DISTRIBUTION OF MAXIMAL CLIQUE SIZE UNDER THE WATTS-STROGATZ MODEL OF EVOLUTION OF COMPLEX NETWORKS

Natarajan Meghanathan

Jackson State University, 1400 Lynch St, Jackson, MS, USA

ABSTRACT

In this paper, we analyze the evolution of a small-world network and its subsequent transformation to a random network using the idea of link rewiring under the well-known Watts-Strogatz model for complex networks. Every link u-v in the regular network is considered for rewiring with a certain probability and if chosen for rewiring, the link u-v is removed from the network and the node u is connected to a randomly chosen node w (other than nodes u and v). Our objective in this paper is to analyze the distribution of the maximal clique size per node by varying the probability of link rewiring and the degree per node (number of links incident on a node) in the initial regular network. For a given probability of rewiring and initial number of links per node, we observe the distribution of the maximal clique per node to follow a Poisson distribution. We also observe the maximal clique size per node in the small-world network to be very close to that of the average value and close to that of the maximal clique size in a regular network. There is no appreciable decrease in the maximal clique size per node when the network transforms from a regular network to a small-world network. On the other hand, when the network transforms from a small-world network to a random network, the average maximal clique size value decreases significantly.

KEYWORDS


1. INTRODUCTION

Network Science is the field of analyzing complex real-world networks from a graph theoretical standpoint. A complex network is abstracted as a graph wherein the nodes and links in the network are modeled respectively as the vertices and edges of the graph. The graph theoretic algorithms run on such complex network graphs (to determine one or more metrics characteristic of the networks) need to be as efficient as possible. The graph theoretic metrics considered for analysis of such complex networks [1] include centrality, clustering coefficient, diameter, clique size, etc. Unlike the polynomial-time efficient algorithms that exist to determine most of the above metrics, the maximum clique size problem is NP-hard [2]. A "clique" is a complete sub graph of a graph such that any two vertices in the sub graph are connected with an edge. Community detection algorithms (e.g., [3-5]) use cliques of various sizes as the basis to determine closely-knit and overlapping communities in complex real-world networks as well as in networks that are simulated using theoretical models.

For an n-vertex graph, the "maximum size clique" is the clique of the largest size k such that the graph does not have any clique of size k+1 (where k ≤ n). Note that there may exist one or more cliques of smaller size in the graph and for a particular vertex i, the largest size clique it is part of need not be the maximum size clique for the entire network graph. Accordingly, we define the
"maximal size clique for a vertex \( i \)" as the largest size clique the vertex is part of. The focus of research in this area has been so far on developing efficient exact algorithms and heuristics (time and space-wise) to determine respectively the maximum size cliques and approximations to the same. Not much work has been conducted on determining the maximal size cliques for the individual vertices of the graph. Specifically, we could not come across any work that has analyzed the distribution of the maximal clique size for the individual vertices of graphs that simulate the evolution of complex networks. In this pursuit, we choose a recently proposed branch-and-bound strategy based efficient exact algorithm [6] to determine maximum size clique for an entire network graph and modify it to determine the maximal size cliques for the constituent vertices of the graph. We use the modified exact algorithm to analyze the distribution of the maximal clique size of the network graphs that span two categories of complex networks: small-world networks and random networks. We choose the well-known Watts-Strogatz model to simulate the evolution of small-world networks (from a regular network) and their subsequent transformation to a random network.

The rest of the paper is organized as follows: In Section 2, we first review and discuss the recently proposed exact algorithm to determine maximum clique size for an entire graph and then explain our modifications to the same to determine the maximal clique for a particular vertex of the graph. Section 3 presents the evolution of small-world networks and their transformation to a random network under the Watts-Strogatz model and describes the results of the diameter and clustering coefficient of the networks that evolve during this transformation. Section 4 presents in detail the distribution of the maximal clique size per node and the average maximal clique size as a function of the probability of link rewiring and the initial number of links per node. Section 5 reviews the literature and discusses related work. Section 6 concludes the paper. In the entire paper, the terms 'vertex' and 'node', 'edge' and 'link' have been used interchangeably. They mean the same.

