Random Models Of Complex Networks

Nice July 8-19, 2019

Topics

Models

-Mean Field Model – $G_{n,p}$

- Fixed degree sequence and
  the Configuration Model
- Forbidden Substructure
- Stochastic Processes:
  Preferential Attachment &
  Triangle-free Graph Process

-Brief Nod to Real World Networks

Structures

-Clustering: Components,
Conductance, Modularity,
Cliques, Stable Sets,(Colouring)

Techniques

-Concentration Inequalities

-Differential Equation Method

-Regularity Lemma

-Switching Arguments

Connectivity & Component Size in $G_{n.p}$

$G_{n,p}$: For each H on $\mathcal{V}_n$ ={v_1,…,v_n}, **P**($G_{n,p}$=H) is $p^{|E(H)|}\ (1-p)^{\binom{n}{2}-|E(H)|}$

The Threshold for Connectivity of $G_{n,p}$

$G_{n,p}$: For each H on $\mathcal{V}_n$ =$\{v_1,\ldots,v_n\}$, **P**($G_{n,p}$=H) is $p^{|E(H)|}\,(1-p)^{\binom{n}{2}-|E(H)|}$

Simple Observations

Letting Y be the (random) number of components of G whose size is between 1 and n/2. G is connected if and only if Y=0.

Letting Y_S be the number of components of size s, $Y=\sum_{s=1}^{\lfloor n/2 \rfloor} Y_s$

The Expected Number of Components of Size s in $G_{n,p}$

Y_s=number of components of size s in $G_{n,p}$

$$\binom{n}{s} s^{s-2} p^{s-1} (1-p)^{(n-s)(s)} \geq \mathbf{E}(Y_s) \geq$$

The Expected Number of Components of Size s in $G_{n,p}$

Y_s=number of components of size s in $G_{n,p}$

Z_s = number of components of size s in $G_{n,p}$ inducing trees.

$$\binom{n}{s} s^{s-2} p^{s-1} (1-p)^{(n-s)(s)} \geq \mathbf{E}(Y_s) \geq \mathbf{E}(Z_s) \geq$$

$$(1-p)^{(s-2)(s-1)/2} \binom{n}{s} s^{s-2} p^{s-1} \quad (1-p)^{(n-s)(s)}$$

The Expected Number of Components of Size s in $G_{n,p}$

Y_s=number of components of size s in $G_{n,p}$

Z_s = number of components of size s in $G_{n,p}$ inducing trees.

$\binom{n}{s} s^{s-2} p^{s-1} (1-p)^{(n-s)(s)} \geq \mathbf{E}(Y_s) \geq \mathbf{E}(Z_s) \geq$

$(1-p)^{(s-2)(s-1)/2} \binom{n}{s} s^{s-2} p^{s-1} \quad (1-p)^{(n-s)(s)}$

Observation: For $p = \frac{\log n}{n}$, $E(Y_1) = \Phi(1)$, $E(Y_2) = \Theta(\frac{\log n}{n})$, and

$E(Y_k) = o(\frac{1}{n})$ for $3 < k \leq \frac{n}{2}$.

$p = \frac{\log n}{n} + \frac{f(n)}{n}$ ***,f***->∞, f(n)<log n =>E(Y)=o(1). So P($G_{n,p}$ connected)=1-o(1)

Expectation and Concentration of The Number of Isolated Vertices

Y_1= |{v s.t. d(v)=0}|

E(Y_1)=n(1-p)$^{n-1}$$\approx e^{pn-\log n}$

This is ω(1) when p=$\frac{\log n-\omega(1)}{n}$

E(Y_1^2)=n(1-p)$^{n-1}$ +n(n-1)(1-p)$^{2n-3}$

$\leq$**E**(Y_1) +(1-p)$^{-1}$**E**(Y_1)^2

So for p=$\frac{\log n-\omega(1)}{n}$

E(Y_1^2)-**E**(Y_1)^2 =o(**E**(Y_1)^2) and

P(Y_1=0)$\leq$ $\boldsymbol{P}$(|Y_1-**E**(Y_1)|$\geq$ $\boldsymbol{E}$(Y_1))
=o(1)

Component Size in $G_{n,p}$

The Components of $G_{n,p}$ Are Usually

VERY BIG or very small

Expected Number of Components of Size s in $G_{n,p}$

$\forall\ \varepsilon > 0\ \exists A_\varepsilon, b_\varepsilon$ >0 s.t. (i) if $p < (1-\varepsilon)/n$ or $p > (1+\varepsilon)/n$ then a.s. no components of $G_{n,p}$ have size between $A_\varepsilon \log n$ and $b_\varepsilon n$.

