Static Random Network Models

Social and Economic Networks

MohammadAmin Fazli
ToC

- Static Random Network Models
  - Different static random network models
  - Some basic mathematics
  - Properties of random networks

- Readings:
  - Chapter 4 from the Jackson book
Static Random Network Models

• Static random network models:
  • Static random model refers to a typed model in which all nodes are established at the same time and then links are drawn between them according to some probabilistic rule
  • Important to capture features of SENs which are produced by random processes
  • Different random graph models:
    • Poisson networks & related models
    • Watts-Strogatz model for small-world networks
    • Markov graphs & p* networks
    • The Configuration model
    • The Expected Degree model
Poisson Networks

• $G(n, p)$:
  • Consider a set of nodes $N = \{1,2, \ldots, n\}$
  • Connect each pair $i, j$ of nodes with probability $p$
  • The expected number of edges: $\binom{n}{2}p$
  • The expected degree of nodes: $(n - 1)p$

• $G(n, M)$:
  • Choose $M$ edges out of all $\binom{n}{2}$ pair of nodes: $\binom{n}{2}$ choices
  • Number of edges: $M$
  • The expected degree of nodes: $\frac{M}{\binom{n}{2}} \times (n - 1) = \frac{2M}{n}$
Poisson Networks

• Binomial Distribution:
  • Consider a sequence of Bernoulli trials. What is the probability of \( m \) heads out of \( n \) flips?

\[
\binom{n}{m} p^m (1 - p)^{n-m}
\]

• Expected number of heads: \( np \)
• The variance: \( npq = np(1-p) \)
• Standard deviation: \( \sqrt{np(1 - p)} \)
Poisson Networks

• Degree distribution: Binominal distribution
  • The probability of having $d$ neighboring edges is equal to:

\[
P(d) = \binom{n-1}{d} p^d (1 - p)^{n-1-d}
\]

• Binominal distribution can be approximated by $\lambda = (n - 1)p$ for large $n$

\[
P(d) \approx \frac{e^{-\lambda} \lambda^d}{d!}
\]
Watts-Strogatz Model

• Consider a cycle and connect each node to its $2m$ nearest nodes
• Diameter: $\frac{n}{4}$
• Clustering Coefficient: $\frac{1}{2}$
• Diameter is high, while the clustering coefficient is also high
Watts-Strogatz Model

• Watts & Strogatz show that with a few random rewiring the diameter will be decreased a lot.
• We will speak about small-world models deeply later
Watts-Strogatz Model

$P = 0$  increasing randomness  $P = 1$

Cluster, Path Length

$L_p/L_0$  $C_p/C_0$
Markov Graphs & P* Networks

• Think about building a random graph in which the formation of the link $ij$ is correlated with formation of the links $jk$ and $ik$?

• Frank & Strauss method using Clifford & Hammersley theorem:
  • Build a graph $D$ whose nodes is the potential links in $G$
  • If $ij$ and $jk$ are linked in $D$, it means that there exist some sort of dependency between them
  • $C(D)$ is the set of $D$’s cliques
  • $I_A(G) = 1$ for $A \in C(D)$, $A \subseteq G$ (consider $G$ as a set of edges) and else $I_A(G) = 0$
  • The probability of a given network $G$ depends only on which cliques of $D$ it contains:

$$\log(\text{Pr}[G]) = \sum_{A \in C(D)} \alpha_A I_A(G) - c$$
Markov Graphs & P* Networks

• An example: a symmetric case
  • Build a random graph with controllability on the number of its edges \(n_1(G)\) and its triads \(n_3(G)\)
  • \(C(D)\) consists of \(n_3(G)\) triads and \(n_1(G)\) edges. So if we weight them equally we have:
    \[
    \log(\Pr(G)) = \alpha_1 n_1(G) + \alpha_3 n_3(G) - c
    \]
  • We can calibrate with different parameters to have different random networks with different number of triangles and edges.
    • \(\alpha_3 = 0\) is the Poisson networks case
The Configuration Model

• A sequence of degrees is given \((d_1, d_2, d_3, \ldots, d_n)\) and we want to build a random graph having these degrees

• We generate the following sequence of numbers

\[
\begin{align*}
&1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 2, 2, 2, 2, 2, \ldots \\
&\underbrace{n, n, n, n, n, n, n, n, n, n, n, n, n, n, n}_{d_1 \text{ entries}}, \underbrace{n, n, n, n, n, n, n, n, n, n, n, n, n, n, n}_{d_2 \text{ entries}}, \ldots
\end{align*}
\]

• Randomly pick two number of elements and connect corresponding nodes

• The result is a multigraph
An Expected Degree Model

• Form a link between node $i$ and node $j$ with probability
  \[
  \frac{d_i d_j}{\sum_k d_k}
  \]