2. CLIQUE

A clique is a subset of the vertices of a graph such that any two vertices in the subset are connected with an edge. The maximal clique size and maximum clique size problems are both NP-hard [2]. There are several exact algorithms (that take exponential run-time at the worst case) have been proposed to determine the maximum clique size for sparse graphs. With the surge of research interests in the analysis of complex real-world networks from a graph theoretic standpoint, we come across efficient algorithms (e.g., [6-9]) to extract the largest size cliques in large/dense graphs. The "branch-and-bound" strategy is a commonly used strategy [10] behind these algorithms and the idea is to explore the neighborhood of only those vertices that have scope for being part of a clique that can be larger than the clique known until then during the execution of the algorithm. The variation among the various exact algorithms is the pruning strategy adopted to decide on the bounds for branching through the solution space and limiting the search. In this section, we will first review a recently proposed exact branch-and-bound based exact algorithm to determine maximum size clique for an entire network graph and then explain our modifications to it to determine the maximal size cliques for the individual vertices of a graph.

The pseudo code of the exact algorithm (proposed in [6]) to determine the maximum size clique for a graph is outlined in Figure 1. The idea behind the algorithm is to keep track of the maximum size clique found so far (variable \( \text{max} \) in the pseudo code, the initial value is 0) and update it whenever an even larger clique is found. The algorithm proceeds in iterations, with iteration \( i \) attempting to determine whether vertex \( v_i \) could be part of a clique that is larger than the maximum size clique known until then. The algorithm decides to explore the neighborhood of a vertex only if the degree of the vertex is larger than the currently known maximum size clique;
moreover only the neighbors whose degree is at least the size of the currently known maximum clique size are considered for further exploration (these neighbors form the set $U$ passed to the subroutine CLIQUE). The vertices are explored in the increasing order of their IDs (one iteration per vertex).

---

**Subroutine MAXCLIQUE**

```plaintext
Procedure MAXCLIQUE ($G = (V, E)$)
max ← 0
for $i : 1$ to $|V|$ do 
  if $\text{degree}(v_i) \geq \text{max}$ then 
    $U ← \emptyset$
    for each $v_j \in \text{Neighbor}(v_i)$ do 
      if $\text{degree}(v_j) \geq \text{max}$ then 
        $U ← U \cup \{v_j\}$
    CLIQUE($G, U, 1$)
  end if
end for
end if
while $U \neq \emptyset$ do
  if $\text{size} + |U| \leq \text{max}$ then
    return
  end if
  select any vertex $u$ from $U$
  $U ← U \setminus \{u\}$
  $N(u) := \{w \mid w \in \text{Neighbor}(u) \wedge \text{degree}(w) \geq \text{max}\}$
  CLIQUE($G, U \cap N(u), \text{size} + 1$)
end while
```

---

Figure 1. Exact Algorithm to Extract the Maximum Size Clique in a Graph (adapted from [6])

The sub routine CLIQUE (called for a particular vertex $v_i$ and its set of neighbors $U$ whose degree is at least the size of the maximum clique, $\text{max}$, found so far) runs via a combination of iterations and recursions. In each iteration, the sub routine randomly removes a node $u$ in the set $U$, adds it to the clique found so far and recursively calls the sub routine CLIQUE (with vertex $u$ and the set $U$ comprising of vertices that are the neighbors of $u$ as well as the neighbors of the vertices that are part of the clique found so far). If the set $U$ passed to the CLIQUE sub routine is empty, the value of $\text{max}$ is updated if it is less than the size of the clique found so far and the recursion is terminated; otherwise, another recursive call to the sub routine CLIQUE is made with a new vertex $u$ randomly chosen/removed from the set $U$ of vertices passed during the latest recursion. The sub routine CLIQUE (called with vertex $v_i$) proceeds in iterations until all the candidate neighbor vertices in the set $U$ (that was initially passed to it) are explored through a sequence of recursive calls as explained above. At each stage, the sub routine CLIQUE (called for a particular vertex $v_i$) only explores those neighbors of $v_i$ and their individual neighborhoods that have scope for being part of a clique that is larger than the maximum size clique found so far (whose size is kept track of using the variable $\text{max}$).

