NETWORK CODING IN UNDIRECTED GRAPHS IS EITHER VERY HELPFUL OR NOT HELPFUL AT ALL

Sumegha Garg

Joint work with Mark Braverman and Ariel Schwartzman

Princeton University
MULTICOMMODITY FLOW

- Graph $G = (V, E)$
- Capacity function $c: E \rightarrow R^+$
- Set $I$ of $k$ commodities: $\{(s_i, t_i), i \in [k]\}$
- Rate is $r$:
  - Flow between source-sink pair is at least $r$
  - Total flow through an edge is upper bounded by its capacity
- $MCF(G)$ denotes the multicommodity flow rate
FLOWS OF PACKETS

- No re-encodings at intermediate nodes
- The Question: Can we get better information rate if the commodities are bits and we use bit tricks on them?
NETWORK CODING

• First introduced in [ACLY ’00]
• $M = \{M_i\}_{i \in [k]}$ is the messages by sources
• Each edge has $f: M \rightarrow \Delta(e)$ (alphabet for $e$)
  • Function of alphabets on incoming edges
  • Entropy($e$) is upper bounded by capacity
• Each sink edge $t_i$ carries $M_i$
\( NC(G) \) denotes the network coding rate

\( r = \lim_{b \to \infty} \left( \frac{\max r_b}{b} \right) \) (All capacities are multiplied by \( b \))
• $NC(G)$ denotes the network coding rate

• $r = \lim_{b \to \infty} \left( \frac{\max r_b}{b} \right)$ (All capacities are multiplied by $b$)

\[ r_2 = 1, \ NC = \frac{1}{2} \]
• $NC(G)$ denotes the network coding rate

• $r = \lim_{b \to \infty} \left( \frac{\max r_b}{b} \right)$ (All capacities are multiplied by $b$)

• Decidable?

• $\text{Gap}(G) = \frac{NC(G)}{MCF(G)}$

• Is there a $G$ with $\text{Gap}(G) > 1$?

$\begin{align*}
r_2 &= 1, \quad NC = \frac{1}{2} \\
\end{align*}$
DIRECTED GRAPHS

• Yes! [HKL ‘04] [LL ‘04]
• $NC/MCF$ gap can be as large as $O(|G|)$
SPARSITY BOUND FOR UNDIRECTED

• $U \subseteq V, \text{Sparsity}(U, V \setminus U) = \frac{\text{Capacity}(U, V \setminus U)}{\text{Demand}(U, V \setminus U)}$
SPARSITY BOUND FOR UNDIRECTED

- \( U \subseteq V, \text{Sparsity}(U, V \setminus U) = \frac{\text{Capacity}(U, V \setminus U)}{\text{Demand}(U, V \setminus U)} \)
SPARSITY BOUND FOR UNDIRECTED

- \( U \subseteq V, \text{Sparsity}(U, V \setminus U) = \frac{\text{Capacity}(U, V \setminus U)}{\text{Demand}(U, V \setminus U)} \)
- \( \text{Sparsity}(G) = \min_{U \subseteq V} \text{Sparsity}(U, V \setminus U) \)
SPARSITY BOUND FOR UNDIRECTED

- $U \subseteq V, \text{Sparsity}(U, V \setminus U) = \frac{\text{Capacity}(U, V \setminus U)}{\text{Demand}(U, V \setminus U)}$

- $\text{Sparsity}(G) = \min_{U \subseteq V} \text{Sparsity}(U, V \setminus U)$

- $\text{MCF}(G) \leq \text{NC}(G) \leq \text{Sparsity}(G)$

- $\frac{\text{Sparsity}(G)}{O(\log |G|)} \leq \text{MCF}(G) \leq \text{Sparsity}(G)$ [LR ‘99]

- Expander

- Information Upper Bound
MAXIMUM GAP IN UNDIRECTED GRAPHS

\[
\frac{NC(G)}{O(\log |G|)} \leq \frac{Sparsity(G)}{O(\log |G|)} \leq MCF(G) \leq NC(G)
\]

- Maximum gap can be \(O(\log |G|)\)
MAXIMUM GAP IN UNDIRECTED GRAPHS

- Maximum gap can be $O(\log |G|)$
- Li and Li conjectured that $MCF(G) = NC(G) \forall G$ [LL '04]
  - Coding gives no advantage
• [SYC ‘03] [KS ‘03] [K ‘03] [JFY ‘05] [KS ‘06] [HKL ‘06] produced techniques for lower and upper bounding NC
Either the conjecture is true or it must be nearly ‘completely false’

