sum_{v \in V} X_v] && \text{By construction of } S \\ &= \textstyle\sum_{v \in V} \mathbb{E}[X_v] && \text{Linearity of expectation} \\ &= \textstyle\sum_{v \in V} p && \text{Since } \mathbb{E}[X_v] = \Pr[X_v = 1] = p \\ &= np && \text{Since } |V| = n \end{aligned}$$

2.
$$\begin{aligned} \mathbb{E}[|N(v) \cap S|] &= \mathbb{E}[\textstyle\sum_{v \in N(v)} X_v] && \text{By definition of } N(v) \cap S \\ &= \textstyle\sum_{v \in N(v)} \mathbb{E}[X_v] && \text{Linearity of expectation} \\ &= \textstyle\sum_{v \in N(v)} p && \text{Since } \mathbb{E}[X_v] = \Pr[X_v = 1] = p \\ &= d(v) \cdot p && \text{Since } |N(v)| = d(v) \end{aligned}$$

   By one-sided Chernoff bound,

$$\begin{aligned} \Pr[|N(v) \cap S| = 0] &= \Pr[|N(v) \cap S| \leq (1 - 1) \cdot \mathbb{E}[|N(v) \cap S|]] \\ &\leq e^{-\frac{\mathbb{E}[|N(v) \cap S|]}{2}} \\ &= e^{-\frac{d(v) \cdot p}{2}} \end{aligned}$$

□

Remark As a reminder, $\widetilde{\mathcal{O}}$ hides logarithmic factors. For example, $\mathcal{O}(n \log^{1000} n) \subseteq \widetilde{\mathcal{O}}(n)$.

Theorem 4. *[ACIM99] For a fixed $k \geq 1$, every graph G on n vertices has a 2-additive spanner with $\widetilde{\mathcal{O}}(n^{3/2})$ edges.*

Proof.
Construction Partition vertex set V into *light vertices* L and *heavy vertices* H, where

$$L = \{v \in V : \deg(v) \leq n^{1/2}\} \text{ and } H = \{v \in V : \deg(v) > n^{1/2}\}$$

1. Let E_1' be the set of all edges incident to some vertex in L.
2. Initialize $E_2' = \emptyset$.
   - Choose $S \subseteq V$ by independently putting each vertex into S with probability $10n^{-1/2} \log n$.
   - For each $s \in S$, add a Breadth-First-Search (BFS) tree rooted at s to E_2'

Select edges in spanner to be $E' = E_1' \cup E_2'$.
Number of edges

1. Since there are at most n light vertices, $|E_1'| \leq n \cdot n^{1/2} = n^{3/2}$.
2. By Claim 3 with $p = 10n^{-1/2} \log n$, $\mathbb{E}[|S|] = n \cdot 10n^{-1/2} \log n = 10n^{1/2} \log n$. Then, since every BFS tree has $n - 1$ edges[2], $|E_2'| \leq n \cdot |S|$, thus

$$\mathbb{E}[|E'|] = \mathbb{E}[|E_1' \cup E_2'|] \leq \mathbb{E}[|E_1'| + |E_2'|] = \mathbb{E}[|E_1'|] + \mathbb{E}[|E_2'|] \leq n^{3/2} + n \cdot 10n^{1/2} \log n \in \widetilde{\mathcal{O}}(n^{3/2})$$

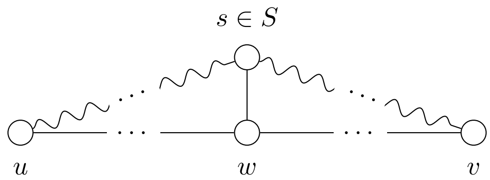
[2] Though we may have repeated edges

Stretch factor Consider two arbitrary vertices u and v with the shortest path $P_{u,v}$ in G. Let h be the number of heavy vertices in $P_{u,v}$. We split the analysis into two cases: (i) $h \leq 1$; (ii) $h \geq 2$. Recall that a heavy vertex has degree at least $n^{1/2}$.

Case (i) All edges in $P_{u,v}$ are adjacent to a light vertex and are thus in E_1'. Hence, $d_{G'}(u,v) = d_G(u,v)$, with additive stretch 0.