• Self links are allowed

• The expected degree of node $i$ will be $d_i$

• Maximum of $d_i^2 < \sum_k d_k$
Configuration Model vs Expected Degree Model

- Consider the degree sequence $< k, k, \ldots, k >$

- In configuration model:
  - The probabilities of self links and multi links is negligible
  - The probability of a node to have degree $k$ will converge to

- In expected degree model:
  - The probability of a node to have degree $k$ will converge to
  
  \[
  \frac{e^{-k} (k)^k}{k!}
  \]
  
  whose maximum value is $1/2$.
  - Why? See the blackboard.
Distribution of the Degree of Neighboring Nodes

- Consider a given graph with degree distribution $P(d)$
- A related calculation $\tilde{P}(d)$: the probability that a randomly chosen edge has a (randomly chosen) neighbor with degree $d$
- $P(d) = \tilde{P}(d)$?
  - $P(1) = P(2) = \frac{1}{2}$
  - $\tilde{P}(1) = \frac{2}{3} \times \frac{1}{2} = \frac{1}{3}$
  - $\tilde{P}(2) = \frac{2}{3} \times \frac{1}{2} + \frac{1}{3} \times \frac{1}{3} = \frac{2}{3}$
- We can formulate $\tilde{P}(d)$

\[
\tilde{P}(d) = \frac{P(d)d}{\langle d \rangle}
\]

See the blackboard
Distribution of the Degree of Neighboring Nodes

• Consider the degree sequence <1,1,2,2,1,1,2,2,...>. Compare two cases
  • In random models such as the configuration model: The distribution of the neighboring nodes have the same distribution as $\tilde{P}(d)$ for all nodes.
  • In networks with correlation properties: The graph is highly segregated by degrees
Multiple & Self-Links

• In the configuration model the probability of self links and multiple links is negligible
  • The same is true for the expected degree model
• Theorem: Let $q_i^n$ denote the number of self or duplicate links & $Q_i^n = \Pr[q_i^n > 0]$. If a degree sequence $(d_1, d_2, ..., d_n)$ is such that $\max_{i \leq n} d_i \to 0$ then $\max_{i \leq n} Q_i^n \to 0$. 
Threshold Effects

- We define the function $t(n)$ as the threshold for the parameter $p(n)$ for the property $A(N)$ if

$$\Pr[A(N)|p(n)] \to 1 \text{ if } p(n)/t(n) \to \infty$$

and

$$\Pr[A(N)|p(n)] \to 0 \text{ if } p(n)/t(n) \to 0$$
Example

- $A(N) = \{g \mid d_1(g) \geq 1\}$
- In Poisson model the only model parameter is $p(n)$
- $\Pr[A(N) \mid p(n)] = 1 - (1 - p(n))^{n-1}$
- The threshold is $t(n) = \frac{1}{n-1}$, on the blackboard see why?
  - Hint: write $p(n) = \frac{r(n)}{n-1}$
A Summary of Thresholds for Poisson Random Network Model

• $t(n) = n^{-2}$: The first edge emerges

• $t(n) = n^{-1.5}$: The first component with three nodes emerges

• $t(n) = n^{-1}$: Cycles emerge as does the giant component

• $t(n) = \frac{\log(n)}{n}$: The network becomes connected
Threshold for Connectedness

• For Poisson Random Network Model
  • Erdos & Renyi Theorem: A threshold function for the connectedness of the Poisson random network is $t(n) = \frac{\log(n)}{n}$
    • Proof: See the blackboard
      • First we show that if $\frac{p(n)}{t(n)} \to 0$ then the probability that there exist isolated nodes tends to 1 and thus the network is clearly unconnected.
      • Second we show that if $\frac{p(n)}{t(n)} \to \infty$ then the chance of having any component of size less than $n/2$ tends to zero and thus the network is clearly connected
  • For other static random models it is very hard to find well defined thresholds
Threshold for Connectedness

- For the expected degree model:
  - $Vol^n = \sum_{i=1}^{n} d_i$ denote the total expected degree of nodes
  - The probability of a link between nodes $i$ and $j$ is $\frac{d_i d_j}{Vol^n}$
  - The probability that node $i$ is isolated
    \[
    \prod_j \left(1 - \frac{d_i d_j}{Vol^n}\right) \approx e^{-d_i}
    \]
  - The probability that no node is isolated
    \[
    \prod_i \left(1 - e^{-d_i}\right) \approx e^{-\sum_i e^{-d_i}}
    \]
  - If $\sum e^{-d_i} \to \infty$ isolated nodes will occur
  - If $\sum e^{-d_i} \to 0$ there will be no isolated node
Size of the Giant Component

