Finding Cliques in Networks: Supplementary

Cassandra Overney and Emma Pan

November 2019

1 Maximum Number of Cliques Proof

We start by proving the lemma:

$\binom{n}{l}x^l \leq \binom{x}{l}n^l$ *for all integers* $x \geq n \geq l \geq 0$.

Proof: Let $P(l)$ be the statement "$\binom{n}{l}x^l \leq \binom{x}{l}n^l$ for all integers $x \geq n \geq l \geq 0$." We want to show $P(l)$ is true for all $l \geq 0$. First we show that $P(0)$ is true. On the left side of the inequality, $\binom{n}{0}x^0 = 1 \times 1 = 1$. The right side of the inequality simplifies to $\binom{x}{0}n^0 = 1 \times 1 = 1$. $1 \leq 1$, so $P(0)$ is true.

For the inductive step, we assume $P(k)$ is true for some $k \geq 0$. We want to show that $P(k+1)$ is true. Since $k \geq 0$, if we add 1 to both sides, we get $k+1 \geq 1$. We can multiply both sides of the inequality by $x-n$ to get $(k+1)(x-n) \geq x-n$. We can then add xn to both sides, resulting in $xn + (k+1)(x-n) \geq xn + x - n$. We can simplify the inequality by performing the following arithmetic operations

$$\begin{aligned}
xn + (k+1)(x-n) &\geq xn + x - n && \text{from above} \\
xn + (k+1)x - (k+1)n &\geq xn + x - n && \text{distribute k+1} \\
xn - (k+1)n + n &\geq xn + x - (k+1)x && \text{move n and (k+1)x to other side} \\
n(x - (k+1) + 1) &\geq x(n + 1 - (k+1)) && \text{factor out n and x}
\end{aligned}$$

Since $P(k)$ is true, we know $\binom{n}{k}x^k \leq \binom{x}{k}n^k$, so we can take this inequality and merge it with the simplified inequality from above by multiplying the respective greater and smaller sides together. We would get

$$\begin{aligned}
\binom{x}{k}n^k n(x-(k+1)+1) &\geq \binom{n}{k}x^k x(n+1-(k+1)) \\
\frac{x!n^{k+1}(x+1-(k+1))}{(x-k)!k!} &\geq \frac{n!x^{k+1}(n+1-(k+1))}{(n-k)!k!} && \text{write out combinatorics} \\
\frac{x!n^{k+1}(x-k)}{(x-k)!k!} &\geq \frac{n!x^{k+1}(n-k)}{(n-k)!k!} && \text{simple arithmetic} \\
\frac{x!n^{k+1}}{(x-(k+1))!k!} &\geq \frac{n!x^{k+1}}{(n-(k+1))!k!} && \text{simplify factorial} \\
\frac{x!n^{k+1}}{(x-(k+1))!(k+1)!} &\geq \frac{n!x^{k+1}}{(n-(k+1))!(k+1)!} && \text{divide both sides by k+1} \\
\binom{x}{k+1}n^{k+1} &\geq \binom{n}{k+1}x^{k+1} && \text{convert back to combinatorics}
\end{aligned}$$

Since we were able to obtain $\binom{x}{k+1}n^{k+1} \geq \binom{n}{k+1}x^{k+1}$, $P(k+1)$ is true. In summary we showed $P(0)$ is true. We also showed that if $P(k)$ is true for some $k \geq 0$ then $P(k+1)$ is true. Thus, $P(l)$ is true for all $l \geq 0$.

We will use $\binom{n}{l}x^l \leq \binom{x}{l}n^l$ *for all integers* $x \geq n \geq l \geq 0$ to prove the following theorem:

For all integers $x \geq l \geq 0$, every graph G with $n \geq l$ vertices and no $(x+1)-$clique has at most $\binom{x}{l}\left(\frac{n}{x}\right)^l$ l-cliques.

Let $P(n)$ be the statement "for all $x \geq l \geq 0$ a graph G with $n \geq l$ vertices and no (x+1)-cliques has at most $\binom{x}{l}\left(\frac{n}{x}\right)^l$ l-cliques." We want to show $P(n)$ is true for all $n \leq x$ and $n > x$.

