# Expected Subgraph Capacity for Graphs (Draft)

Chandradeep Chowdhury
Patrick Perrine

November 26, 2023

In this article, we explore the notion of *interference* of nodes shared between induced subgraphs in a basic sense. We define and generalize our framework as a result of the hypergeometric distribution, and then arrive at results regarding the *capacity* of graphs based on those notions of interference.

## 1 Basic Interference Framework

**Definition 1.** (**Interference**) Given two sets $U, W$, we say $U$ *interferes* with $W$ if

$$|U \cap W| \geq |W|. \tag{1}$$

**Lemma 2.** *Given a vertex set $V$ with $n$ vertices and two subsets $U, W$ of respective sizes $r_u, r_w$, the probability of an interference between $U$ and $W$, denoted as $Y$, follows a hypergeometric distribution with $Y \sim Hypergeometric(n, r_u, y)$.*

*Proof.* If $V = \{v_1, ..., v_n\}$, we can represent a subset $U$ as a vector $u$ of length $n$ defined by

$$u_i = \begin{cases} 1 & \text{if } v_i \in U \\ 0 & \text{if } v_i \notin U. \end{cases}$$

With this representation, $U, W$ intersect at the indices where both vectors $u, w$ have a 1. Let $Y$ be a discrete random variable denoting the number of indices where both $u, w$ have a 1. Then

$$\mathbb{P}(Y = y) = \frac{\binom{r_u}{y}\binom{n-r_u}{r_w-y}}{\binom{n}{r_w}} \tag{2}$$

This follows from the fact that given the first array $U$, we already know where the 1's are located. We can pick the $y$ intersecting 1's for the second array in $\binom{r_u}{y}$ ways implicitly placing 0's in the remaining spots. We then fill the remaining $n - r_u$ indices corresponding to the 0's in the first array with $r_w - y$ 1's in $\binom{n-r_u}{r_w-y}$ ways. Finally we divide by the total number of possible subgraphs $\binom{n}{r_w}$.

We then use Vandermonde's identity to realize this probability as a result of the hypergeometric distribution, with $Y \sim Hypergeometric(n, r_u, y)$. □

## 2 Generalizing Interference for $k$ Instances

**Definition 3.** (***k*-Interference**) Given two sets $U, W$, and for some $k \in (0, |W|]$, we say $U$ $k-$*interferes* with $W$ if

$$|U \cap W| \geq \frac{|W|}{k}. \tag{3}$$

**Corollary 4.** *If $|U| = |W|$, then $U$ $k$-interferes with $W$ if and only if $W$ $k$-interferes with $U$.*

We restrict the upper range of $k$ to $|W|$ for convenience, as beyond that all values of $\frac{|W|}{k}$ will be less than 1. If $\frac{|W|}{k} = 1$, we could arrive at a variation of the Hitting Set problem.

**Lemma 5.** *Given a vertex set $V$ with $n$ vertices and two subsets $U, W$ of respective sizes $r_u, r_w$, the probability that $U$ $k$-interferes with $W$, denoted by $Y$, is the tail distribution or survival function of $Y$ at $\lceil \frac{r_w}{k} \rceil$, or simply $\bar{F}_Y\left(\lceil \frac{r_w}{k} \rceil\right)$.*

*Proof.* The probability that $U$ $k$-interferes with $W$ is

$$\sum_{y=\lceil \frac{r_w}{k} \rceil}^{r_w} \mathbb{P}(Y=y) = 1 - \sum_{y=0}^{\lceil \frac{r_w}{k} \rceil} \mathbb{P}(Y=y) = 1 - \mathbb{P}\left(Y \leq \left\lceil \frac{r_w}{k} \right\rceil\right) = \mathbb{P}\left(Y > \left\lceil \frac{r_w}{k} \right\rceil\right) = \bar{F}_Y\left(\left\lceil \frac{r_w}{k} \right\rceil\right). \tag{4}$$

□

**Lemma 6.** *If $r_w$ is drawn from a distribution with mean $r$, then the expected value of the tail distribution of $Y \sim Hypergeometric(n, r_u, y)$, $\bar{F}_Y\left(\lceil \frac{r_w}{k} \rceil\right)$ is*