• In the Poisson Random Network Model:
  • Assume that there exist a giant component and it consists $q$ fraction of nodes
  • Each node is in the giant component if it has no neighbor in the giant component. Thus approximately we have

$$1 - q = \sum_{d} (1 - q)^d P(d) = \sum_{d} \frac{e^{-(n-1)p}((n-1)p)^d}{d!} (1 - q)^d P(d)$$

• Thus since

$$\sum_{d} \frac{((n-1)p(1-q))^d}{d!} = e^{(n-1)p(1-q)}$$

• Thus $q = 1 - e^{-q(n-1)p}$
Threshold for the Emergence of the Giant Component

• The probability that the link connects two nodes that were already connected in a component with $s$ nodes is the probability that both of its end nodes lie in that component, which is proportional to $\left(\frac{s}{n}\right)^2$

$$\sum_i \left(\frac{s_i}{n}\right)^2 \leq \sum_i \left(\frac{s_iS}{n^2}\right) \leq \frac{S}{n}$$

• The portion of links that lie on cycles is of an order no more than $\frac{S}{n}$

• Before the threshold for the emergence of the giant component, the components are essentially tree.
Threshold for the Emergence of the Giant Component

• Define $\phi$ as the number of nodes which can be reached by the random process.

• Under some constraints we have

$$\phi = 1 + \sum_{d=1}^{\infty} (d - 1) \tilde{P}(d)\phi = 1 + \sum_{d=1}^{\infty} \frac{P(d)d}{\langle d \rangle} (d - 1)\phi$$

$$\phi = 1 + \frac{\langle d^2 \rangle - \langle d \rangle}{\langle d \rangle} \phi \quad \quad \phi = \frac{1}{2 - \frac{\langle d^2 \rangle}{\langle d \rangle}}$$

• The boundaries where $\phi$ is NOT well defined then the giant component emerges (the mentioned constraints come here). That is where $\langle d^2 \rangle > 2\langle d \rangle$.
Threshold for the Emergence of the Giant Component

• For Poisson random network model
  \[
  \begin{cases}
  \langle d \rangle^2 > 2 \langle d \rangle \\
  \langle d^2 \rangle = \langle d \rangle + \langle d \rangle^2 \\
  \Rightarrow \langle d \rangle^2 > \langle d \rangle \Rightarrow \langle d \rangle > 1 \Rightarrow t(n) = \frac{1}{n}
  \end{cases}
  \]

• For k-Regular network model
  \[k^2 = 2k \Rightarrow k = 2\]

• For Scale-Free network model
  • \( P(d) = cd^{-\gamma} \)
  • For \( \langle d^2 \rangle \) diverges for \( \gamma < 3 \)
Diameter Estimation

• Estimating Diameter for Random Networks
  • Choose a random node
  • The expected number of reachable nodes with 1 edge: $\langle d \rangle$
  • The expected number of reachable nodes with 2 edges:
    \[
    \langle d \rangle \sum_d (d - 1) \bar{P}(d) = \langle d \rangle \frac{\langle d^2 \rangle - \langle d \rangle}{\langle d \rangle}
    \]
  • The expected number of reachable nodes with k edges:
    \[
    \langle d \rangle \left( \frac{\langle d^2 \rangle - \langle d \rangle}{\langle d \rangle} \right)^{k-1}
    \]
Diameter Estimation

• Estimating Diameter for Random Networks
  • The total number of nodes is approximately:

\[
\sum_{k=1}^{\ell} \langle d \rangle \left( \frac{\langle d^2 \rangle - \langle d \rangle}{\langle d \rangle} \right)^{k-1}
\]

• Thus

\[
\langle d \rangle \left( \frac{\langle d^2 \rangle - \langle d \rangle}{\langle d \rangle} \right)^{\ell} - 1 = n - 1
\]

• Or

\[
\ell = \frac{\log \left[ (n - 1) \left( \langle d^2 \rangle - 2 \langle d \rangle \right) + \langle d \rangle^2 \right] - \log \left[ \langle d \rangle^2 \right]}{\log \left[ \langle d^2 \rangle - \langle d \rangle \right] - \log \left[ \langle d \rangle \right]}
\]
Diameter Estimation

• Estimating Diameter for Random Networks
  • In cases $\langle d^2 \rangle$ is much larger than $\langle d \rangle$
    \[
    \ell = \frac{\log[n] + \log[\langle d^2 \rangle] - 2 \log[\langle d \rangle]}{\log[\langle d^2 \rangle] - \log[\langle d \rangle]} = \frac{\log[n/\langle d \rangle]}{\log[\langle d^2 \rangle/\langle d \rangle]} + 1
    \]
  • In Poisson random network models where $\langle d^2 \rangle - \langle d \rangle = \langle d \rangle^2$
    \[
    \ell = \frac{\log\left((n - 1)\frac{\langle d \rangle - 1}{\langle d \rangle} + 1\right)}{\log(\langle d \rangle)}
    \]
Generating Functions