A characteristic property of the exact algorithm described here is that the maximum size clique for the entire graph is found during a particular iteration involving vertex $v_i$ such that vertices $v_j$ whose neighborhood are explored after this iteration are not part of the clique. If the maximum size clique for the entire graph is found in an earlier iteration itself, then the subsequent iterations could run relative more quickly as all they will do is to simply prune the search space as much as possible. Hence, the time efficiency of the exact algorithm is significantly influenced by the order in which the vertices are considered to be explored for the iterations as explained in the previous paragraphs and illustrated in Figure 1. Thus, if vertices are to be explored in the increasing order of their IDs, the search time could significantly reduce only if a vertex with a smaller ID is part of the maximum size clique. For guaranteed better performance, it would be more apt to run MAXCLIQUE by exploring vertices in the decreasing order of their degree, rather than simply based on the increasing order of their IDs (since the maximum clique size in a graph cannot be larger than the largest value for the degree of any vertex in the graph).
The exact algorithm (describe above to determine the maximum size clique for an entire graph) can be easily modified to determine the maximal size clique for an individual vertex of the graph, as illustrated in Figure 2. In order to determine the maximal clique size distribution of the vertices of a graph, we have to now run the procedure MAXIMALCLIQUE for each vertex \( v_i \) by considering all the neighbors of \( v_i \) for possible inclusion in the clique. This would increase the overall run time of the algorithm; but, it is unavoidable. However, the strategies incorporated by the exact algorithm in the procedure CLIQUE could be retained: the recursive calls in CLIQUE are made only for those neighbors \( u \) that have the potential to be part of a clique that is larger than the maximal size clique found until then for vertex \( v_i \). For guaranteed better performance of procedure CLIQUE called with a vertex \( v_i \), it would be more apt to explore the neighbors of \( v_i \) in the decreasing order of their degree.

---

**Procedure MAXIMALCLIQUE** \((G = (V, E))\)

```plaintext
for i : 1 to |V| do
    maximalCliqueSize[v_i] ← 0
    U ← \(\emptyset\)
    for each \( v_i \in \text{Neighbor}(v_i) \) do
        U ← \( U \cup \{v_i\} \)
        CLIQUE(G, v_i, U, 1)
```

**Subroutine CLIQUE** \((G = (V, E), v_i, U, \text{size})\)  // size is the size of clique found so far for vertex \( v_i \)

```plaintext
if U = \(\emptyset\) then
    if size > maximalCliqueSize[v_i] then
        maximalCliqueSize[v_i] ← size

return
while |U| > 0 do
    if size + |U| ≤ maximalCliqueSize[v_i] then
        return
    select any vertex \( u \) from \( U \)
    U ← \( U \backslash \{u\} \)
    \( N(u) := \{w | w \in \text{Neighbor}(u) \land \text{degree}(u) \geq \text{maximalCliqueSize}[v_i]\} \)
    CLIQUE(G, v_i, U \(\cap\) \( N(u) \), size + 1)
```

---

Figure 2. Exact Algorithm to Determine the Distribution of the Maximal Clique Size of the Vertices in a Graph (adapted from [6])

3. **Small-World Networks and Their Characteristics**

Small-world networks are a category of complex networks that exhibit a smaller diameter (maximum of the number of hops in the shortest paths connecting any two nodes) as well as a larger clustering coefficient (the probability that two nodes that share a common neighbor are connected). The other two categories of complex networks (random networks and scale-free networks) exhibit a significantly lower clustering coefficient. Small-world networks are characteristic of having a majority of the links in the local neighborhood of the vertices (helps to sustain a larger clustering coefficient) and having very few links (but not negligible) that connect vertices that do not have any common neighbors (still sufficient enough to ensure a smaller diameter). As the clustering coefficient of a vertex is measured as the ratio of the number of links among the neighbors of a vertex to that of the maximum possible number of links among the neighbors of a vertex, the characteristic of possessing a majority of the links in the local
neighborhood of the vertices helps the small-world networks to incur a significantly larger clustering coefficient, compared to the random networks and scale-free networks.

Initially, we start with a regular network wherein the number of neighbors per node (i.e., the number of links per node, identified as $K_{\text{regular}}$) is the same (and typically, an even number of links per node) as well as there is a particular pattern in the distribution of the links in the network (typically dependent on the dimension of the network). The regular network envisioned in this paper is a one-dimensional network with a ring as the underlying topological structure. Each node is connected to at least two other nodes (i.e., to the two neighboring nodes that are each one hop away in the ring): if there are more than 2 links per node, then the node is connected to nodes in the increasing order of the hop count in the ring. In general, if the number of links per node is $K_{\text{regular}}$, then a node is connected to neighbor nodes that are $1, 2, \ldots, \frac{K_{\text{regular}}}{2}$ hops away from the node on the ring. Figure 3 displays a 10-node regular network with four links per node (i.e., $K_{\text{regular}} = 4$) and each node is connected to nodes that are 1 and 2 hops away from it in the ring.

