Bipartite Networks

Kiran Kumar(04CS1011)

May 6, 2008

1 Introduction

A bipartite network is a graph which connects two distinct sets (or partitions) of nodes, which we will refer to as the top and the bottom set. An edge in the network runs between a pair of a top and a bottom node but never between a pair of top or a pair of bottom nodes.
Typical examples of this type of networks include collaboration networks such as the movie-actor, article-author, and board-director network. In the movie-actor network, for instance, the movies and actors are the elements of the top and the bottom set respectively, and an edge between an actor a and a movie m indicates that a has acted in m. The actors a and a' are collaborators if both have participated in the same movie, i.e., if both are connected to the same node m'. The concept of collaboration can be extended to include so diverse phenomena represented by bipartite networks as the city-people network, in which an edge between a person and a city indicates that the person has visited that particular city, the word-sentence, bank-company or donor-acceptor network, which accounts for injection and merging of magnetic field lines.

2 The Growth Model

We consider the case in which the top partition grows with time while the number of nodes in the bottom partition N is kept constant. We grow the network in the following way. At each time step a new node is incorporated to the top set. Then, μ edges are connected from the new node to the nodes in the bottom set. The probability of attaching a new edge to the bottom

node i is $A(k_i^t)$, where k_i^t refers to the degree of the bottom node i at time t. We refer to $A(k_i^t)$ as the attachment kernel and define it as:

$$A(k_i^t) = \frac{\gamma k_i^t + 1}{\sum_{j=1}^{N} \gamma k_i^t + 1} \text{ —- (1)}$$

where the sum in the denominator runs over all bottom nodes, and γ is a model parameter which controls the relative weight of random to preferential attachment.

The stochastic process can be performed in such a way that the attachment of the μ incoming nodes is done sequentially, i.e., one edge is attached per time step. This implies that the denominator of $A(k_i^t)$ has to be updated for each incoming node (and hence an edge), and that t refers to the event of incorporating a new edge to the bottom set. Alternatively, the attachment of the μ new edges can be done in parallel. This implies that the new μ edges have all the same probability of attaching to bottom node i. In this case, t refers to the event of incorporating a new node to the top set.

There are two significant limits to consider: $\gamma = 0$ and $\gamma \longrightarrow \infty$. For $\gamma = 0$, Eq. (1) reduces to $A(k_i^t) = 1/N$, which implies that all bottom nodes have the same probability of being selected by an incoming edge. This limit corresponds to pure random attachment. For $\gamma \longrightarrow \infty$ Eq. (1) reduces to,

$A(k_i^t) = \frac{k_i^t}{\sum_{j=1}^{N} k_i^t}$ Which means that higher degree bottom nodes have higher probability of being selected. This case corresponds to pure preferential attachment.

3 Evolution equation for sequential attachment

Now we aim to derive an evolution equation for the degree distribution of the bottom nodes. We focus on sequential attachment. Let $p_{k,t}$ be the probability of finding a randomly chosen bottom node with degree k at time t. We recall that t refers to the t-edge attachment event. $p_{k,t}$ can be defined as $p_{k,t} = (n_{k,t})/N$, where $n_{k,t_{avg}}$ refers to the number of nodes in the bottom set with degree k at time t, and avg denotes ensemble average, i.e. average over realizations of the stochastic process described above. We express the evolution of $p_{k,t}$ in the following way:

$$p_{k,t+1} = (1 - A(k,t))p_{k,t} + A(k-1,t)p_{k-1,t} \text{ —- (2)}$$

where A(k,t) refers to the probability that the incoming edge lands on a bottom node of degree k. A(k,t) can be easily derived from Eq.(1) and takes the form:

$$A(k,t) = \frac{\gamma k+1}{\gamma t+N} \text{ —- (3)}$$

The reasoning behind Eq.(2) is the following. The probability of finding a bottom node with degree k at time $t+1$ decreases due to the number of nodes, which have a degree k at time t and receive an edge at time t+1 therefore acquiring degree $k+1$, i.e., $A(k,t)p_{k,t}$. Similarly, this probability increases due to the number of nodes that at time t have degree $k-1$ and receives an edge at time $t+1$ to have a degree k, i.e., $A(k-1,t)p_{k-1,t}$. Hence the net increase in the probability can be expressed as in Eq.(2).
According to what was done in the stochastic simulations, we assume that at time $t = 0$ all bottom nodes have zero degree, which implies the initial condition $p_{k,t=0} = \rho_{k,0}$, where $\rho_{k,0}$ is the Kronecker delta function.

