

EvoCut : A new Generalization of Albert-Barabási Model for Evolution of Complex Networks

Shailesh Kumar Jaiswal, Manjish Pal, Mridul Sahu, Prashant Sahu, Amal Dev
Sarma National Institute of Technology Meghalaya
Shillong, India

shaileshsr141@gmail.com, manjishster@nitm.ac.in, mridulsahu01@gmail.com, prashant13sahu@gmail.com,
amal.sarma@nitm.ac.in

Abstract—With the evolution of social networks, the network structure shows dynamic nature in which nodes and edges appear as well as disappear for various reasons. The role of a node in the network is presented as the number of interactions it has with the other nodes. For this purpose a network is modeled as a graph where nodes represent network members and edges represent a relationship among them. Several models for evolution of social networks has been proposed till date, most widely accepted being the Barabási-Albert [1] model that is based on *preferential attachment* of nodes according to the degree distribution. This model leads to generation of graphs that are called *Scale Free* and the degree distribution of such graphs follow the *power law*. Several generalizations of this model has also been proposed. In this paper we present a new generalization of the model and attempt to bring out its implications in real life.

I. INTRODUCTION

Our lives are surrounded with several complex phenomena which are at times very difficult to explain or understand. These phenomena usually encompass a lot of spheres of our lives. One such phenomena is how certain biological, geological, physical, astronomical, financial and social systems show a very peculiar similarity that is they all exhibit a so called *Scale Free Property* [4]. This property says that certain features of these systems follow a common pattern. For example in case of geology, most of the earthquakes that occur on this planet are nominal whereas few are gigantic, in case of financial systems most people on earth have small incomes and a few have a monumental incomes, in case of social systems, there are few celebrities who are extremely popular whereas most of the other are not and this list of examples continues on and on. As it turns out this kind of a Scale free pattern has been known to scientists and researchers for long. Herbert Simon [3] showed the existence of such properties way back in the 1950s. Although this property is very common, present literature exhibits that there is not much understanding of as to why and how it occurs. In recent years with the advent of new technologies like the internet, WWW, social media etc., a new stream of research has come up which seeks to explain this phenomenon in several complex systems by employing mathematical models. Here the key is to find the existence of a graph theoretic structure that is present in these systems and show that the Scale Free Property is present in that graph. Researchers have shown that the presence of this property means that the corresponding graph follows the *power law* degree distribution. The most well known among such models is the model of Barabási-Albert that describes a model based on *preferential attachment*. This model ensures that as this graph evolves by the addition of new nodes, each new node

gets attached to a set of m existing nodes where the attaching probability of the new node is proportional to the degree of that node already present in the graph. This model although considered to be fairly satisfactory, is not very realistic at times and has several drawbacks. Due to this several generalizations of this model has been proposed that try to improvise on the properties of the graph generated. In this paper we propose a new model that is a generalization of the AB model and is based on the *cuts* in the graph. Our model is very novel and has no apparent links with the already existing generalization of the Barabási-Albert (BA) model. In this paper, we introduce our model, which we have named as the *EvoCut* model, and describe its properties. We further bring out how this model is more realistic than the already existing models.

A. The scale free property

As mentioned in the previous section, the scale free property [4] basically means that the occurrence of very high is small and there is an abundance of small. In the context of networks and graphs this means that there are very few nodes with high degree and there are a lot of nodes with small degree. A more technical way of putting it is to say that the degree distribution follows the *power law* as shown in Fig 1 i.e. $p(k) \propto k^{-\gamma}$ where $\gamma \in (2, 3]$; $p(k)$ is the number of nodes with degree k divided by $2m$ (which is the total degree of nodes). This alternatively means that on log scale the plot of $p(k)$ vs. k is linear.

