STAT206 From Random Graphs To Complex Networks	**Spring 2003**

Lecture 5: February 4

Lecturer: David Aldous — *Scribe: Shelley Zhuang*

Disclaimer: *These notes have not been subjected to the usual scrutiny reserved for formal publications. They may be distributed outside this class only with the permission of the Instructor.*

Today's lecture discusses Clustering Coefficient and Newman-Watts Small World Model.

5.1 Clustering Coefficient

Consider an undirected graph, then a 2-path is a path of two edges, a triangle has three 2-paths.

We define Clustering Coefficient (CC) to be

$$CC = \frac{3(\text{\# triangles})}{\text{\# 2-paths}}$$

for a non-random graph.

Recall from elementary probability, for random variables $X, Y > 0$, $E(X/Y) \neq EX/EY$. Thus for a random graph, we define

$$CC = \frac{3\text{E(\# triangles)}}{\text{E(\# 2-paths)}}$$

which can be interpreted as "relative density of triangles", or as "chance a typical 2-path is part of a triangle".

Now we want to localize the above definition, and rephrase it in terms of a uniformly random vertex V. Let n be number of vertices in a realization of a graph.

$$\text{E(\# 2-paths with central vertex = V}|\text{realization of a graph) = 1/n * (\# of 2-paths)}$$
$$\text{E(\# triangles containing V}|\text{realization of a graph) = 3/n * (\# of triangles)}$$

Taking expectation of both sides of the above equations, we get

$$\text{E(\# 2-paths with central vertex = V) = 1/n * E(\# of 2-paths)}$$
$$\text{E(\# triangles containing V) = 3/n * E(\# of triangles)}$$

Then

$$
\begin{aligned}
CC &= \frac{\text{3E(\# of triangles)}}{\text{E(\# 2-paths)}} \\
&= \frac{\text{E(\# triangles containing V)}}{\text{E(\# 2-paths with central vertex V)}} \\
&= \frac{\text{E(\# triangles containing V)}}{E\binom{D}{2}}
\end{aligned}
$$

where D = d(V) = degree of V, and we can calculate $E\binom{D}{2}$ in terms of ED and $var(D)$.

$E(D(D-1)) = E(D^2) - ED = var(D) + (ED)^2 - ED$

$E\binom{D}{2} = \frac{var(D)+(ED)^2-ED}{2}$

5.1.1 Erdos-Renyi Random Graph Model

$\mathcal{G}(n, p)$

n-vertices, each possible edge is present with probability p independently.

<u>Sparse Case</u>

$p = \lambda/n, 0 < \lambda < \infty$

Degree of vertex $\sim$ Binomial(n-1, λ/n) $\approx$ Poisson(λ) for n large.

This random graph model has been studied at great depth, and arguments exploit special structure not present in other models. It is a special case of Lecture 3 random graph with prescribed degree distribution.

What is CC in Erdos-Renyi random graph $\mathcal{G}$(n, λ/n)?

$$
\begin{aligned}
\text{E(\# triangles)} &= \binom{n}{3}(\lambda/n)^3 \\
\text{E(\# 2-paths)} &= \frac{n(n-1)(n-2)}{2}(\lambda/n)^2
\end{aligned}
$$

$CC = \lambda/n \to 0$ as $n \to \infty$, which is implied by "locally tree-like" property as $n \to \infty$. $CC \to 0$ as $n \to \infty$ in other tree-like models from lectures 3 and 4.

5.2 Newman-Watts Small World Model

Consider a grid with N^d vertices, $\mathcal{G}(N^d, d, k, \alpha)$

$k > 2d$, such that definition is unambiguous. Put an edge from each vertex to its k Euclidean-nearest neighbors. We already get triangles from this! Now choose a parameter $0 < \alpha < \infty$, for each pair of vertices, create extra-edge with probability $\frac{\alpha}{N^d}$.

We study limits as $N \to \infty$.

D = degree of uniformly random vertex V $\approx k$ + Poisson(α)

This does not have a power-law tail.

$$\begin{aligned}
\text{E(\# 2-paths with central vertex V)} &= E\binom{D}{2} \\
&= \frac{1}{2}(var(D) + (ED)^2 - ED) \\
&= \frac{1}{2}(\alpha + (k+\alpha)^2 - (k+\alpha)) \\
&= \frac{1}{2}(k^2 + (2\alpha - 1)k + \alpha^2)
\end{aligned}$$

# of triangles containing a given vertex in non-random ($\alpha = 0$) case is $\gamma_{d,k}$ by elementary geometry. So what happens when we add extra random edges? We want to compute an upperbound on # extra triangles containing V as a result. There are three possible cases.

1. $\leq k^2 \frac{\alpha}{N^d}$
2. $\leq kN^d (\frac{\alpha}{N^d})^2$
3. $\leq (N^d)^2 (\frac{\alpha}{N^d})^3$

So upperbound E(# extra triangles containing V) $\leq O(1/N^d)$ as $N \to \infty$. Thus,

$$\begin{aligned}
CC &= \frac{\text{E(\# triangles containing V)}}{E\binom{D}{2}} \\
CC(\alpha) &= CC(0)\frac{k^2 - k}{k^2 + (2\alpha - 1)k + \alpha^2}
\end{aligned}$$

because E(# triangles containing V) does not change with α. Note $CC(\alpha)$ stays away from 0 as long as α/k is bounded.

Now we study L = graph distance in $\mathcal{G}(N^d, d, k, \alpha)$ between two uniformly random vertices. If α = 0, then $EL \sim C_{d,k}N$. Just adding 1 extra long edge will change constant $C_{d,k}$.

If we use a different scaling, where P(edge is an extra edge) = β/N^{2d}. E(# extra edges) = $\binom{N^d}{2}\frac{\beta}{N^{2d}} \approx \frac{\beta}{2}$, and # extra edges $\approx$ Poisson($\beta/2$). $EL \sim f_d(\beta)N$ as $N \to \infty$, but there is no hope for a useful formula for $f_d(\beta)$.

Back to P(extra edge) = α/N^d. Ignore original lattice edges, and look at Erdos-Renyi random graph $\mathcal{G}(N^d, \alpha/N^d)$. For this, $EL \sim \log N^d / \log \alpha$. We expect $EL \sim g_{d,k}(\alpha) \log N^d$, where $g_{d,k}(\alpha) \sim 1/\log \alpha$ for large α, which has "small-worlds" feature $EL \approx \log N^d$.