Accounting for Changes in Existing Connections in the Preferential Deletion Model for Web-Like Networks

Carlos Santiago Bañón

College of Engineering and Computer Science - University of Central Florida - Orlando, FL

Abstract

This paper presents a new *Preferential Deletion Model with Changes in Existing Connections (PDCModel)* that can be used to better study the behavior of real-life social networks. It uses the *Preferential Deletion Model* defined by Narsingh Deo and Aurel Cami in their 2005 paper *Preferential Deletion in Dynamic Models of Web-Like Networks* and expands upon it. We then analyze a dynamic graph randomly generated by our new *PDCModel*, and explore its number of nodes, number of edges, average degree, and degree distributions. When analyzing this new degree distribution, it is evident that it peaks at a degree of $k = 6$, which differs significantly with the original model. Therefore, we can conclude that the average number of connections in social accounts actually peaks at $k = 6$, and not $k = 1$ as presented by Deo and Cami in their paper.

Keywords: Web-Like Networks, Interconnection Networks, Dynamic Random Graph Modeling, Preferential Node Deletion, Scale-Free, Power-Law Degree Distribution, Social Networks, Changes in Existing Edges

1. Introduction

When studying networks, the term *web-like* is often used to refer to tangible, real networks that are not random in nature. These networks can have many applications in various fields, such as telephone networks, neural networks, and social networks. This paper focuses on the latter. The study of how social networks behave is essential for understanding today's world, as there is a parallel between real-life connections from person to person and the same connections in the context of social networks such as Facebook, Instagram, Twitter, and others.

To study social networks, many models have been proposed throughout the years. All of these employ the study of graph theory, a very unique, yet essential area of study within the field of computer science. To represent these networks using graphs, we can simply use nodes as accounts in the network and edges as connections, or friendships between these accounts.

This paper provides a new model that can be useful in the study of these networks. In order to arrive to the creation of our new model, another interesting model developed by Narsingh Deo and Aurel Cami in 2005 was first examined. Their model, called the *Preferential Deletion Model*, seeks to also provide a thorough understanding of real-life networks, and is the basis of our new model.[1]

In their paper *Preferential Deletion in Dynamic Models of Web-Like Networks*, they include a very important discussion that can help understand how to create such models. They first state that the Erdös-Rényi model does not provide an accurate picture for these networks, which is a model that develops completely randomized dynamic graphs.[1] Therefore, they established their model with preferential deletion, which is defined in Section 3.

However, their model only accounts for the creation and deletion of nodes. It does not study the case where existing connections between nodes are changed throughout the life cycle of the network. For example, modern social networks give users the ability to friend existing accounts. Therefore, the new model proposed in this paper seeks to evaluate the case where new connections are created between existing nodes.

This new model, described as a *Preferential Deletion Model with Changes in Existing Connections*, is defined in Section 4, and the rest of the paper serves to compare both models, particularly their degree distributions.

2. Notation

The notation used throughout this paper is outlined in Figure 1.

Symbol	Meaning
$G(V, E)$	A graph with set of nodes V and set of edges E
$n(G)$ or n	Number of nodes in G
$m(G)$ or m	Number of edges in G
$d(u)$	Degree of node u
$\bar{d}(G)$	Average degree of graph G
G_t	Graph obtained after the tth step of a random graph generating procedure
n_t	Number of nodes in G_t
m_t	Number of edges in G_t
$d_t(u)$	Degree of node u in G_t
$\bar{d}_t$	Average degree of G_t
$N_{k,t}$	Number of nodes of degree k in G_t
$N_{k,t}^{(1)}$	Number of neighbors of degree k of a node chosen for deletion from G_t
$N_{k,t}^{(1)}(u)$	Number of neighbors of degree k of node u in G_t
$\mathbb{P}[k]$	Fraction of nodes of degree k
$\mathbb{P}_t[u]$	Probability of vertex u being selected as an end-node of a new edge or for deletion during the tth step of the random graph process
$\mathbb{P}[A]$	Probability of event A
$\mathbb{E}[X]$	Expectation of random variable X

Figure 1: List of symbols.

