Update on Projects

Sidney Resnick
School of Operations Research and Information Engineering
Rhodes Hall, Cornell University
Ithaca NY 14853 USA

http://people.orie.cornell.edu/~sid
sir1@cornell.edu

MURI CFEM Update
March 5, 2015
1. Undirected growing preferential attachment graphs

Notation:

- $G_n =$ graph after $n$ changes; node set is $V_n = \{1, \ldots, n\}$.
- $D_n(v)$ = degree of $v \in V_n$.
- $\delta \geq -1 =$ parameter.
- Initialize with one node having a self loop.

Conditional on knowing the graph $G_n$, at stage $n+1$ a new node $n+1$ appears and with a parameter $\delta \geq -1$, either

1. The new node $n+1$ attaches to $v \in V_n$ with probability

$$\frac{D_n(v) + \delta}{n(2 + \delta) + (1 + \delta)}, \quad (1)$$

or

2. $n+1$ attaches to itself with probability

$$\frac{1 + \delta}{n(2 + \delta) + (1 + \delta)}. \quad (2)$$
In the first case

- \( D_{n+1}(n+1) = 1 \) and in the second case
- \( D_{n+1}(n+1) = 2 \).

Frequencies: The number of nodes at time \( n \) with degree \( k \) is

\[
N_n(k) = \sum_{v=1}^{n} 1[D_n(v) = k]
\]

and

\[
\frac{N_n(k)}{n} \to p_k, \quad (n \to \infty)
\]

and

\[
p_k = \left( \frac{(2 + \delta)\Gamma(3 + 2\delta)}{\Gamma(1 + \delta)} \right) \frac{\Gamma(k + \delta)}{\Gamma(k + 3 + 2\delta)} = c(\delta) \frac{\Gamma(k + \delta)}{\Gamma(k + 3 + 2\delta)}.
\]

and

\[
p_k \sim c(\delta)k^{-3-\delta}, \quad (k \to \infty).
\]
1.1. CLT for counts

With Gena: Using MG CLT:

$$\sqrt{n} \left( \frac{N_n(k)}{n} - p_k \right) \Rightarrow N(0, \sigma^2_k(\delta))$$

and also jointly in $k$. Requires knowing order of

$$\left| \frac{E(N_n(k))}{n} - p_k \right|.$$

1.2. Tail estimation: Estimate $\delta$.

MG convergence theorem: For each $v$:

$$\frac{D_n(v)}{n^{1/(2+\delta)}} \rightarrow \xi_v, \quad (n \rightarrow \infty).$$

Known

$$\xi_v \overset{d}{=} \xi_1 \prod_{j=1}^{v} B_i,$$

where $B_i$ are beta rv’s.

Known

$$\bigvee_{v \in V_n} \frac{D_n(v)}{n^{1/(2+\delta)}} \rightarrow \bigvee_v \xi_v < \infty.$$
Should imply

$$\sum_{v \in V_n} \epsilon \frac{D_n(v)}{n^{1/(2+\delta)}} \Rightarrow \sum_v \epsilon \xi_v$$

in $M_+(0, \infty]$.

1.2.1. Goals:

With node based data:

- Not much is known about the $\xi_v$'s. Is there any connection to Poisson processes. (Yes when $\delta = 0$. [Bollobas & Riordan])

- Connection to tail empirical measure? Use this to get convergence to mean measure.

- Imply the Hill estimator is consistent?
2. Directed edge preferential attachment

2.1. Model description

- Model parameters: $\alpha, \beta, \gamma, \delta_{in}, \delta_{out}$ with $\alpha + \beta + \gamma = 1$.
- $G(n)$ is a directed random graph with $n$ edges, $N(n)$ nodes.
- Set of nodes of $G(n)$ is $V_n$; so $|V_n| = N(n)$.
- Set of edges of $G(n)$ is $E_n = \{(u, v) \in V_n \times V_n : (u, v) \in E_n\}$.
- In-degree of $v$ is $D_{in}(v)$; out-degree of $v$ is $D_{out}$. Dependence on $n$ is suppressed.
- Obtain graph $G(n)$ from $G(n-1)$ in a Markovian way as follows:
1. With probability $\alpha$, append to $G(n-1)$ a new node $v \notin V_{n-1}$ and create directed edge $v \rightarrow w \in V_{n-1}$ with probability

$$\frac{D_{in}(w) + \delta_{in}}{n - 1 + \delta_{in}N(n - 1)}.$$ 

2. With probability $\gamma$, append to $G(n-1)$ a new node $v \notin V_{n-1}$ and create directed edge $w \in V_{n-1} \rightarrow v \notin V_{n-1}$ with probability

$$\frac{D_{out}(w) + \delta_{out}}{n - 1 + \delta_{out}N(n - 1)}.$$ 

3. With probability $\beta$, create new directed edge between existing nodes

$$v \in V_{n-1} \rightarrow w \in V_{n-1}$$

with probability

$$\left(\frac{D_{out}(v) + \delta_{out}}{n - 1 + \delta_{out}N(n - 1)}\right)\left(\frac{D_{in}(w) + \delta_{in}}{n - 1 + \delta_{in}N(n - 1)}\right).$$
2.2. CLT for count ratios.