When does $G_{n,p}$ Have A LARGE COMPONENT?

$|\frac{p}{n}-1| > \varepsilon$ for some positive ε.

A Truncated Breadth First Search of $G_{n,p}$

At the start of iteration i E_i is a set of Explored Vertices & O_i is a set of open vertices. $O_1=v_1$ & $E_1=\emptyset$.

For i:=1 to t we:

choose v_i in O_i and expose the

set S_i of vertices of $\mathcal{V}_n - O_i - E_i$

joined to v_i by an edge.

set $O_{i+1} = O_i + S_i - v$, $E_{i+1} = E_i + v_i$

(*) if $|O_{i+1}|=0$ we add a vertex of $\mathcal{V}_n$-

$O_{i+1} - E_{i+1}$ to O_{i+1}

A Truncated Breadth First Search of $G_{n,p}$

At the start of iteration i E_i is a set of Explored Vertices & O_i is a set of open vertices. $O_1=v_1$ & $E_1=\emptyset$.

For i:=1 to t we:

choose v_i in O_i and expose the

set S_i of vertices of $\mathcal{V}_n-O_i-E_i$

joined to v_i by an edge.

set $O_{i+1}=O_i+S_i-v$, $E_{i+1}=E_i+v_i$

(*) if $|O_{i+1}|=0$ we add a vertex of $\mathcal{V}_n-$

$O_{i+1}-E_{i+1}$ to O_{i+1}

Ignoring the vertices added in (*)_ the number of vertices added in iteration i is $Bin(n-|E_i|-|O_i|,p)$. This lies between Bin(n,p) and Bin(n+1-i-|O_i|,p).

A Truncated Breadth First Search of $G_{n,p}$

At the start of iteration i E_i is a set of Explored Vertices & O_i is a set of open vertices. $O_1=v_1$ & $E_1=\emptyset$.

For i:=1 to t we:

choose v_i in O_i and expose the set S_i of vertices of $\mathcal{V}_n$-O_i-E_i joined to v_i by an edge.

set $O_{i+1}=O_i+S_i-v$, $E_{i+1}=E_i+v_i$

(*) if $|O_{i+1}|=0$ we add a vertex of $\mathcal{V}_n$-O_{i+1}-E_{i+1} to O_{i+1}

Ignoring the vertices added in (*)_ the number of vertices added in iteration i is Bin(n-$|E_i|$-$|O_i|$,p). This lies between Bin(n,p) and Bin(n+1-i-|O_i|,p).

Hence, setting t=$\frac{10}{\varepsilon^2}$ log n and n'=nt,

P(v_1 is in a component of size > A log n)<

P(Bin(n',p)>t-1) {=o($\frac{1}{n}$) if p=$\frac{1-\epsilon}{n}$}

A Truncated Breadth First Search of $G_{n,p}$

At the start of iteration i E_i is a set of Explored Vertices & O_i is a set of open vertices. $O_1=v_1$ & $E_1=\emptyset$.

For i:=1 to t we:

choose v_i in O_i and expose the set S_i of vertices of $\mathcal{V}_n$-O_i-E_i joined to v_i by an edge.

set $O_{i+1}=O_i+S_i-v$, $E_{i+1}=E_i+v_i$

(*) if $|O_{i+1}|=0$ we add a vertex of $\mathcal{V}_n$-O_{i+1}-E_{i+1} to O_{i+1}

Ignoring the vertices added in (*)_ the number of vertices added in iteration i is Bin(n-$|E_i|$-$|O_i|$,p). This lies between Bin(n,p) and Bin(n+1-i-|O_i|,p).

Hence, setting t=$\frac{10}{\varepsilon^2}$ log n and n'=nt,

P(v_1 is in a component of size > A log n)<

P(Bin(n',p)>t-1) {=o($\frac{1}{n}$) if p=$\frac{1-\epsilon}{n}$}

Setting ε'=min($\frac{1}{4}$,ε), t'=$\frac{\varepsilon' n}{4}$ & n''= (1-$\frac{\varepsilon'}{2}$)nt',

and terminating if |O_i|>$\frac{\varepsilon' n}{4}$

P(There is a component of size > $\frac{\varepsilon' n}{4}$)>

P(Bin(n'',p)>$\frac{\varepsilon' n}{4}$) {=1-o($\frac{1}{n}$) if p=$\frac{1+\epsilon}{n}$}

Random Graphs on a Fixed Degree Sequence & The Configuration Model

Configuration Model on $\mathcal{D}=\{d_1,\ldots,d_n\}$

$C_i = \{v^1_i, v^2_i, \ldots, v_i^{d_i}\}$

Take a random matching $M_{\mathcal{D}}$ on $\bigcup_{i=1}^n C_i$

Create $H(\mathcal{D})$: The number of edges between v_i and v_j is the number of edges of $M_{\mathcal{D}}$ between C_i and C_j.

