1 Proof Continued from the Previous Class

\[ \text{Prob}[X > (1 + \delta) \mu] < e^{-\frac{\mu \delta^2}{4}} \quad \text{when} \delta \leq 2e-1 \]

Now, \( \mu = C|S_i| \)

\( \delta = \frac{1}{4} \)

So, plugging the values in the bound provide us with the following equation:

\[ e^{-\frac{C|S_i| \frac{1}{4} \delta^2}{4}} = e^{-\frac{C|S_i|}{64}} \]

[Ensures every time that the elements are independent]

Now, \( p(n) \geq C \cdot \frac{\ln n}{n} \) where, C is constant

So, \( C = p(n)(n-1) \geq C \cdot \ln n \)

Plugging this into the bounds, we get an upper bond of

\[ e^{-\frac{C \ln n |S_i|}{64}} \quad \text{As,} \quad e^{\ln} = \frac{1}{n} \]

Picking C large enough gives a bound \( (\frac{1}{n}) |S_i| \leq \frac{1}{n} \)

By this upper bound is bound as \( (\frac{5C}{4}) |S_i| \leq \frac{1}{n} \)

So, \( \text{Prob}[|S_{i+1}| > \frac{5C}{4} |S_i|] \leq \frac{1}{n} \)

Similarily, \( \text{Prob}[|S_{i+1}| < \frac{C}{4} |S_i|] \leq \frac{1}{n} \)

\[ \square \]

Lemma  Let, \( |T_i| \geq \frac{n-1}{2} \) then,

\[ \text{Prob}[\frac{C}{4}^{i+1} \leq |S_i + 1| \leq \frac{5C}{4}^{i+1}] \geq 1 - \frac{1}{n} \]

This will be true, only if \( S_0, S_1, \ldots, S_{i+1} \) follow this rule. Now as we know, this is true with a very high probability. As the probability of its non-occurrence is only \( \frac{2(i+1)}{n} \). With changes in the value of C[constant] this probability becomes \( \frac{2}{n} \), which is very small.

Theorem  Let \( u,v \in V \). Then, with probability \( \geq 1 - \frac{1}{n} \)
\[ d(u,v) \leq O\left(\frac{\ln n}{\ln C}\right) \]

This goes on until we reaches \(|T_i| < \frac{n-1}{2}\).

Now, as we reached to the end set, say \(S_{i+1}\) then the ball having the reached nodes strictly has more than half of the nodes of the graph.

\[ d(u,w) = \ln n \frac{n}{\ln C} \]

And, \(d(w, v) = \frac{\ln n}{\ln C}\)

The above relation was proved assuming, \(p(n) \geq C\frac{\ln n}{n}\)

But, the claim is actually true for, \(p(n) > \frac{1}{n}\)

Now, a big problem occurs due to this. As, these two points occur at different times.
2 Phase Transition In ER Graphs

Lemma If, \( p(n) < \frac{\ln n}{n} \) then with probability \( \to 1 \) as \( n \to \infty \), the graph has atleast one isolated vertex. This shows disconnectivity of the graph.

Proof For a vertex \( u \in V \), let:

\[
I_u = \begin{cases} 
1 & \text{if } u \text{ is isolated} \\
0 & \text{otherwise}
\end{cases}
\]

\[
\text{Prob}[I_u = 1] = (1-p)^{n-1} \sim e^{-p(n-1)}
\]

So, \( E[I_u] \sim e^{-p(n-1)} \)

Let, \( X = \sum I_u = ne^{-p(n-1)} \)

Suppose, \( p = \lambda \frac{\ln n}{n} \) where, \( \lambda < 1 \)

then, \( E[X] = ne^{-\lambda \ln n} = n^{1-\lambda} \)

if, \( \lambda = 0.9 \), then \( n^{0.1} \) is expected isolated graph.
Note that, As $n \to \infty$, $E[X] \to \infty$

Further calculation involving the variance can be used to show that the,

$\text{Prob}[\text{there exist an isolated vertex}] \to 1$ as $n \to \infty$  \hspace{1cm} \square$

## 3 Watts Strogatz Model

The table below describes the various networks with their degree and cluster coefficient.\[1\]

<table>
<thead>
<tr>
<th></th>
<th>$N$</th>
<th>Average degree</th>
<th>$CC$</th>
<th>CC of corresponding ER Graph</th>
</tr>
</thead>
<tbody>
<tr>
<td>Actor Network</td>
<td>225,226</td>
<td>61</td>
<td>0.79</td>
<td>0.00027</td>
</tr>
<tr>
<td>Power Grid</td>
<td>4941</td>
<td>2.67</td>
<td>0.080</td>
<td>0.005</td>
</tr>
<tr>
<td>C.elegance</td>
<td>282</td>
<td>14</td>
<td>0.28</td>
<td>0.05</td>
</tr>
</tbody>
</table>

There are observed networks that show the properties like:
1. Sparse (average degree is small relative to $N$)
2. Small average path length relative to $N$
3. Cluster coefficient is high relative to that of corresponding ER graph.

Ques) Is there a simple random graph model with these 3 characteristics?

Answer) If we take a Circular Graph : $C(n,k)$

for $n=10$, $k=4$, $C(10,4)$ is like:

![Circular Graph C(10,4)](image)

Then, the above discussed points are satisfiable as:

1. Sparsity is controlled by $k$, if $k$ is large, sparsity is less.

2. Cluster coefficient is high, as every V-node is connected to the neighbours, half on one side and half on other side, and these neighbours are also connected in similar way.

3. Average path length depends on $k$, for high value of $k$, it will be small.

**Watts Strogatz Model** $WS(n,k,p)$ where $0 \leq p \leq 1$

Now, we choose 2 vertex and one edge and reconnect it.

As, $p$ goes from 0 to 1, randomness of the graph increases.

At, $p = 1$, Original graph is completely lost and we have a totally random graph $ER(n, ?)$

At, $p = 0$, we have an original graph $C(n,k)$.

In intermediate region the property of graph is like:
So, at very small randomness, we see a lot of decrease in the path length but we do not see a lot of change in clustering coefficient.

References