Case (ii)

Claim 5. *Suppose there exists a vertex $w \in P_{u,v}$ such that $(w,s) \in E$ for some $s \in S$, then $d_{G'}(u,v) \leq d_G(u,v) + 2$.*



Proof.

$$\begin{aligned}
d_{G'}(u,v) &\leq d_{G'}(u,s) + d_{G'}(s,v) && \text{By triangle inequality} \\
&= d_G(u,s) + d_G(s,v) && \text{Since we add the BFS tree rooted at } s \\
&\leq d_G(u,w) + d_G(w,s) + d_G(s,w) + d_G(w,v) && \text{By triangle inequality} \\
&\leq d_G(u,w) + 1 + 1 + d_G(w,v) && \text{Since } (s,w) \in E,\ d_G(w,s) = d_G(s,w) = 1 \\
&\leq d_G(u,v) + 2 && \text{Since } u,w,v \text{ lie on } P_{u,v}
\end{aligned}$$

□

Let w be a heavy vertex in $P_{u,v}$ with degree $d(w) > n^{1/2}$. By Claim 3 with $p = 10n^{-1/2}\log n$, $\Pr[|N(w) \cap S| = 0] \leq e^{-\frac{10\log n}{2}} = n^{-5}$. Taking union bound over all possible pairs of vertices u and v,

$$\Pr[\exists u,v \in V, P_{u,v} \text{ has no neighbour in } S] \leq \binom{n}{2} n^{-5} \leq n^{-3}$$

Then, Claim 5 tells us that the additive stretch factor is at most 2 with probability $\geq 1 - \frac{1}{n^3}$.

Therefore, with high probability ($\geq 1 - \frac{1}{n^3}$), the construction yields a 2-additive spanner. □

Theorem 6. *[Che13] For a fixed $k \geq 1$, every graph G on n vertices has a 4-additive spanner with $\widetilde{O}(n^{7/5})$ edges.*

Proof.
Construction Partition vertex set V into *light vertices* L and *heavy vertices* H, where

$$L = \{v \in V : \deg(v) \leq n^{2/5}\} \text{ and } H = \{v \in V : \deg(v) > n^{2/5}\}$$

1. Let E_1' be the set of all edges incident to some vertex in L.
2. Initialize $E_2' = \emptyset$.
   - Choose $S \subseteq V$ by independently putting each vertex into S with probability $30n^{-3/5}\log n$.
   - For each $s \in S$, add a Breadth-First-Search (BFS) tree rooted at s to E_2'
3. Initialize $E_3' = \emptyset$.
   - Choose $S' \subseteq V$ by independently putting each vertex into S' with probability $10n^{-2/5}\log n$.
   - For each heavy vertex $w \in H$, if there exists edge (w,s') for some $s' \in S'$, add (w,s') to E_3'.
   - $\forall s,s' \in S'$, add the shortest path between s and s' with $\leq n^{1/5}$ internal heavy vertices to E_3'. Note: If all paths between s and s' contain $> n^{1/5}$ heavy vertices, do not add any edge to E_3'.

Select edges in spanner to be $E' = E'_1 \cup E'_2 \cup E'_3$.

Number of edges

- Since there are at most n light vertices, $|E'_1| \leq n \cdot n^{2/5} = n^{7/5}$.
- By Claim 3 with $p = 30n^{-3/5} \log n$, $\mathbb{E}[|S|] = n \cdot 30n^{-3/5} \log n = 30n^{2/5} \log n$. Then, since every BFS tree has $n-1$ edges[3], $|E'_2| \leq n \cdot |S| = 30n^{7/5} \log n \in \widetilde{\mathcal{O}}(n^{7/5})$.
- Since there are $\leq n$ heavy vertices, $\leq n$ edges of the form (v, s') for $v \in H, s' \in S'$ will be added to E'_3. Then, for shortest $s - s'$ paths with $\leq n^{1/5}$ heavy internal vertices, only edges adjacent to the heavy vertices need to be counted because those adjacent to light vertices are already accounted for in E'_1. By Claim 3 with $p = 10n^{-2/5} \log n$, $\mathbb{E}[|S'|] = n \cdot 10n^{-2/5} \log n = 10n^{3/5} \log n$. So, E'_3 contributes $\leq n + \binom{|S'|}{2} \cdot n^{1/5} \leq n + (10n^{3/5} \log n)^2 \cdot n^{1/5} \in \widetilde{\mathcal{O}}(n^{7/5})$ edges to the count of $|E'|$.

Stretch factor Consider two arbitrary vertices u and v with the shortest path $P_{u,v}$ in G. Let h be the number of heavy vertices in $P_{u,v}$. We split the analysis into three cases: (i) $h \leq 1$; (ii) $2 \leq h \leq n^{1/5}$; (iii) $h > n^{1/5}$. Recall that a heavy vertex has degree at least $n^{2/5}$.