The number of loops on v_i is the number of edges of $M_{\mathcal{D}}$ within C_i.

If each $d_i > 0$ and $\sum_{i=1}^n d_i^2 \leq Bn$ then the expected number of loops and multiple edges in $H(\mathcal{D})$ is at most B^2+B

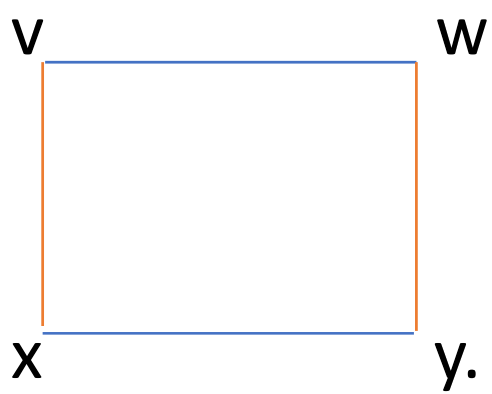
So the probability there are more than $2B^2+2B$ loops and multiple edges is at most ½.

A Switching Argument

Lemma: If $\mathcal{A}$ and $\mathcal{B}$ are families of (multi)graphs such that there are at least δ switches from each graph in $\mathcal{A}$ which result in a graph in $\mathcal{B}$ and at most Δ switches from each graph in $\mathcal{B}$ resulting in a graph in $\mathcal{A}$ then

$$\frac{\delta}{\Delta}|\mathcal{A}| \leq |\mathcal{B}|$$

Switch Delete {vw,xy} add {vx,wy}.



Lower Bounding The Probability A Random d-Regular Multigraph Is Loopless

Let $\mathcal{D}$={d,d,d…,d}

Let A_i be the set of matchings M s.t. if $\mathcal{M}_{\mathcal{D}}$ =M then $\mathcal{H}_{\mathcal{D}}$ has i loops

$\sum_{i=0}^{2d} \mathbf{P}(\mathrm{H}_{\mathcal{D}}$ in $A_i) > \frac{1}{2}$.

By our switching argument:

$$|A_{i-1}| \geq \frac{i(dn-d^2)}{\binom{d}{2}n}|A_i|$$

Hence, for large n,

$\forall$ j with $1 \leq j \leq 2d$:

$$|A_0| > \frac{|A_j|}{d^j}$$

So

$$\mathbf{P}(\mathrm{H}_{\mathcal{D}} \text{ has 0 loops}) > \frac{\mathbf{P}(H_D\ has\ j\ loops)}{d^{2d}}$$

So, $\mathbf{P}(\mathrm{H}_{\mathcal{D}}$ has 0 loops$) > \frac{1}{2(2d+1)d^{2d}}$

Bounding The Probability A Uniformly Random 2-factor has a cycle of length> $\boldsymbol{\varepsilon n}$

A_i 2-factors with i cycles of length between $\frac{\varepsilon n}{3}$ and εn and at least one cycle of length exceeding εn.

B_i number of 2-factors with i cycles of length between $\frac{\varepsilon n}{3}$ and εn and no cycle of length exceeding εn.

By a switching argument:

$|A_i| \leq \frac{n^2}{\varepsilon^2 n^2/8} (|A_{i+1}|+|B_{i+1}|)$

Since $|A_i|$ = 0 for i>$\frac{\varepsilon}{3}$ This gives an upper bound on **P**($\bigcup_i A_i$)

Expected number of cycles of length exceeding εn=

$$\sum_{i=\lceil \varepsilon n \rceil}^{n} \frac{n!}{2i((n-i)!)} \left(\frac{2^{i-1}}{\prod_{j=1}^{i}(2n-2j+1)}\right)$$

Which exceeds $\sum_{i=\lceil \varepsilon n \rceil}^{n} \frac{1}{4i} \approx \ln(1/\varepsilon)/4$

Gives lower bound of $\varepsilon \ln(1/\varepsilon)/4$