For our base case, we will show that $P(n)$ is true when $n \leq x$. This is the same as showing $P(0)...P(x)$ is true for all $x \geq 0$. Since the number of vertices is less than than or equal to x, it is not possible to form a (x+1)-clique because there aren't enough vertices. As a result, we know that G would have no (x+1)-cliques, so we can focus on finding the number of l-cliques. We define $c_l(G)$ to be the number of l-cliques in graph G. $c_l(G) \leq \binom{n}{l}$ because $\binom{n}{l}$ gives all possible groups of l vertices from G. Only a subset of them will actually be complete subgraphs. From the lemma we proved above, $\binom{n}{l}x^l \leq \binom{x}{l}n^l$. We can divide both sides of the inequality by x^l to get $\binom{n}{l} \leq \binom{x}{l}\left(\frac{n}{x}\right)^l$. Since $c_l(G) \leq \binom{n}{l}$, we also know that $c_l(G) \leq \binom{x}{l}\left(\frac{n}{x}\right)^l$. Thus, the number of l-cliques in G is $\leq \binom{x}{l}\left(\frac{n}{x}\right)^l$ and $P(n)$ is true when $n \leq x$.

For the inductive step where we have to consider if $n > k$, we assume that $P(0)...P(k)$ is true for some $k \geq 0$. We want to show that $P(k+1)$ is true. Let G be a graph with k+1 vertices, no (x+1)-cliques, and a maximum $c_l(G)$ value. (We care about the maximum because we are trying to prove a maximal bound on the number of l-cliques.) We can add edges to G until it contains an x-clique, X. Every l-clique in G is the union of some i-clique of $G \backslash X$ and some $(l-i)$-clique of $G[X]$ for $0 \leq i \leq l$. $G \backslash X$ gives the subgraph of G that doesn't contain the vertices and edges in X. $G[X]$ is the complement of $G \backslash X$. In other words, we can break up every l-clique into two parts, one in $G \backslash X$ and one in $G[X]$. The vertices in each i-clique of $G \backslash X$ has at most $x-i$ common neighbors in X because X is a clique and G has no $(x+1)$-clique. If a vertex in an i-clique had more than

$x-i$ common neighbors in X, then a clique with $\geq x+1$ vertices can be formed with the i-clique and its neighbors in X.

Thus, for each i-clique of $G\backslash X$, we can pick $l-i$ vertices from its $x-i$ common neighbors in $\binom{x-i}{l-i}$ ways to form l-cliques of G. Based on our assumption that $P(0)...P(k)$ is true, the number of i-cliques in $G\backslash X$ with smaller than $k+1$ vertices is $\leq \binom{x}{i}\left(\frac{k+1-x}{x}\right)^i$ where $k+1-x$ gives us the number of vertices in $G\backslash X$, which is $\leq k$. By the product rule, we have at most $\binom{x}{i}\left(\frac{k+1-x}{x}\right)^i\binom{x-i}{l-i}$ l-cliques for a given i. We can use the sum rule to get the total number of l-cliques from $0 \leq i \leq l$, resulting in $c_l(G) \leq \sum_{i=0}^{l}\binom{x}{i}\left(\frac{k+1-x}{x}\right)^i\binom{x-i}{l-i}$. We can simplify this expression with the following operations

$$
\begin{aligned}
c_l(G) &\leq \sum_{i=0}^{l}\binom{x}{i}\left(\frac{k+1-x}{x}\right)^i\binom{x-i}{l-i} && \text{from above}\\
c_l(G) &\leq \sum_{i=0}^{l}\frac{x!}{(x-i)!\,i!}\,\frac{(x-i)!}{(x-l)!(l-i)!}\left(\frac{k+1-x}{x}\right)^i && \text{write out combinatorics}\\
c_l(G) &\leq \sum_{i=0}^{l}\frac{x!}{i!}\,\frac{1}{(x-l)!(l-i)!}\,\frac{l!}{l!}\left(\frac{k+1-x}{x}\right)^i && \text{multiply by } \frac{l!}{l!}\\
c_l(G) &\leq \sum_{i=0}^{l}\frac{x!}{(x-l)!\,l!}\,\frac{l!}{(l-i)!\,i!}\left(\frac{k+1-x}{x}\right)^i && \text{rearrange terms}\\
c_l(G) &\leq \sum_{i=0}^{l}\binom{x}{l}\binom{l}{i}\left(\frac{k+1-x}{x}\right)^i && \text{convert back to combinatorics}\\
c_l(G) &\leq \binom{x}{l}\sum_{i=0}^{l}\binom{l}{i}\left(\frac{k+1}{x}-1\right)^i && \text{simple arithmetic}\\
c_l(G) &\leq \binom{x}{l}\left(\frac{k+1}{x}\right)^l && \text{binomial theorem}
\end{aligned}
$$