*placeholder*

*Proof.* Observe that

$$\begin{aligned}
\mathbb{E}\left(\bar{F}_Y\left(\left\lceil \frac{r_w}{k} \right\rceil\right)\right) &= \mathbb{E}\left(\sum_{y=\lceil \frac{r_w}{k} \rceil}^{r_w} \mathbb{P}(Y=y)\right) \\
&= \mathbb{E}\left(\sum_{y=\lceil \frac{r_w}{k} \rceil}^{r_w} \frac{\binom{r_u}{y}\binom{n-r_u}{r_w-y}}{\binom{n}{r_w}}\right) \\
&= \sum_{y=\lceil \frac{r_w}{k} \rceil}^{r_w} \mathbb{E}\left(\frac{\binom{r_u}{y}\binom{n-r_u}{r_w-y}}{\binom{n}{r_w}}\right) \\
&\geq \sum_{y=\lceil \frac{r_w}{k} \rceil}^{r_w} \mathbb{E}\left(\frac{\left(\frac{r_u}{y}\right)^y \left(\frac{(n-r_u)}{r_w-y}\right)^{r_w-y}}{\left(\frac{n}{r_w}\right)^{r_w}}\right) \\
&= \sum_{y=\lceil \frac{r_w}{k} \rceil}^{r_w} \frac{\left(\frac{r}{y}\right)^y \left(\frac{(n-r)}{r-y}\right)^{r_w-y}}{\left(\frac{n}{r}\right)^{r_w}}
\end{aligned} \tag{5}$$

□

# 3 Capacity Results

**Definition 7.** (**$(r, T, k)$-Subgraph Capacity**) Given a vertex set $V = \{v_1, ..., v_n\}$, the $(r, T, k)$-*subgraph capacity* of $V$ is the *expected maximum* number of subgraphs of size $r$ that can be collected subject to the constraint that for any randomly picked subgraph $U$,

$$\mathbb{E}[X] \leq T \tag{6}$$

where $X$ is the number of interferences caused due to picking $U$.

## 3.1 Capacity for Exactly $r$-sized Subgraphs

**Theorem 8.** *Given a vertex set $V$ with $n$ vertices and the propery that every generated subgraph will have size exactly $r$, the $(r, T, k)$-subgraph capacity of $V$ is*

$$\left\lfloor \frac{T}{\bar{F}_Y\left(\lceil \frac{r}{k} \rceil\right)} + 1 \right\rfloor.$$

*Proof.* Suppose we have $M$ subgraphs in the collection. Pick an arbitrary subgraph $U$. From lemma 5, we know that the probability of $U$ $k$-interfering with another subgraph is $\bar{F}_Y\left(\lceil \frac{r}{k} \rceil\right)$. Since there are $M-1$ other subgraphs, the expected number of $k$-interferences caused by picking $U$ is $(M-1)\bar{F}_Y\left(\lceil \frac{r}{k} \rceil\right)$.

From inequality 6, we have

$$(M-1)\bar{F}_Y\left(\left\lceil \frac{r}{k} \right\rceil\right) \leq T \implies M \leq \frac{T}{\bar{F}_Y\left(\lceil \frac{r}{k} \rceil\right)} + 1. \tag{7}$$

The $(r, T, k)$-subgraph capacity of $V$ is the largest integer $M$ that satisfies inequality 7. □

*Alternate proof.* Suppose we have $M$ subgraphs in the collection. Pick two subgraphs $U, W$. From lemma 5, we know that the probability of $U$ $k$-interfering with another subgraph is $\bar{F}_Y\left(\lceil \frac{r}{k} \rceil\right)$. Since all subgraphs have the same size, by corollary 4 this becomes the probability that $U, W$ pair will cause exactly 2 $k$-interferences. So the expected number of interferences caused by one pair is

$$2\bar{F}_Y\left(\left\lceil \frac{r}{k} \right\rceil\right).$$

We know that there are $\binom{M}{2} = M(M-1)/2$ such pairings so the expected number of total interferences is

$$2 \cdot \frac{M(M-1)}{2}\bar{F}_Y\left(\left\lceil \frac{r}{k} \right\rceil\right) = M(M-1)\bar{F}_Y\left(\left\lceil \frac{r}{k} \right\rceil\right).$$