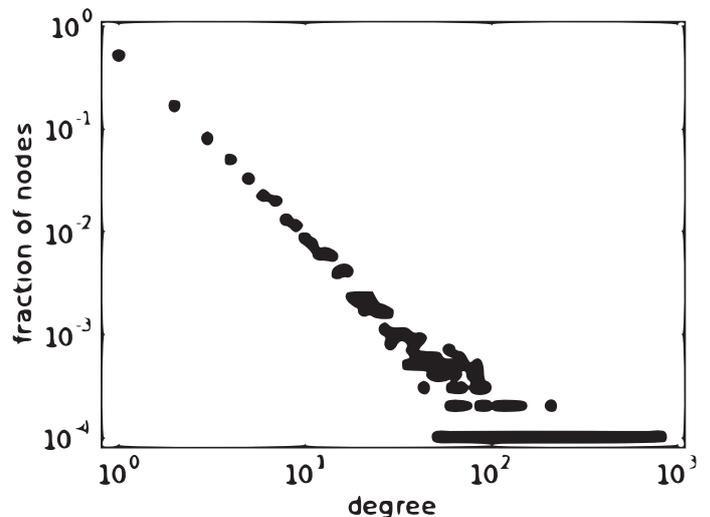


Fig. 1. Plot of power-law degree distribution on log-log scale

II. PRIOR WORK ON GENERALIZATION OF ALBERT-BARABÁSI MODEL

In this section we present some prior models that are based on the generalization of the Barabási-Albert model.

A. Dorogovtsev-Mendes-Samukhin (DMS) Model

Dorogovtsev, Mendes and Samukhin present a model [5] that generalizes the BA model in the following sense, they incorporate the initial state of the nodes in the network and the degree (k) is dependent on the initial state (s) and time (t). They come up with a dynamics that achieves power law when the degree is very large i.e. $p(k) \propto k^{-\gamma}$ when $k \rightarrow \infty$ and $\gamma \in (2, \infty)$. The average connectivity $k(s, t) \propto (s/t)^{-\beta}$ as $(s/t) \rightarrow 0$. This generalization makes the model very unrealistic because $\gamma \in (2, \infty)$ whereas in reality $\gamma \in (2, 3]$, as supported by BA model.

B. Antal-Krapivsky-Redner (AKR) Model

The AKR model [6] generalizes the BA model by considering a graph in which the links are associated with a notion of friendship (+1) or enmity (-1) and the notion of a node is replaced by a collection of three nodes linked with either (+1) or (-1) links, which is called a triad. The model defines a notion of a balance which is the product of all the values of the links in a particular triad. If this product is equal to 1 then it is called *balanced* otherwise it is called *imbalanced*. In this model, the authors define a dynamics that tries to maintain the balance of the triads and describes how the degree of a link (which is the number of triads in which it is involved) changes with time as the network is allowed to evolve according to the dynamics. They derive a set of differential equations that governs this rate and solve them to get a distribution same as the power law.

C. Sole-Pastor-Satorras-Smith-Kepler (SPSK) Model

The Solé, Pastor-Satorras, Smith, Kepler (SPSK) model [7] uses 3 mechanisms *duplicate*, *divergence* and *mutate* for the process of evolution shown in Fig 2. Using the operation *Duplicate* one can copy a randomly selected node along with its connections, using *Divergence* operation one can delete some connections made after the duplicate operation and the *Mutate* operation allows us to add connections once the duplicate operation is applied. It has been shown by them that this generalization produces the *power law* but leads to unrealistic assortativity and clustering. This is a generalization of the BA model because it allows more possible ways of connection of a new node with the existing graph.

It is customary to note that there have been some generalizations of preferential attachment model that are based on the addition and deletion of nodes [8], [9].

