

Opinion-dependent Rewiring Model

Sam Magura, Vitchyr Pong

June 26, 2012

1 Model Definition

The process operates on a graph $G = (V, E)$. Each vertex $v = \langle v_1, v_2, \dots, v_D \rangle$ where v_1, v_2 , and so on represent the opinions of v . Initially, each vertex is placed randomly in the *opinion space* $\Omega = [0, 1]^D$. The distance function $d(x, y)$ gives the *opinion distance* between two vertices x and y . We use

$$d(x, y) = |x_1 - y_1| + |x_2 - y_2| + \dots \quad (1)$$

Let the constant $p = 1/D$ so that $p d(x, y) \leq 1$ for any possible $\{x, y\}$.

Each iteration, do the following.

1. Randomly choose an edge from the graph.
2. Apply a random orientation (x, y) to the edge.
3. Randomly choose a vertex z from the set of all vertices that are neither x nor adjacent to x .
4. With probability

$$p d(x, y) \min \left(1, \frac{d(x, y)}{d(x, z)} \right), \quad (2)$$

remove the edge $\{x, y\}$ and add the edge $\{x, z\}$.

In Expression 2, $p d(x, y)$ represents the chance that $\{x, y\}$ is marked for removal and $\min(1, \frac{d(x, y)}{d(x, z)})$ represents the chance that the addition of $\{x, z\}$ is accepted. We say “ $\{x, y\}$ is marked for removal” since the edge is only removed if the addition of $\{x, z\}$ is also accepted. Thus, the greater the opinion distance of the original edge $\{x, y\}$, the greater the chance that that edge is marked for removal. The addition of the new edge $\{x, z\}$ is always accepted if the new distance is less than the original distance. However, if the new distance is greater, the addition is accepted with probability $d(x, y)/d(x, z)$.

In many network models, multiple edges are allowed because they are relatively unlikely in large networks. However, if multiedges are allowed in this model, many edges will accumulate between pairs of vertices that have especially low opinion distances. Because of this, we prevent the creation of multiedges by refusing to select a z that is already adjacent to x in Step 3.

2 Stationary Distribution

The model is a Markov Chain because the probability of transitioning from a state G to a state H is dependent only G and H .

Definition 1. *Graphs $G = (V, E_G)$ and $H = (V, E_H)$ are the same state in the Markov Chain if and only if $E_G = E_H$.*

Transition probability Consider states G and H such that $\{x, y\}$ is in G but not H , and $\{x, z\}$ is in H but not G . For a transition from G to H , the following must occur.

1. The edge $\{x, y\}$ is selected. This occurs with probability $\frac{1}{m}$.
2. The orientation (x, y) is chosen. This occurs with probability $\frac{1}{2}$.
3. Vertex z is selected in Step 3 of the model definition. This occurs with probability $\frac{1}{m-k(x)-1}$ where $k(x)$ gives the degree of x , since z is selected randomly from the set of all vertices that are neither x nor adjacent to x .
4. The rewiring is accepted. This occurs with probability $p \frac{d(x, y)}{d(x, z)} \min(1, \frac{d(x, z)}{d(x, y)})$.

Therefore the transition probability is

$$P(G, H) = \frac{p \frac{d(x, y)}{d(x, z)} \min(1, \frac{d(x, z)}{d(x, y)})}{2m(m-k(x)-1)}. \quad (3)$$

Theorem 1. *The detailed balance condition is satisfied by the stationary distribution*

$$\pi(G) = b \prod_e \frac{1}{d(e)^2} \quad (4)$$

where b is some normalization coefficient, e is any edge in graph G , and $d(e)$ gives the distance between the endpoints of edge e .

Proof. The detailed balance condition is

$$\pi(G) P(G, H) = \pi(H) P(H, G). \quad (5)$$

Without loss of generality, suppose $d(x, y)$ is greater than $d(x, z)$. Substituting in the transition probabilities from Equation 3 and the proposed stationary distribution π , we have

$$\frac{p \frac{d(x, y)}{d(x, z)} \min(1, \frac{d(x, z)}{d(x, y)})}{2m(m-k(x)-1)} \left(b \prod_{e \in E_G} \frac{1}{d(e)^2} \right) = \frac{p \frac{d(x, z)}{d(x, y)} \min(1, \frac{d(x, y)}{d(x, z)})}{2m(m-k(x)-1)} \left(b \prod_{e \in E_H} \frac{1}{d(e)^2} \right). \quad (6)$$

This simplifies to

$$d(x, y)^2 \prod_{e \in E_G} \frac{1}{d(e)^2} = d(x, z)^2 \prod_{e \in E_H} \frac{1}{d(e)^2} \quad (7)$$

since the degree of x does not change. Because the only difference between G and H is that G has an edge $\{x, y\}$ that H does not and H has an edge $\{x, z\}$ that G does not, the detailed balance condition holds for the proposed stationary distribution π . \square

3 Constructive Model

Theorem 2. *When the rewiring model has reached the steady state, the probability that an edge e^* is present is approximately*

$$\frac{c}{d(e^*)^2 + 1} \quad \text{with} \quad c = \frac{m}{\sum_{e \in K_n} \frac{1}{d(e)^2 + 1}}$$

where $n = |V|$, $m = |E|$, and K_n is the complete graph on n vertices.