3. Definition of the Preferential Deletion Model (*PDModel*)

Let G_1 be a graph that begins with a single node and a self-loop. For each discrete time step $t + 1$, $t > 0$, either one of these processes can occur:

3.1. Birth Process

Deo and Cami define the birth process as follows: with probability p, a new node is added, together with a new edge incident on it. The other end-node u of the new edge is chosen preferentially from among all the existing nodes based on a *linear preferential attachment rule*.[1]

This linear preferential attachment can be defined as:

$$\mathbb{P}_{t+1}[u] = \frac{d_t(u)}{\sum_{w \in V_t} d_t(w)} = \frac{dt_u}{2m_t}. \tag{1}$$

3.2. Death Process

If a new node is not created, then a node will be deleted from the list. This process is defined as follows: with probability $q = 1 - p$, a node u is chosen for deletion along with all the edges incident on it

in G_t. To make small-degree nodes more likely candidates for deletion than the higher-degree ones, node u is chosen according to the probability distribution defined below.[1]

$$\mathbb{P}_{t+1}[u] = \frac{n_t - d_t(u)}{n_t^2 - 2m_t} \tag{2}$$

Finally, they state that, if during any step $t < 0$, the graph becomes empty, then it starts anew with a single node and a self-loop.

4. Definition of the Preferential Deletion Model with Changes in Existing Nodes (*PDCModel*)

To define this new model, extensive research was made to paint an accurate picture of how social networks behave and evolve.

Therefore, this new model uses the original *Preferential Deletion Model* as a foundation and builds on it. To account for new connections, the following question had to be answered first: *What is the probability that a given two people will meet?* In their paper *The Structure of Growing Social Networks*, Emily M. Jin, Michelle Girvan, and M.E.J. Newman provide an excellent answer to this question, formally providing a thorough mathematical answer.[2]

Therefore, the probability per unit time p_{ij} of a given two people i, j meeting is:

$$p_{ij} = f(d_i)f(d_j)g(z_{ij}), \tag{3}$$

where d_i and d_j are the degrees of i and j, respectively, and z_{ij} is their number of mutual connections.

Further, they use the *Fermi* function to define $f(d_i)$. In the world of physics, the *Fermi* function models the probability that an electron energy state will be occupied at a particular temperature T.[2] However, they found that it can be very useful in this context. Thus, they formally define $f(d_i)$ as

$$f(d_i) = \frac{1}{e^{\beta(d_i - d*) + 1}}, \tag{4}$$

where $d*$ is a parameter that makes the equation fall sharply if large, and β is a temperature-like arbitrary value that controls such sharpness.

As for $g(z)$, the function serves to show the increase we can expect in the likelihood that two people if they have one or more mutual friends, and is defined as such:

$$g(z) = 1 - (1 - p_0)e^{-\alpha z}, \tag{5}$$

where p_0 represents the probability of two people meeting with no mutual acquaintances, and α is a parameter that controls the rate at which the function increases.

This new model creates a new degree distribution that presents a more accurate representation of modern-day *web-like* structures as they relate to social media networks. Further, this paper also analyzes the expected number of nodes, the number of neighbors, the average degree of a graph G created using this new model, and its degree distribution.

5. Number of Nodes

Upon using this model, the first important conclusion drawn is that the expected number of nodes in graphs G_1 and G_2 created using the *PDModel* and the *PDCModel*, respectively, is roughly the same. This suggest that the order of growth for both is:

$$\mathbb{E}[n_t] = \Theta[(p - q)t].^{[1]} \tag{6}$$

Further, as the original paper discusses, the expected number of nodes in both resulting random graphs seldom vanish. That is, there is a very low probability $\mathbb{P}[n_t = 0]$, that the graph will disappear after some step $t > 0$.[1]