Theorem 1. Given a graph $G$ that achieves a gap of $1 + \epsilon$ between the multicommodity flow rate and the network coding rate, we can construct an infinite family of graphs $\tilde{G}$ that achieve a gap of $O\left(\log |\tilde{G}|\right)^c$ for some constant $c < 1$ that depends on the original graph $G$. 
Given two graphs $G_1$ and $G_2$ with gaps $(1 + \epsilon_1)$ and $(1 + \epsilon_2)$ respectively, we construct a new graph $G$ with gap $(1 + \epsilon_1)(1 + \epsilon_2)$ while keeping a check on size of $G$.

Apply this construction repeatedly on the starting graph with the gap
• Replace each edge of $G_1$ by a source-sink pair of $G_2$
• Replace each edge of $G_1$ by a source-sink pair of $G_2$
• Keep the source/sink pairs of $G_1$ and edges of $G_2$
• Keep the source/sink pairs of $G_1$ and edges of $G_2$
Idea: Effective capacity seen by $G_1$ under network coding is greater than that seen under flows.

- Information transferred grows linearly with capacity.
  - Gaps should multiply.
• For $G_2$, there is a gap only when all source-sink pairs send flows simultaneously.
• We have multiple copies of $G_1$ and each source-sink pair in a copy of $G_2$ replaces an edge in a different copy of $G_1$. 
UPDATED GRAPH TENSOR

First copy of $G_1$
Second copy of $G_1$
Not the final tensor: Final Construction based on high girth bipartite graphs
Main Theorem

- Start with a graph $G_0 = G$ with gap $(1 + \epsilon)$

**Theorem 2.** Given a graph $G$ of size $n$ with a gap of $1 + \epsilon$ between the multicommodity flow rate and the network coding rate, we can create another graph $G'$ of size $n^{c^2}$ and a gap of $(1 + \epsilon)^2$, where $c$ depends on the diameter of the graph $G$. 
ITERATIVE TENSORING

• For iteration $j$, $G_j = G_{j-1} \otimes G_{j-1}$ (Applying theorem 2 to $G_{j-1}$ to get $G_j$)
• Gap = $(1 + \epsilon)^{2^j}$
• Size grows like $n^{c^2j}$
• Gap grows as $O(\log |G_j|)^{c_1}$ where $c_1$ is a constant < 1!
OPEN PROBLEMS

• Proving/ Disproving the Li and Li conjecture
  • Even for linear codes?
• Computing optimal network coding for directed graphs
• \( n \) Random variables \( X_1, X_2, \ldots, X_n \) (joint distribution)

• Entropic vector is \( 2^n - 1 \) dimensional vector with \( S^{th} \) coordinate holding \( H(X_{i_1}, X_{i_2}, \ldots, X_{i_k}) \) where \( S = \{i_1, i_2, \ldots, i_k\} \subseteq [n] \)

• \( n = 2, \ [H(X_1), H(X_2), H(X_1X_2)] \)
• $n$ Random variables $X_1, X_2, ..., X_n$ (joint distribution)

• Entropic vector is $2^n - 1$ dimensional vector with $S^{th}$ coordinate holding $H(X_{i_1}, X_{i_2}, ..., X_{i_k})$ where $S = \{i_1, i_2, ..., i_k\} \subseteq [n]$

• $n = 2$, $[H(X_1), H(X_2), H(X_1X_2)]$

• Entropic region is the set of all such vectors
  • $H(X_1) \leq H(X_1X_2) \leq H(X_1) + H(X_2)$
• n Random variables $X_1, X_2, \ldots, X_n$ (joint distribution)

• Entropic vector is $2^n - 1$ dimensional vector with $S^{th}$ coordinate holding $H(X_{i_1}, X_{i_2}, \ldots, X_{i_k})$ where $S = \{i_1, i_2, \ldots, i_k\} \subseteq [n]$

• $n = 2, \ [H(X_1), H(X_2), H(X_1X_2)]$

• Entropic region is the set of all such vectors

  • $H(X_1) \leq H(X_1X_2) \leq H(X_1) + H(X_2)$

• Given a $2^n - 1$ dimensional vector, is it $\epsilon$-close to a vector in entropic region? (Even decidability)
THANKS