Case (i) All edges in $P_{u,v}$ are adjacent to a light vertex and are thus in E'_1. Hence, $d_{G'}(u,v) = d_G(u,v)$, with additive stretch 0.

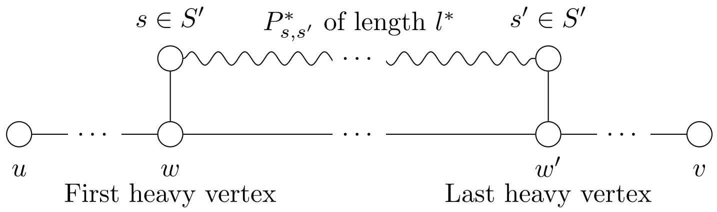
Case (ii) Denote the first and last heavy vertices in $P_{u,v}$ as w and w' respectively. Recall that in Case (ii), including w and w', there are at most $n^{1/5}$ heavy vertices between w and w'. By Claim 3, with $p = 10n^{-2/5} \log n$, $\Pr[|N(w) \cap S'| = 0] = \Pr[|N(w') \cap S'| = 0] \leq e^{-\frac{n^{2/5} \cdot 10n^{-2/5} \log n}{2}} = n^{-5}$.

Let $s, s' \in S'$ be adjacent vertices to w and w' respectively. Observe that $s - w - w' - s'$ is a path between s and s' with at most $n^{1/5}$ internal heavy vertices. Let $P^*_{s,s'}$ be the shortest path of length l^* from s to s' with at most $n^{1/5}$ internal heavy vertices. By construction, we have added $P^*_{s,s'}$ to E'_3. Observe:

- By definition of $P^*_{s,s'}$, $l^* \leq d_G(s,w) + d_G(w,w') + d_G(w',s') = d_G(w,w') + 2$.
- Since there are no internal heavy vertices between $u - w$ and $w' - v$, Case (i) tells us that $d_{G'}(u,w) = d_G(u,w)$ and $d_{G'}(w',v) = d_G(w',v)$.

Thus,

$$
\begin{array}{lll}
 & d_{G'}(u,v) & \\
= & d_{G'}(u,w) + d_{G'}(w,w') + d_{G'}(w',v) & \text{Decomposing } P_{u,v} \text{ in } G' \\
\leq & d_{G'}(u,w) + d_{G'}(w,s) + d_{G'}(s,s') + d_{G'}(s',w') + d_{G'}(w',v) & \text{Triangle inequality} \\
= & d_{G'}(u,w) + d_{G'}(w,s) + l^* + d_{G'}(s',w') + d_{G'}(w',v) & P^*_{s,s'} \text{ is added to } E'_3 \\
\leq & d_{G'}(u,w) + d_{G'}(w,s) + d_G(w,w') + 2 + d_{G'}(s',w') + d_{G'}(w',v) & \text{Since } l^* \leq d_G(w,w') + 2 \\
= & d_{G'}(u,w) + 1 + d_G(w,w') + 2 + 1 + d_{G'}(w',v) & \text{Since } (w,s) \in E' \text{ and } (s',w') \in E' \\
 & & d_{G'}(w,s) = d_{G'}(s',w') = 1 \\
= & d_G(u,w) + 1 + d_G(w,w') + 2 + 1 + d_G(w',v) & \text{Since } d_{G'}(u,w) = d_G(u,w) \text{ and} \\
 & & d_{G'}(w',v) = d_G(w',v) \\
\leq & d_G(u,v) + 4 & \text{By definition of } P_{u,v}
\end{array}
$$

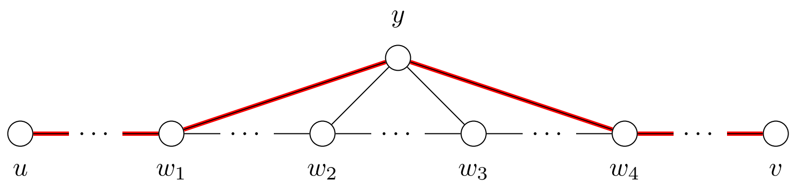


Case (iii)

[3] Though we may have repeated edges

Claim 7. *There cannot be a vertex y that is a common neighbour to more than 3 heavy vertices in $P_{u,v}$.*

Proof. Suppose, for a contradiction, that y is adjacent to $w_1, w_2, w_3, w_4 \in P_{u,v}$ as shown in the picture. Then $u - w_1 - y - w_4 - v$ is a shorter $u - v$ path than $P_{u,v}$, contradicting the fact that $P_{u,v}$ is the shortest $u - v$ path.