The last simplification uses the binomial theorem, which is $a^t = \sum_{j=0}^{t}\binom{t}{j}(a-1)^j$ with $a = \frac{k+1}{x}$ and $t = l$. Since $c_l(G) \leq \binom{x}{l}\left(\frac{k+1}{x}\right)^l$, $P(k+1)$ is true. In summary, we showed that $P(n)$ is true if $n \leq x$ (base case). We also showed that if $P(0)...P(k)$ is true for some $k \geq 0$, then $P(k+1)$ is true. Thus, for all $x \geq l \geq 0$, a graph G with $n \geq l$ vertices and no (x+1)-cliques has at most $\binom{x}{l}\left(\frac{n}{x}\right)^l$ l-cliques.

2 Relaxed Cliques

Since cliques are complete subgraphs, they are often overly restrictive for network analysis, especially in community detection. As a result, different variants of relaxed cliques can be defined in terms of more flexible vertex degree, distance, edge density, and connectivity characteristics. Relaxed cliques allow us to create application-specific constraints for network analysis. In this section, we will explain 8 different types of relaxed cliques.

2.1 Mathematical Notation

Here is a list of mathematical notation that we will reference throughout our explanations.

- G is a graph with a vertex set V and an edge set E
- S is a subset of V
- $G[S]$ is a subgraph of G
- $E(S)$ is the set of edges in G with both endpoints in S

2.2 Relaxing the Degree Constraint

The degree of vertex i in G is denoted as $\deg_G(i)$. The *minimum vertex degree of* G is $\delta(G)$. We can relax the number of degrees required for each vertex in $G[S]$ to get *k-cores* and *s-plexes*.

k-core. For $k \geq 0$, $G[S]$ is a *k-core* if $\delta(G[S]) \geq k$. In other words the minimum degree in the subgraph is at least k. Each vertex is connected to at least k other vertices, and can have at most $k-1$ non-neighbors inside S. In standard, unrelaxed cliques, the minimum degree in $G[S]$ is the number of vertices in S subtracted by 1 because every vertex is connected to every other vertex in the subgraph, so $k = |S| - 1$. *k-cores* are designed to capture cohesive subgroups, as well as the regions surrounding them.

s-plex. For $s \geq 1$, $G[S]$ is a *s-plex* if $\delta(G[S]) \geq |S| - s$, which means that the minimum degree in the subgraph is at least the total number of vertices in $G[S]$ subtracted by a constant, s. Every s vertices form a *dominating set*. For standard cliques, $\delta(G[S]) = |S| - 1$, so $s = 1$. Low *s-plex* cliques, $s = 1, 2, 3$, are used to closely resemble the cohesive subgroups that can be found in social networks. They also retain other desirable properties of a clique such as low diameter (reachability) and high connectivity (robustness).