III. OUR CONTRIBUTION

In this paper we present a new model that generalizes the BA model. As compared to the previous models our model is a natural generalization of the BA model and is more intuitive than the previous generalizations. It uses the combinatorial properties of the given graph and does not create unrealistic links as in the case of SPSK model and DMS model. Our

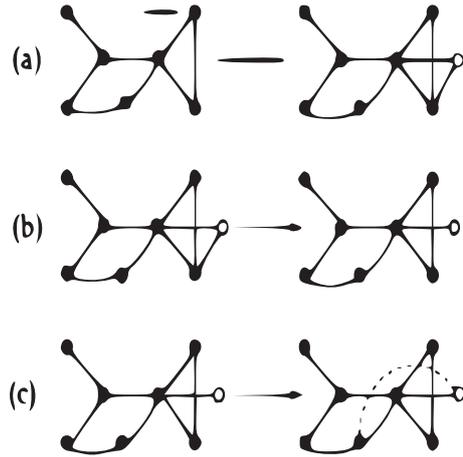


Fig. 2. (a) Duplicate, (b) Divergence and (c) Mutate

model has two variants, one of which generates scale free networks (as in BA model) and the other one gives rise to a family of graphs in which the degree distribution follows the stretched exponential distribution rather than the power law, which is similar to the SPSK model. In the following sections we describe in detail our proposed model and also discuss the properties of the two variants of the model.

IV. PRELIMINARIES

We model the network as an undirected graph in which nodes represent the network members and an undirected edge represents a relationship between them. Initially the graph is considered to have n_0 number of nodes with a few links m_0 . Our evolution process crucially uses the notion of cuts which is an important combinatorial property of a graph. Given a graph $G = (V, E)$ a subset of vertices S and $\bar{S} = V \setminus S$, a *Cut* is defined as the set $E(S, \bar{S}) = \{(a, b) | a \in S \text{ and } b \in \bar{S}\}$. Our evolution process uses the size of the cut crucially making it substantially different from all the prior models that do not use the notion of cut in any way. We also define a k -neighborhood of a particular vertex v as the number of nodes of G which are within a distance k from the node v . Thus $B(v, k) = \{u | d(v, u) \leq k\}$. We also define the set $B'(v, k)$ which we call the *boundary* of a node v as the set of all nodes which satisfy the following

$$\{u \in B(v, k) | (u, v) \in E(B(v, k), \overline{B(v, k)}); v \in \overline{B(v, k)}\}$$

We also define the *pulling power* of a node v as probability that a new node gets attached to an already existing node v . In this paper we would use this term irrespective of the model that we are looking into.

V. THE EVOCUT MODEL

In this section we introduce our model and describe in detail its two variants. Both these variants use preferential attachment of a different kind to select a node from the graph, but have different strategies for attachment of the new incoming node.

Algorithm 1: Model A

```

1  $G_0 = (V_0, E_0)$ ;  $t = 0$ ;  $|V_0| = n_0$  and  $E_0 = m_0$ ,
    $k = k_0$ ,  $Y = 0$ ;
2 while ( $n_t \leq N$ ) do
3   Let  $v_t$  be the new node at time  $t$ ;
4   for  $v \in V_t$  do
5     Compute  $x = |B(v, k)|$ ;
6      $Y = Y + x$ ;
7   end
8   Pick a  $u$  randomly in which the probability of
     selecting  $u$  is  $\{\frac{|B(u, k)|}{Y}\}$ ;
9    $E_t = E_t \cup (v_t, u)$ ;
10   $t = t + 1$ ;
11   $Y = 0$ ;
12 end

```

Algorithm 2: Model B

```

1  $G_0 = (V_0, E_0)$ ;  $t = 0$ ;  $|V_0| = n_0$  and  $E_0 = m_0$ ,
    $k = k_0$ ,  $Y = 0$ ;
2 while ( $n_t \leq N$ ) do
3   Let  $v_t$  be the new node at time  $t$ ;
4   for  $v \in V_t$  do
5     Compute  $x = |B(v, k)|$ ;
6      $Y = Y + x$ ;
7   end
8   Pick a  $u$  randomly in which the probability of
     selecting  $u$  is  $\{\frac{|B(u, k)|}{Y}\}$ ;
9   Select a node  $v'$  randomly uniformly from the set
      $\{z \in B(u, k) | (z, u') \in E(B(u, k), B(u, k)); u' \in$ 
      $B(u, k)\}$ ;
10   $E_t = E_t \cup (v_t, v')$ ;
11   $t = t + 1$ ;
12   $Y = 0$ ;
13 end