• Let $\pi(.)$ is a probability distribution over the set $\{0,1,2,\ldots\}$
• Define $G_\pi$ as the generating function related to $\pi$ as follows:

$$G_\pi(x) = \sum_{k=0}^{\infty} \pi(k)x^k$$

• $G_\pi(1) = 1$
• Taking a derivative from $G_\pi(x)$

$$G_\pi'(x) = \sum_{k=0}^{\infty} \pi(k)kx^{k-1}$$

• Thus

$$G_\pi'(1) = \sum_{k=1}^{\infty} \pi(k)k = E_\pi[k] = \langle k \rangle$$
Generating Functions

• More generally,

\[
\left( \frac{d}{dx} \right)^m G_\pi = \sum_{k=1}^{\infty} \pi(k) k^m x^k
\]

\[
E[k^m] = \langle k^m \rangle = \left( \frac{d}{dx} \right)^m G_\pi \bigg|_{x=1}
\]

• Assume that \( \pi_2(k) \) is the probability that the sum of two draws from \( \pi \) is \( k \). Thus

\[
\pi_2(k) = \sum_{i=0}^{k} \pi(i) \pi(k-i)
\]

• Thus

\[
G_{\pi_2}(x) = \sum_{k=0}^{\infty} \pi_2(k) x^k = \sum_{k=0}^{\infty} \sum_{i=0}^{k} \pi(i) \pi(k-i) x^k
\]
Generating Functions

• The relation between $G_{\pi_2}$ and $G_{\pi}$:

$$[G_{\pi}(x)]^2 = \left[ \sum_{k=0}^{\infty} \pi(k) x^k \right]^2 = \sum_{i,j} \pi(i) \pi(j) x^{i+j} = G_{\pi_2}(x)$$

• More generally,

$$G_{\pi_m}(x) = [G_{\pi}(x)]^m$$

• Consider a distribution $\pi$ that is derived by first randomly selecting a distribution from a series of distributions $\pi_1, \pi_2, \pi_3, \ldots$ picking each with probability $\gamma_i$.

$$G_\pi = \sum_i \gamma_i G_{\pi_i}$$
Generating Functions

• Define $\overline{\pi}(k) = \pi(k - 1)$

$$G_{\overline{\pi}}(x) = \sum_{k=1}^{\infty} \pi(k - 1)x^k = xG_{\pi}(x)$$

• Let’s define the generating function for $\tilde{P}(d) = \frac{dP(d)}{\langle d \rangle}$

$$G_{\tilde{P}}(x) = \sum_{k=0}^{\infty} \tilde{P}(d)x^d = \sum_{k=0}^{\infty} \frac{P(d)d}{\langle d \rangle}x^d = \frac{xG'_{P}(x)}{G'_{P}(1)}$$
Distribution of Component Sizes

• Randomly choose an edge and then one of its nodes. What is the distribution of the number nodes of the component including this edge ($Q$)?
  • There is probability $\tilde{P}(d)$ that the node at the end of the randomly selected edge has degree $d$
  • It has $d - 1$ edges emanating from this node
  • The number of vertices found in the owner component is the sum of nodes found by these edges, plus one. Thus, the generating function of additional nodes found through the node if it happens to have degree $d$ is

$$G_Q(x) = x \sum_d \tilde{P}(d)[G_Q(x)]^{d-1}$$
Distribution of Component Sizes

• Randomly choose an edge and then one of its nodes. What is the distribution of the number nodes of the component including this edge ($Q$)?
  • With some calculations (see the blackboard) we have the following:
    \[ G_Q(x) = \left( xG_P \left( G_Q(x) \right) \right)^{1/2} \]
  • and
    \[ G'_Q(x) = \frac{1}{2} \left( xG_P \left( G_Q(x) \right) \right)^{-1/2} \left( G_P \left( G_Q(x) \right) + xG'_P \left( G_Q(x) \right) G'_Q(x) \right) \]
  • For 1 we have
    \[ G'_Q(1) = \frac{1}{2 - \langle d^2 \rangle} \]
Distribution of Component Sizes

• What is the distribution of the component including a randomly chosen vertex (H)?
  • Using $G_Q$ we have
    $$G_H(x) = x \sum_d P(d) \left(G_Q(x)\right)^d = x G_P \left(G_Q(x)\right)$$
  • Thus
    $$G'_H(1) = 1 + G'_P(1)G'_Q(1) = 1 + \frac{\langle d \rangle^2}{2 \langle d \rangle - \langle d^2 \rangle}$$