![Figure 3. Example for an One-Dimensional Regular Network (K_{regular} = 4 Links per Node)](image)

The WS model operates based on a tuning parameter called the probability of link rewiring ($P_{\text{rewire}}$). We rewire each link in the regular network with the probability $P_{\text{rewire}}$. For each link $u-v$ in the regular network, we generate a random number (in the range 0 to 1) and if it is less than $P_{\text{rewire}}$, we decide to rewire the link. When a link $u-v$ is chosen for rewiring, we choose a target node $w$ uniform-randomly among the nodes in the network (such that $w$ is neither $u$ nor $v$), remove the link $u-v$ and connect node $u$ with node $w$ (i.e., add the link $u-w$ to the network). We repeat the above procedure for every link in the initial regular network. Note that the newly added links are not considered for rewiring.

We conduct simulations to transform a regular network to a small-world network and subsequently to a random network according to the WS model. The simulations are conducted for networks of 100 nodes and 200 nodes; the probability of rewiring is varied from values of 0.01 to 0.1, in increments of 0.01 (referred to as small-world network zone), and from values of 0.1 to 1.0, in increments of 0.1 (referred to as random network zone). The reasoning behind the above distinction for the probability of rewiring is based on our observations from the simulation results: for $P_{\text{rewire}}$ values of 0.01 to 0.1, the average diameter of any node in the network (average of the maximum of the number of hops from a node to any other node) reduces significantly, but with only a moderate reduction in the clustering coefficient - a phenomenon characteristic of small-world networks. On the other hand, as we vary the probability of rewiring from 0.1 to 1.0, the average diameter of any node in the network reduces only marginally, whereas the clustering coefficient reduces significantly, indicating the transformation of the small-world network to a random network. We also vary the initial number of links per node ($K_{\text{regular}}$) in the regular network from 4 to 20, in increments of 2. The results presented in Figures 4-10 are the average of
the results observed for 100 network graphs, simulated for each value of the number of nodes (100 and 200 nodes) and each value of the probability of rewiring as mentioned above.

Figure 4 captures the absolute values of the average diameter of any node in the network as well as the ratio of the average diameter with and without rewiring. For a given probability of rewiring, we observe the absolute average value for the diameter to be smaller when we start with a regular network with a larger number of links per node. As we do rewiring, within the small-world zone, we observe the differences in the average diameter per node (for different values of $K_{\text{regular}}$) to reduce significantly (in an exponential fashion); in the random network zone, the average diameter per node for different values of $K_{\text{regular}}$ does not vary appreciably. Based on the results for the ratio of the average diameter per node with and without rewiring, we observe that the percentage decrease in the average diameter per node is much higher for regular networks with fewer numbers of initial links, indicating the effectiveness of rewiring in reducing the path length. With increase in the number of nodes (from 100 nodes to 200 nodes), we observe the network diameter to reduce further (for a given $P_{\text{rewire}}$ and $K_{\text{regular}}$), especially in the random network zone.

Figure 5 illustrates the variation in the average clustering coefficient per node (averaged over all the nodes in the network) in the small-world and random network zones. In the small-world zone, we observe the percentage reduction in the clustering coefficient is by about only 25% (compared to the value observed for the originating regular network, without any rewiring) and the rate of decrease is the same for all values of $K_{\text{regular}}$. On the other hand, as we enter the random network zone, the reduction in the clustering coefficient is significantly larger and specifically, the clustering coefficient of networks that started with a lower $K_{\text{regular}}$ value reduces much faster.
compared to networks that started with a larger $K_{\text{regular}}$ value. As the number of nodes is increased from 100 to 200, we observe the reduction in the clustering coefficient in the random network zone (in terms of both the magnitude as well as the rate of decrease) to be larger. For a given value of $K_{\text{regular}}$, the fraction of the number of links between any two neighbors of a node (compared to the maximum number of links between the neighbors of a node) is bound to be lower for networks with a relatively larger number of nodes, thus reducing the clustering coefficient.