With Tiandong Wang:
Let

\[ N_n(k, l) = \sum_{v \in V_n} 1[D_{\text{in}}(v) = k, D_{\text{out}}(v) = l] \]

\[ = \sum_{v \in V_n} \# \text{ nodes with in-degree } k, \text{ out-degree } l. \]

Know

\[ \frac{N_n(k, l)}{\# \text{ nodes at stage } n} \to p(k, l) \]

and \{p(k, l)\} has a distribution which is regularly varying.

2.2.1. Goals:

• Use MG CLT to study asymptotic normality of

\[ \sqrt{n} \left( \frac{N_n(k, l)}{\# \text{ nodes at stage } n} - p(k, l) \right) \]

• Eventually to find estimators for the parameters \( \alpha_{\text{in}}, \alpha_{\text{out}} \)
and

$$\delta_{in}, \delta_{out}, \alpha, \beta, \gamma.$$
2.3. Reciprocity

With Gena.

For the directed edge model,

What proportion of nodes have reciprocated edges? Find an asymptotic behavior as the number of edges $\to \infty$. Express this asymptotic as a function of input parameters.

Markov chain approach: Let the set of edges be,

$$E_n = \{(u, v) : u \in V_n, v \in V_n, u \to v \text{ or } v \to u\}.$$

Define

$$S_n(u, v) = \begin{cases} 
0, & \text{if } (u, v) \notin E_n, (v, u) \notin E_n, \\
1, & \text{if } (u, v) \in E_n, (v, u) \notin E_n, \\
2, & \text{if } (u, v) \notin E_n, (v, u) \in E_n, \\
3, & \text{if } (u, v) \in E_n, (v, u) \in E_n,
\end{cases}$$
3. $r$th largest

With Ross Maller, Boris Buchmann (ANU).

Setup:

- $\{X_n\}$ iid with a continuous df.
- $M_n^{(r)} = r$th largest among $X_1, \ldots, X_n$.
- $M^{(r)} = \{M_n^{(r)}, n \geq r\} \in \mathbb{R}^\infty$.

Fact: $\{M^{(r)}, r \geq 1\}$ is a Markov chain on $\mathbb{R}^\infty$.

- Relative ranks:

$$R_n = \sum_{j=1}^{n} 1_{[X_j \geq X_n]}$$

= relative rank of $X_n$ among $X_1, \ldots, X_n$

= rank of $X_n$ at “birth”.

- Define the $r$-record times of $\{X_n\}$ by

$$L_0^{(r)} = 0, \quad L_n^{(r)} = \inf\{j > L_n^{(r)} : R_j = r\}$$

and the $r$-records are $\{X_{L_n^{(r)}}, n \geq 1\}$ which are points of PRM $(R(dx))$ by Ignatov’s theorem.
3.1. Facts:

- $\mathcal{R}_r$, the range of $M^{(r)}$ is
  \[ \mathcal{R}_r := \bigcup_{p=1}^{r} \{ X_{L_n^{(p)}}, n \geq 1 \}, \]
  By Ignatov’s theorem, this is a sum of $r$ independent PRM(R) processes and therefore the range of $M^{(r)}$ is PRM($rR$).

- $\mathcal{R}_r$, the range of $M^{(r)}$, converges as a random closed set in the Fell topology to $\mathcal{R}$, the support of the measure $R$:
  \[ \mathcal{R}_r \Rightarrow \mathcal{R}, \]
  as $r \to \infty$.

- $M^{(r)}$ jumps at time $k$ iff
  \[ R_k \in \{1, \ldots, r\}, \]
  so
  \[ \{ [M^{(r)} \text{ jumps at } k], k \geq r \} \]
  are independent events over $k$ and
  \[ P[M^{(r)} \text{ jumps at } k] = \frac{r}{k}. \]
3.1.1. Question

Is there anything corresponding to a stationary distribution in the Markov chain sense. Is there a non-trivial limit as $r \to \infty$ for $M^{(r)}$?
Contents

Undirected
Directed
Extremes