3.1 Exact solution for sequential attachment

Eq.(2)can be written as:

$$P_{t+1} = M_t P_t = [\prod_{\tau=0}^{t} M_\tau]P_0 \text{ — (4)}$$

where p_t denotes the degree distribution at time t and is defined as $p_t = [p_{0,t}p_{1,t}p_{2,t}...]^T$, p_0 is the initial condition expressed as $p_0 = [100...]^T$, and M_τ is the evolution matrix at time τ.
Since our initial condition is a vector with zeros in all position except in the first one, all the relevant information, i.e., the degree distribution of the bottom nodes, is in the first column of $[\prod_{\tau=0}^{t}]$. A close inspection to the evolution of this column, explicitly using Eq. (3), reveals that the $k-th$ element of it, which corresponds to $p_{k,t}$ can be expressed as:

$$pk,t = (tk)\frac{\prod_{i=0}^{k-1}(\gamma i+1)\prod_{j=0}^{t-1-k}(N-1-\gamma j)}{\prod_{m=1}^{t-1}(\gamma m+N)} \text{ —- (6)}$$

for $k <= t$, and $p_{k,t} = 0$ for $k > t$, and where we have defined $\prod_{i=0}^{-1}(...) = 1$, and (tk) refers to the combinatorial number $t!/[(t-k)!k!]$.
Eq.(6) is the exact solution of Eq.(2) using as initial condition $p_{k,t=0} = \delta_{k,0}$, i.e., the analytical expression of the degree distribution of the bottom nodes

when sequential attachment is applied.
In the limit of $\gamma = 0$ Eq.(6) reduces to:

$$p_{k,t} = (tk)(\tfrac{1}{N})^k(1 - \tfrac{1}{N})^{t-k} \text{ — (7)}$$

for $k <= t$, and $p_{k,t} = 0$ for $k > t$. In other words, Eq.(7) is the solution of the sequential problem when pure random attachment is applied.

3.2 Parallel Attachment

We focus on parallel attachment, i.e., when more than one edge are added per time step. We do not aim to derive an exact analytical expression for the degree distribution of this problem but provide a reasonable approximation. We recall that for parallel attachment t refers to the event of incorporating a new top node. We assume that $\mu << N$ and expect Eq. (2) to be a good approximation of the process after replacing $A(k,t)$ with $A_p(k,t)$. We define $A_p(k,t)$ as

$$A_p(k,t) = \tfrac{(\gamma k+1)\mu}{\gamma\mu t+N}$$

The term μ in the denominator appears since in this case the total degree of the bottom nodes at any point in time is μt rather than t as in Eq. (3). The numerator contains a μ since at each time step there are μ edges that are being incorporated into the network rather than a single edge.
It is important to mention here that Eq.(2) cannot exactly represent the stochastic parallel attachment because it explicitly assumes that in one time step a node of degree k can only get converted to a node of degree $k+1$. Clearly, the incorporation of μ edges in parallel allows the possibility for a node of degree k to get converted to a node of degree $k+\mu$. The correct expression for the evolution of $p_{k,t}$ reads:

$$pk, t+1 = (1 - \textstyle\sum_{i=1} \mu A(k,i,t))p_{k,t} + \sum i = 1\mu A(k-i,i,t))p_{k-i,t} \text{ — (9)}$$

where $A(k,i,t)$ represents the probability at time t of a node of degree k of receiving i new edges in the next time step. We expect Eq.(2) to be a good approximation of Eq. (9) when $A(k,1,t) >> A(k,i,t)$ where $i > 1$.
The solution of Eq.(2) with the attachment kernel given by Eq.(8) reads:

$$pk, t = (tk)\frac{\prod_{i=0}^{k-1}(\gamma i+1)\prod_{j=0}^{t-1-k}(\frac{N}{\mu}-1-\gamma j)}{\prod_{m=1}^{t-1}(\gamma m+\frac{N}{\mu})} \text{ —- (10)}$$

We expect Eq.(10) to approximate the degree distribution of the stochastic process with parallel attachment for $\mu << N$. This means that we cannot expect the approximation to hold for large values of γ or $/N$.
In the limit of random attachment, i.e., $\gamma = 0$, Eq.(10) becomes

$$p_{k,t} = (tk)(\tfrac{\mu}{N})^k(1 - \tfrac{\mu}{N})^{t-k} \text{ – (11)}$$

For large values of γ, the approximation fails. However, for $\mu = 1$ Eq.(10) reduces to Eq.(6), which in this case is the exact solution, and then the theory works for all values of γ.