Since there are $M$ subgraphs, the expected number of interferences by picking one subgraph is

$$\frac{M(M-1)}{M}\bar{F}_Y\left(\left\lceil \frac{r}{k} \right\rceil\right) = (M-1)\bar{F}_Y\left(\left\lceil \frac{r}{k} \right\rceil\right).$$

From inequality 6, we have

$$(M-1)\bar{F}_Y\left(\left\lceil \frac{r}{k} \right\rceil\right) \leq T \implies M \leq \frac{T}{\bar{F}_Y\left(\lceil \frac{r}{k} \rceil\right)} + 1.$$

The $(r, T, k)$-subgraph capacity of $V$ is the largest integer $M$ that satisfies inequality 7. □

## 3.2 Capacity for Expected $r$-sized Subgraphs

**Theorem 9.** *Given a vertex set $V$ with $n$ vertices, the $(\mathbb{E}[r], T, k)$-subgraph capacity of $V$ is*

$$\left\lceil \frac{1}{T} \sum_{(i,j)\in\mathbb{Z}\times\mathbb{Z}, 1\leq i,j\leq M, i\neq j} \left(\bar{F}_Y\left(\left\lceil \frac{r_j}{k} \right\rceil\right) + \bar{F}_Y\left(\left\lceil \frac{r_i}{k} \right\rceil\right)\right)\right\rceil.$$

*Proof.* Suppose we have $M$ subgraphs $U_1, ..., U_M$ with sizes $r_1, ..., r_M$. Pick two subgraphs $U_i, U_j$. From lemma 5, we know that the expected number of interferences caused by this pair is

$$\bar{F}_Y\left(\left\lceil \frac{r_j}{k} \right\rceil\right) + \bar{F}_Y\left(\left\lceil \frac{r_i}{k} \right\rceil\right)$$

We then sum over all possible pairings to get the expected number of total interferences:

$$\sum_{(i,j)\in\mathbb{Z}\times\mathbb{Z}, 1\leq i,j\leq M, i\neq j} \left(\bar{F}_Y\left(\left\lceil \frac{r_j}{k} \right\rceil\right) + \bar{F}_Y\left(\left\lceil \frac{r_i}{k} \right\rceil\right)\right)$$

Since there are $M$ subgraphs, the expected number of interferences by picking one subgraph is

$$\frac{1}{M} \sum_{(i,j)\in\mathbb{Z}\times\mathbb{Z}, 1\leq i,j\leq M, i\neq j} \left(\bar{F}_Y\left(\left\lceil \frac{r_j}{k} \right\rceil\right) + \bar{F}_Y\left(\left\lceil \frac{r_i}{k} \right\rceil\right)\right).$$

From inequality 6, we have

$$\frac{1}{M} \sum_{(i,j)\in\mathbb{Z}\times\mathbb{Z}, 1\leq i,j\leq M, i\neq j} \left(\bar{F}_Y\left(\left\lceil \frac{r_j}{k} \right\rceil\right) + \bar{F}_Y\left(\left\lceil \frac{r_i}{k} \right\rceil\right)\right) \leq T$$

which implies

$$M \geq \frac{1}{T} \sum_{(i,j)\in\mathbb{Z}\times\mathbb{Z}, 1\leq i,j\leq M, i\neq j} \left(\bar{F}_Y\left(\left\lceil \frac{r_j}{k} \right\rceil\right) + \bar{F}_Y\left(\left\lceil \frac{r_i}{k} \right\rceil\right)\right). \tag{8}$$

The $(\mathbb{E}[r], T, k)$-subgraph capacity of $V$ is the smallest integer $M$ that satisfies inequality 8.

□

# 4 Conclusion

We intend for these results to be useful in domains where induced subgraphs represent items of information. One example of an application could be that of the Neuroidal Model by Leslie Valiant, as induced subgraphs are used as memories of a neural system within a computational neuroscience perspective.