Note that if y is on $P_{u,v}$, it immediately contradicts that $P_{u,v}$ was the shortest path involving all of $\{y, w_1, w_2, w_3, w_4\}$. □

Claim 7 tells us that $|\bigcup_{w \in \text{Heavy}} N(w)| \geq \sum_{w \in \text{Heavy}} |N(w)| \cdot \frac{1}{3}$. Let

$$N_{u,v} = \{x \in V : (x, w) \in P_{u,v} \text{ for some } w \in P_{u,v}\}$$

Applying Claim 3 with $p = 30 \cdot n^{-3/5} \cdot \log n$ and Claim 7, we get

$$\mathbb{E}[|N_{u,v} \cap S|] \geq n^{1/5} \cdot n^{2/5} \cdot \frac{1}{3} \cdot 30 \cdot n^{-3/5} \cdot \log n = 10 \log n$$

and

$$\Pr[|N(v) \cap S| = 0] \leq e^{-\frac{10 \log n}{2}} = n^{-5}$$

Taking union bound over all possible pairs of vertices u and v,

$$\Pr[\exists u, v \in V, P_{u,v} \text{ has no neighbour in } S] \leq \binom{n}{2} n^{-5} \leq n^{-3}$$

Then, Claim 5 tells us that the additive stretch factor is at most 4 with probability $\geq 1 - \frac{1}{n^3}$.

Therefore, with high probability ($\geq 1 - \frac{1}{n^3}$), the construction yields a 4-additive spanner.

□

Remark Suppose the shortest $u - v$ path $P_{u,v}$ contains a vertex from S, say s. Then, $P_{u,v}$ is contained in E' since we include the BFS tree rooted at s because it is the shortest $u - s$ path and shortest $s - v$ path by definition. In other words, the triangle inequality between u, s, v becomes tight.

Concluding remarks

	Additive stretch factor β	Number of edges	Remarks
[ACIM99]	2	$\widetilde{\mathcal{O}}(n^{3/2})$	Tight [Woo06]
[Che13]	4	$\widetilde{\mathcal{O}}(n^{7/5})$	Open: Is $\widetilde{\mathcal{O}}(n^{4/3})$ possible?
[BKMP05]	≥ 6	$\widetilde{\mathcal{O}}(n^{4/3})$	Tight [AB17]

The additive stretch factors appear to be in even numbers because current constructions "leave" the shortest path, then "re-enter" it later, introducing an even number of extra edges. Regardless, a k-additive spanner is also a $(k-1)$-additive spanner.

References

[AB17] Amir Abboud and Greg Bodwin. The &frac; 43 additive spanner exponent is tight. *Journal of the ACM (JACM)*, 64(4):28, 2017.

[ACIM99] Donald Aingworth, Chandra Chekuri, Piotr Indyk, and Rajeev Motwani. Fast estimation of diameter and shortest paths (without matrix multiplication). *SIAM Journal on Computing*, 28(4):1167–1181, 1999.

[ADD+93] Ingo Althöfer, Gautam Das, David Dobkin, Deborah Joseph, and José Soares. On sparse spanners of weighted graphs. *Discrete & Computational Geometry*, 9(1):81–100, 1993.

[BKMP05] Surender Baswana, Telikepalli Kavitha, Kurt Mehlhorn, and Seth Pettie. New constructions of (α, β)-spanners and purely additive spanners. In *Proceedings of the sixteenth annual ACM-SIAM symposium on Discrete algorithms*, pages 672–681. Society for Industrial and Applied Mathematics, 2005.

[Che13] Shiri Chechik. New additive spanners. In *Proceedings of the twenty-fourth annual ACM-SIAM symposium on Discrete algorithms*, pages 498–512. Society for Industrial and Applied Mathematics, 2013.

[Erd64] P. Erds. Extremal problems in graph theory. In *IN THEORY OF GRAPHS AND ITS APPLICATIONS, PROC. SYMPOS. SMOLENICE*, pages 29–36, 1964.

[Woo06] David P Woodruff. Lower bounds for additive spanners, emulators, and more. In *Foundations of Computer Science, 2006. FOCS'06. 47th Annual IEEE Symposium on*, pages 389–398. IEEE, 2006.