Figure 5. Impact of the Probability of Link Rewiring and the Initial Number of Links per Node on the Average Clustering Coefficient per Node: Transition from Regular Network to Small-World Network and Random Network

4. ANALYSIS OF THE DISTRIBUTION OF MAXIMAL CLIQUE SIZE

Figure 6 captures the variation in the average maximal clique size per node (average of the maximal clique size of all the nodes, measured at the end of rewiring) for various values of the probability of rewiring and the initial number of links per node in the originating regular network. We observe that the small-world zone does not suffer any noticeable decrease in the average maximal clique size per node and the ratio of the average maximal clique size per node with and without rewiring is close to 1. As we transition from the small-world zone to the random network zone, we observe the average maximal clique size to reduce relatively at a much faster rate, with increase in the probability of rewiring. An interesting observation is that the average maximal clique size of random networks that start with a larger $K_{\text{regular}}$ value decreases at a much faster rate compared to the rate of decrease of the average maximal clique size of random networks that start with a lower $K_{\text{regular}}$ value (though the absolute values for the average maximal clique size is still larger for random networks that start with a larger $K_{\text{regular}}$ value). This indicates that the larger cliques present in the small-world networks that started with regular networks of larger $K_{\text{regular}}$ values are more likely to quickly get dismantled as the probability of link rewiring increases.
beyond the threshold value of \( P_{\text{rewire}} \) for small-world network zone. Thus, the average maximal clique size per nodes for different values of \( K_{\text{regular}} \) tend to get closer as we increase the probability of rewiring in the random network zone, and such a convergence is relatively more pronounced for networks with 200 nodes, compared to 100 nodes. Accordingly, for a given \( P_{\text{rewire}} \) and \( K_{\text{regular}} \), the rate of decrease in the average maximal clique size per node is much more steeper for networks with 200 nodes.

Figures 6-10 capture the variation in the maximal clique size for the nodes in the small-world zones and random network zones. For a given value of \( P_{\text{rewire}} \) and \( K_{\text{regular}} \), we observe the distribution of the maximal clique size is Poisson for both the zones. Figures 7 and 9 capture the distribution of the maximal clique size in the small-world zone. For smaller values of \( K_{\text{regular}} \) (4 and 6 links per node), we observe the maximal clique size per node to be very close to the average value for all the nodes; as we increase the value of \( K_{\text{regular}} \), we observe the maximal clique size per node to vary slightly, but not much different from the average value for the maximal clique size - coinciding with the invariant nature of the average maximal clique size per node observed in Figure 6. For a given value of \( K_{\text{regular}} \), the average maximal clique size of a regular network is \( 1 + \frac{K_{\text{regular}}}{2} \) and the average maximal clique size per node in a small-world network is very close to this value (minimal variation) observed for its predecessor regular network. For networks with larger \( K_{\text{regular}} \) values, the values for the maximal clique size per node is less than the average value by at most 2 and greater than the average value by at most 1, and as observed in Figures 7 and 9, these deviations occur with a vary small probability. The Poisson curve for the maximal clique size per node shifts to the right in such a way that the peak for the curve increases by a value of 1 as we increase the value of \( K_{\text{regular}} \) by 2. For a given value of \( K_{\text{regular}} \), the Poisson curve for the maximal clique size per node is more steep (minimal variation) and the tallest for a \( P_{\text{rewire}} \) value of 0.01 and becomes more spread out (shallow and more variations) as the \( P_{\text{rewire}} \)
value increases to 0.10 and above. Coinciding with the observations made in Figure 6, for a given value of $K_{\text{regular}}$ and $P_{\text{rewire}}$, there is not much variation in the distribution of the maximal clique size per node for networks of 100 nodes and 200 nodes.