2.3 Relaxing the Distance Constraint

For two vertices, $i, j, \in V$, the minimum distance between i and j is the minimum length of a path from i to j in G. The minimum distance is defined as $\mathrm{dist}_G(i, j)$. We can increase the distances between vertices in $G[S]$ to get *s-cliques* and *s-clubs*.

s-clique. For $s \geq 1$, $G[S]$ is a *s-clique* if $\mathrm{dist}_G(i, j) \leq s$ for all $i, j \in S$. In other words, every pair of vertices in S are at most s distance away from each other. For standard cliques, every pair of vertices in S are connected so the minimum distances are all 1, so $\mathrm{dist}_G(i, j) \leq 1$ for all $i, j \in S$ and $s = 1$. *s-cliques* are good to use when retaining reachability.

s-club. For $s \geq 1$, $G[S]$ is a *s-club* if $\mathrm{dist}_{G[S]}(i, j) \leq s$ for all $i, j \in S$. In other words, the minimum distance between every pair of vertices in S when only considering the edges in $G[S]$ does not exceed s. We can view S as a subset of vertices that induces a subgraph with a diameter of at most s. Two disconnected vertices have a minimum distance of ∞. Similar to *s-cliques*, standard cliques have $\mathrm{dist}_{G[S]}(i, j) \leq 1$ for all $i, j \in S$ and $s = 1$. *s-clubs* are also good to use for reachability in social networks and chain reactions.

2.4 Relaxing the Density Constraint

To begin, we define the *edge density* of a subgraph $G[S]$ as $\rho(G[S]) = |E(S)|/\binom{|S|}{2}$. In other words, edge density gives the ratio of edges in the subgraph to the total possible number of edges in $G[S]$, $\binom{|S|}{2}$. In cliques, $\rho(G[S]) = 1$ because cliques are complete subgraphs. We can relax the density constraint to obtain *ρ-quasi-cliques* and *s-defective cliques*.

ρ-quasi-clique. For $0 \leq \rho \leq 1$, $G[S]$ is a *ρ-quasi-clique* if $\rho(G[S]) \geq \gamma$. If ρ is less than 1, then the relaxed clique has less edges compared to the standard clique with the same number of vertices. *ρ-quasi-cliques* provide a relative measure for existing/ missing edges, which is good for robustness.

s-defective clique. For $s \geq 0$, $G[S]$ is a *s-defective clique* if $|E(G[S])| \geq \binom{|S|}{2} - s$. Here we are comparing the number of edges in $G[S]$ to the total number of possible edges subtracted by a constant s. In a standard clique, $s = 0$ since $|E(G[S])| = \binom{|S|}{2}$. If $s > 0$, then we have more flexibility in finding subgraphs with a lower number of edges. *s-defective cliques* are used to identify massive protein interaction networks from large-scale experiments.

2.5 Relaxing the Connectivity Constraint

We first define what connectivity means. A subset of vertices C is a *vertex cut* of G if $G[V \setminus C]$ is a disconnected graph. $G[V \setminus C]$ refers to the subgraph of G discluding C. Any vertex cut C has at most $|V| - 2$ elements because if we take away $|V| - 1$ or $|V|$ vertices in G, then we would end up with a graph with 1 or 0 vertices, which would not qualify as a disconnected graph. The *vertex connectivity* $\kappa(G)$ is the size of the minimum vertex cut. Both *k-blocks* and *s-bundles* are defined based on connectivity, which are required in social networks.

k-block. For $k \geq 1$, $G[S]$ is a *k-block* if $\kappa(G[S]) \geq k$. For cliques, $G[S]$ has no vertex cuts because every vertex is connected to every other vertex, so no disconnected graphs could form by removing any set of vertices. In other words $\kappa(G[S]) = |S| - 1$, which is one more than the maximum size of C. To satisfy the inequality, $k = |S| - 1$.

s-bundle. For $s \geq 1$, $G[S]$ is a *s-bundle* if $\kappa(G[S]) \geq |S| - s$. In other words the size of the minimum vertex cut is greater than or equal to the number of vertices in the subgraph subtracted by a constant, s. The closer s is to 1, the more connectivity $G[S]$ has. An original clique has $s = 1$ because $\kappa(G[S]) = |S| - 1$.

2.6 Illustrated Examples of Relaxed Cliques

The minimum degree of the graph in figure 1 is 2, which means that it is a 2-core. Also, $|S| = 6$, so if $s = 4$, then the minimum degree would be ≥ 2, which also makes the graph a 4-plex.

The minimum distances between vertices 1,2 1,3 and 3,4 are all 1. The minimum distance between 2,3 is 2, and the minimum distance between 2,4 is 3. This means that the minimum distance between any two vertices in figure 2 is at least 3, making it a 3-club and 3-clique. Here we are assuming $G[S]$ has the same minimum distances as G.