Proof. For a constructive model in which each edge e is independently present with probability $P(e)$, the probability that a particular graph G is produced is

$$\mu(G) = \prod_{e \in G} \frac{P(e)}{1 - P(e)} \prod_{e \in K_n} (1 - P(e)). \quad (8)$$

We want to find the function $P(e)$ such that $\mu(G)$ is approximately equal to $\pi(G)$. Because the expressions for $\mu(G)$ and $\pi(G)$ both contain a product over all edges $e \in G$ multiplied by some factor which is independent of G , we can write

$$\frac{P(e)}{1 - P(e)} \propto \frac{1}{d(e)^2} \quad \longrightarrow \quad P(e) = \frac{c}{d(e)^2 + 1}. \quad (9)$$

Since each state in the rewiring model contains exactly m edges, the expected the number of edges of a graph produced by the constructive model should be m :

$$\sum_{e \in K_n} P(e) = m. \quad (10)$$

By substituting Equation 9 into Equation 10, we reach an expression for the coefficient c :

$$c = \frac{m}{\sum_{e \in K_n} \frac{1}{d(e)^2 + 1}}. \quad (11)$$

□

Theorem 3. *For large n ,*

$$c = \frac{2m}{n^2} \left[\int_{\Omega} \int_{\Omega} \frac{du dv}{d(u, v)^2 + 1} \right]^{-1} \quad (12)$$

where Ω is the opinion space $[0, 1]^D$.

Proof. The sum in the denominator of our expression for c (Equation 11) can be rewritten as

$$\frac{1}{2} \sum_u \sum_{v \neq u} \frac{1}{d(u, v)^2 + 1} \quad (13)$$

where u and v are any vertices in the graph. The $1/2$ accounts for the fact that both edge (x, y) and edge (y, x) are included in the sum. For large n , vertices are evenly distributed within the opinion space since they are placed randomly. Therefore, there are $n \cdot du_1 \cdot du_2 \cdots du_D$ vertices in an interval with dimensions du_1, du_2, \dots, du_D . Because the sums in Expression 13 are computed

over n and $n - 1$ vertices respectively, and n is large, we can express the two sums as a double integral:

$$\frac{1}{2} n^2 \int_{\Omega} \int_{\Omega} \frac{du dv}{d(u, v)^2 + 1}. \quad (14)$$

□

Proposition 1. *For $D = 1$ and large n ,*

$$c = \frac{4m}{n^2 (\pi - \ln 4)}.$$

Proof. For $D = 1$, the double integral in Theorem 3 reduces to

$$\frac{1}{2} n^2 \int_0^1 \int_0^1 \frac{du_1 dv_1}{(u_1 - v_1)^2 + 1}. \quad (15)$$

We evaluate this integral to reach our result. □

4 Expected Degree

Here we derive an expression for the expected degree of a vertex $x = \langle x_1, x_2, \dots, x_D \rangle$. For simplicity, we present our argument for the one-dimensional case and then generalize the result for any dimension D .

4.1 One-dimensional case

Proposition 2. *The expected degree of a vertex $x = x_1$ is*

$$k(x) = c n \left(\tan^{-1} x + \tan^{-1}(1 - x) \right). \quad (16)$$

Proof. The expected number of vertices in an interval $I = [x + s, x + s + ds]$ is $n ds$. From Theorem 2, we see that the expected number of edges $\{x, y\}$ where $y \in I$ is $\frac{k}{s^2 + 1} n ds$ since $s = d(x, y)$. Therefore the degree of x is given by

$$k(x) = c n \left(\int_{-x}^0 \frac{ds}{s^2 + 1} + \int_0^{1-x} \frac{ds}{s^2 + 1} \right). \quad (17)$$

Evaluating the integrals yields our result. □

Proposition 3. *The expected degree distribution is given by*

$$N(k) = n \frac{dx}{dk}. \quad (18)$$

Proof. Let $x(k)$ give the opinion of a vertex with expected degree k and always return a value in $[0, 1/2]$. With $dk \rightarrow 0$, all vertices in the interval $[x(k), x(k + dk)]$ have expected degree k . Because vertices are spread evenly within the opinion space, the number of vertices that have

degree k is proportional to $dx = x(k + dk) - x(k)$. We claim that the proportionality constant is $\frac{n}{dk}$. If this constant is correct, the relationship

$$\int_0^{\infty} N(k) dk = n \tag{19}$$

will hold. Substituting $n \frac{dx}{dk}$ for $N(k)$ gives us

$$\int_0^{\infty} N(k) dk = n \int_0^1 dx = n. \tag{20}$$

The proposed proportionality constant is correct. □