Figures 7 and 8 capture the variation in the maximal clique size for the nodes in the random network zone for a given probability of rewiring and varying the initial number of links per node with values of 4, 12 and 20 links - scenarios that exhibit minimal, moderate and maximum variation in the maximal clique size per node as the probability of rewiring increases. For lower values of the probability of rewiring (0.1 and 0.2; when the network is still in the small-world zone), the distribution of the maximal clique size per node is taller for each value of $K_{\text{regular}}$ and the distributions are non overlapping (as the $K_{\text{regular}}$ values are 4, 12 and 20, the average maximal
clique size is around 3, 7 and 11 - vindicating the non-overlapping nature of the peaks and the distribution of the maximal clique size for lower values of $P_{\text{rewire}}$. With increase in $P_{\text{rewire}}$, we start observing the distributions of the maximal clique size for the three fairly different values of $K_{\text{regular}}$ to start overlapping; the distributions tend to shift to the left - coinciding with a decrease in the average maximal clique value. With increase in $P_{\text{rewire}}$, the shift towards lower values of the maximal clique is more pronounced for networks with a larger $K_{\text{regular}}$ value, vindicating the rapid fall in the average maximal clique size; also for larger values of $K_{\text{regular}}$, the distributions for the maximal clique become more spread out with increase in $P_{\text{rewire}}$ - lowering the probability of finding the maximal clique size per node to be close to the average value. On the other hand, for networks with lower values of $K_{\text{regular}}$, the distribution for the maximal clique size remains fairly narrow (even with increase in $P_{\text{rewire}}$), indicating that it is still possible to observe the maximal clique size for any node to be close to the average value.

Figure 9. Distribution of the Maximal Clique Size vs. Initial Number of Links per Node: Transition from Regular Network to Small World Network [200 Node Network]

Figure 10. Distribution of the Maximal Clique Size vs. Probability of Link Rewiring: Transition from Small World Network to Random Network [200 Node Network]
5. RELATED WORK

The focus of research in the literature so far has been on developing heuristics (both time and space efficient) for determining the maximum size for the entire network graphs, based on strategies (like branch and bound) that reduce the solution search space. For example, branch and bound strategies based on node degree [6], vertex coloring [7] and vertex ordering [8] have been proposed as potential strategies for effectively and efficiently pruning the solution search space. In addition, there have been efforts to develop parallelized versions of branch and bound (e.g., [9]) that could be applied to determine cliques in large real-world network graphs with 1000 to 100 million nodes. In [12], the authors explore the use of maximal cliques of size at least $k$ nodes to identify overlapping communities based on percolation; two cliques of size $k$ are said to percolate into each other if they share $k$ nodes. With regards to analyzing the distribution of clique-related metrics for complex networks, in [13], the authors have analyzed the distribution of clique-degree of the nodes (the clique-degree of a node is the number of cliques of a particular size the node is part of) in various real-world networks and have observed them to exhibit a power-law distribution. As far as we know, there is no work that has analyzed the distribution of the maximal clique size of the vertices for complex network graphs that evolve under any well-known theoretical model (like the Watts-Strogatz model for small-world networks that eventually transform to a random network with increase in the probability of link rewiring).

6. CONCLUSIONS

The following significant conclusions could be made from the research conducted in this paper. with regards to the distribution for the maximal clique size per node for small-world networks and random networks that evolve from a regular network. As we transform from a regular network (with $K_{regular}$ number of links per node) to a small-world network through link rewiring, the maximal clique size of the nodes is invariant and very close to that of the average maximal clique size per node as well as close to that of the average maximal clique size per node in the regular network. As we transform from a small-world network to a random network (by increasing the probability of rewiring), the distribution of the maximal clique size per node becomes more broader and thereby the probability of observing a maximal clique size per node close to that of the average maximal clique size is relatively much lower. Also, with increase in the probability of rewiring, the distributions for the maximal clique size obtained for different $K_{regular}$ values overlap each other and shift towards a lower average value. Nevertheless, for all the scenarios/values for the probability of rewiring and the initial number of links per node, the distribution for the maximal clique size reflects that of a Poisson distribution. As the vertices in a random network exhibit a Poisson style distribution for the node degree, we conjecture a high positive correlation between the maximal clique size per node and node degree in small-world networks and random networks, and this will be further analyzed in future work.

REFERENCES


Author

Dr. Natarajan Meghanathan is currently an Associate Professor of Computer Science at Jackson State University, USA. His areas of research interests are Network Science and Graph Theory, Wireless Ad hoc Networks and Sensor Networks, Cyber Security and Machine Learning. He has published more than 150 peer-reviewed articles and obtained grants from several federal agencies. He serves as the editor-in-chief of three international journals as well as serves in the organizing committees of several international conferences.