To better visualize this, Figure 2 shows the expected number of nodes obtained from the new *PDCModel*, and Figure 3 shows a comparison between both the old *PDModel* and the new *PDCModel*.
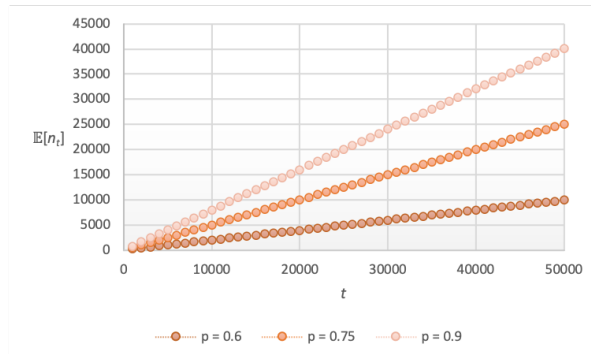
Figure 2: Expected number of nodes using the new *PDCModel*.

6. Number of Edges

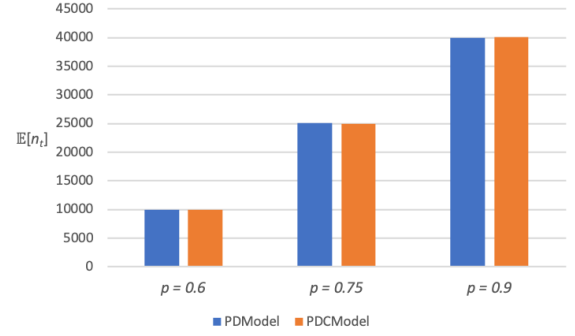When analyzing the expected number of edges when generating a graph with the new *PDCModel*, this is

Figure 3: Comparison of the expected number of nodes using both the *PDModel* and the new *PDCModel*.

where the first major change can be observed. Overall, the new model provides graphs with a much higher number of edges. This is due to the nature of the new model, as it continually connects existing nodes throughout the life cycle of the graph. Figure 4 shows how the number of expected edges as t increases.

In addition, Figure 5 shows a comparison between the old and new models. As it is evident, the expected number of nodes is significantly higher in the new *PDCModel*.
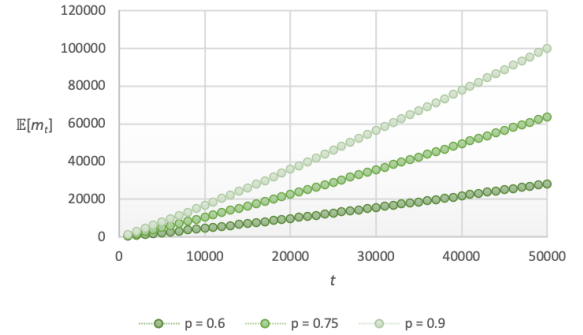
Figure 4: Expected number of edges using the new *PDCModel*.

7. Degree Distribution in the First Neighborhood of the Deleted Node

We then turn to the degree distribution of a random graph G generated with the new model. First, we focus on the degree distribution in the first neighborhood of a deleted node in G_t. Deo and Cami formally define this idea as the expected number of neighbors of degree k of a node chosen for deletion from G_t.[1] This is where the new model also significantly differs from the old model.
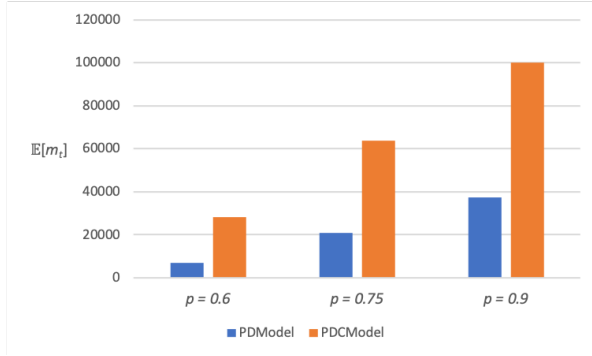
3

Figure 5: Comparison of the expected number of edges using both the *PDModel* and the new *PDCModel*.