The edge density of figure 3 is 0.5 because $3/6 = 0.5$. There are 3 edges and 6 possible edges in a graph with 4 vertices. An edge density of 0.5 results in a 0.5-quasi-clique. Additionally, since

$\binom{|S|}{2} = 6$ and $|E(G[S])| = 3$, $6 - 3 = 3$, so the graph is also a 3-defective clique.

The vertex connectivity of figure 4 is 1. If we remove vertex 1 from the graph, then we will obtain a disconnected graph since there will no longer be a path between vertex 2 and vertices 3 and 4. The size of the vertex cut is 1, so figure 4 is a 1-block. Since $|S| = 4$ and $4 - s = 1$, we get $s = 3$, which corresponds to a 3-bundle.

Table 1 gives a summary of the different types of relaxed cliques we covered above.

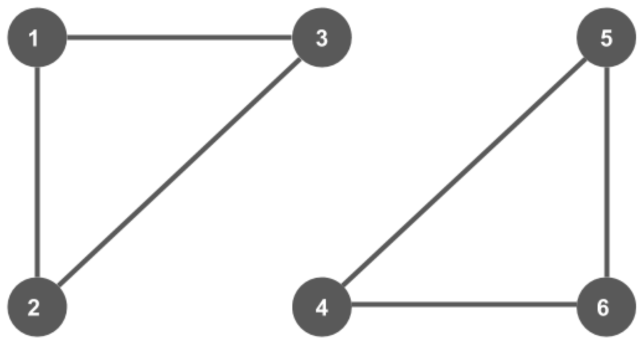

Figure 1: 2-core and 4-plex

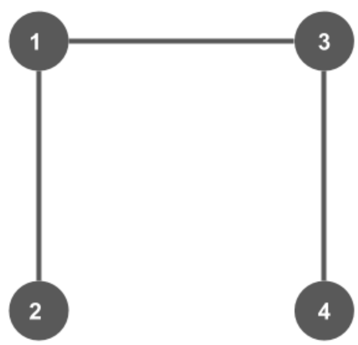

Figure 2: 3-club and 3-clique

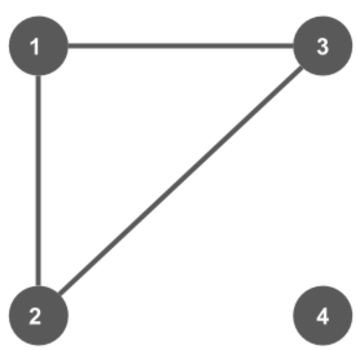

Figure 3: 0.5-quasi-clique and 3-defective clique

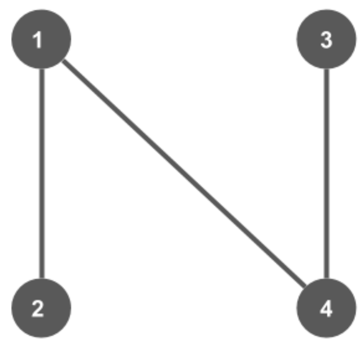

Figure 4: 1-block and 3-bundle

Table 1: Summary of different clique relaxations

Name	**Definition**	**Based on**	**Clique**
k-core	$\delta(G[S]) \geq k$	Degree	$k = \|S\| - 1$
s-plex	$\delta(G[S]) \geq \|S\| - s$	Degree	$s = 1$
s-clique	$\text{dist}_G(i, j) \leq s$ for all $i, j \in S$	Distance	$s = 1$
s-club	$\text{dist}_{G[S]}(i, j) \leq s$ for all $i, j \in S$	Distance	$s = 1$
y-quasi-clique	$\rho(G[S]) \geq \gamma$	Density	$\rho = 1$
s-defective clique	$\|E(G[S])\| \geq \binom{\|S\|}{2} - s$	Density	$s = 0$
k-block	$\kappa(G[S]) \geq k$	Connectivity	$k = \|S\| - 1$
s-bundle	$\kappa(G[S]) \geq \|S\| - s$	Connectivity	$s = 1$