Figure 7: Comparison of the expected number of neighbors of degree k of a node chosen for deletion using both the *PDModel* and the new *PDCModel*.

Figure 6 shows the expected number of neighbors of degree k of a node chosen for deletion in a graph generated by the new *PDCModel*. An interesting observation is that the peak in the expected number of neighbors does not occur in degree $k = 1$, as it does in the orginal *PDModel*. Instead, it peaks at roughly $k = 6$, and subsequently falls sharply. This is again due to the nature of the model itself. It can be concluded that most nodes peak at a degree of at most $k = 6$, and thus it can be difficult to find many nodes with more connections than that.
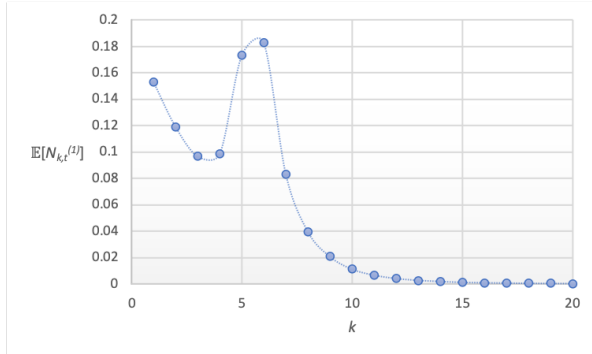
The cumulative degree distribution obtained by the new model can be seen in Figure 8, and it is evident that the same peak phenomenon occurs. The number of nodes of degree k seems to peak at around $k = 6$, again due to the definition of the model itself. Figure 9 then shows a comparison between the old and new models.

Figure 8: Cumulative degree distribution of the graph generated by the new *PDCModel*.

Figure 6: The expected number of neighbors of degree k of a node chosen for deletion using the new *PDCModel*.

Further, Figure 7 shows a comparison of the expected number of neighbors with graphs generated by both the *PDModel* and the *PDCModel*. Here, it is very evident that the peaks are different, due to the definition of both models.

8. Degree Distribution

In terms of the degree distribution of the new *PDCModel*, the results are very similar to those in Section 7.
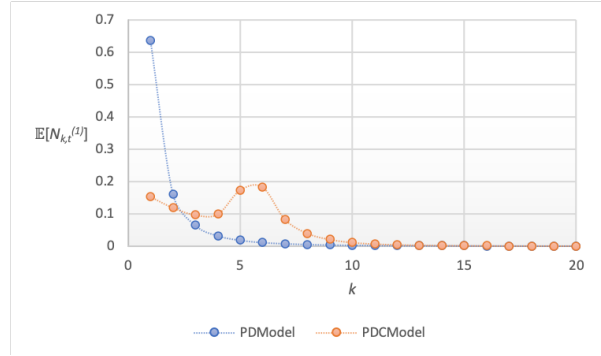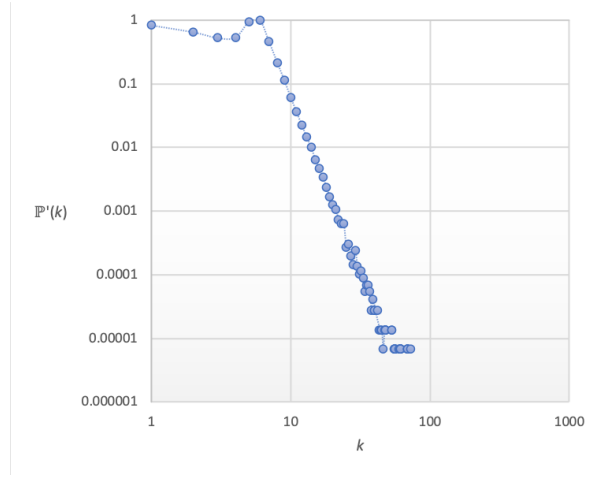
9. Average Degree and Number of New Connections

Another interesting difference we can see between both models is the average degree of a given node in the graphs generated by both models. The average degree is much higher in the graph generated by the *PDCModel*. Figure 10 shows this relationship.

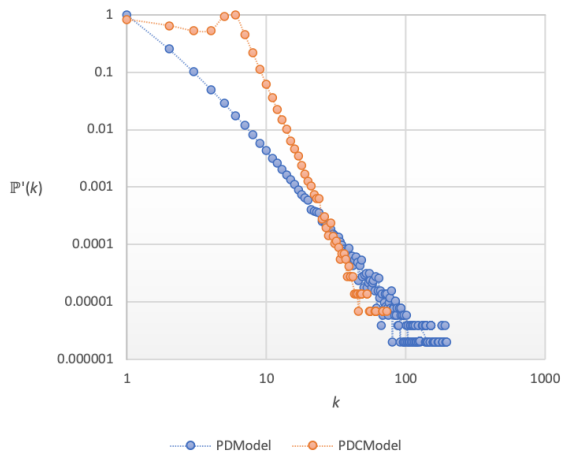Finally, we can take a look at the number of new connections made by the model. As the value of p in-

4

Figure 9: Comparison of the cumulative degree distributions in graphs generated by both the *PDModel* and the *PDCModel*.
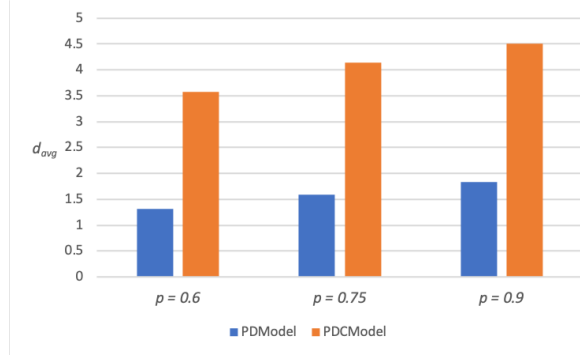

Figure 10: Comparison of the average degree in graphs generated by both the *PDModel* and the new *PDCModel*.

creases, we can see a sharp increase in the number of additional connections made between existing edges using the *PDCModel*. Figure 11 helps visualize this.

10. Conclusion

With this paper, we can conclude that indeed the new *PDCModel* produces different results when compared to the original *Preferential Deletion Model* developed by Deo and Cami. This new model presents a degree distribution that paints a much more accurate picture of the behavior of accounts in social networks. Dynamic graphs generated with the new model produce degree distributions that peak at around $k = 6$, which differs significantly with the original model. Therefore, we can conclude that the average number of connections in social accounts actually peaks at $k = 6$, and not $k = 1$ as presented by Deo and Cami in their paper. This then
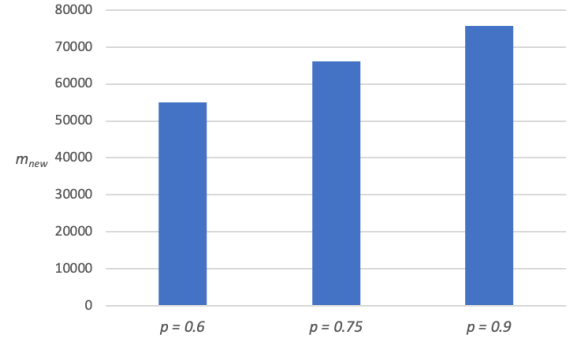

Figure 11: The number of extra edges created by the new *PDCModel*.

establishes this model as a sound way of dynamically generating random graphs for further study of *web-like* networks.

11. References

[1] N. Deo, A. Cami, Preferential deletion in dynamic models of web-like networks, in: Information Processing Letters, Volume 102, Issue 4, 2007, Pages 156-162.

[2] E.M. Jin, M. Girvan, M.E.J. Newman, The structure of growing social networks, in: American Physical Society, Volume 64, Issue 4, 2001, Pages 1-9.