```

VI. PROPERTIES OF THE EVOCUT MODEL

In this section we describe in detail the properties and the reasoning behind coming up with these models. In this model we allow the nodes to be attached one by one and the pulling power of a particular node is defined by the size of the cut $E(B(v, k), \overline{B(v, k)})$ which the number of edges in the k -neighborhood of the node v where k is a parameter between $[0 - (n - 1)]$. The pulling power is same for both the models A and B. The other features of the models is being mentioned in the following subsections.

A. Model A - Deterministic Case

In this model a particular node is attached to a node which selected randomly from a probability distribution that is different from the one used in the BA- model. The process starts with an initial graph $G_0 = (V_0, E_0)$ and goes until the number of nodes in the graph is less than N . Every time when a new node arrives, the **for** loop computes the normalizing factor

$$\sum_v |E(B(v, k), \overline{B(v, k)})| \quad (1)$$

and the later part of the model selects a node u randomly according to the probability distribution in which the probability of selecting a node u is the ratio

$$\frac{B(u, k)}{\sum_w |E(B(w, k), \overline{B(w, k)})|} \quad (2)$$

and attaches the new node with this chosen node u . Thus this model is thus *deterministic* in the sense that there is no randomness while attaching the new node with the existing graph, once a node has been selected from the aforementioned probability distribution.

B. Model B - Randomized Case

In this model a particular node is attached to a randomly chosen node on the boundary of the set $B(v, k)$. The boundary is defined as the nodes in $B(v, k)$ which are incident to the edges in the set $E(B(v, k), \overline{B(v, k)})$. This model not only uses randomness according to preferential attachment but also uses randomness in choosing the node to which the incoming node should be added. This step makes it different from the choice made in Model A. As in the previous case the model first selects a node u randomly according to the probability distribution in which the probability of selecting a node u is the ratio

$$\frac{B(u, k)}{\sum_w |E(B(w, k), \overline{B(w, k)})|} \quad (3)$$

and then randomly chooses a node from the set $B'(u, k)$ and attach the new node with that node.

C. As a Generalization of BA Model

One can observe that in both the models mentioned above if we fix the parameter k as 0, then we get the (BA) model. This is simply because when $k = 0$, $B(v, k) = v$ and the set $E(B(v, k), \overline{B(v, k)})$ is equal to the degree of v and in that case the attaching probability distribution is exactly equal to the degree distribution of the graph at time t which is same as the BA model. Thus, in our model, only the special case of $k = 0$ results in the BA model.

VII. EXPERIMENTAL RESULTS AND ANALYSIS

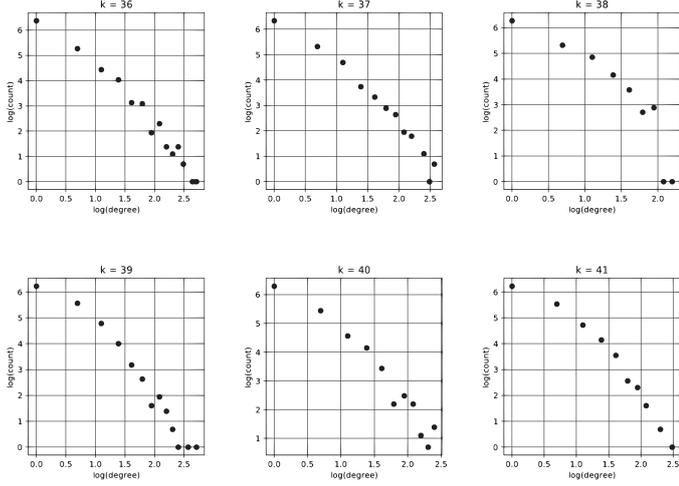
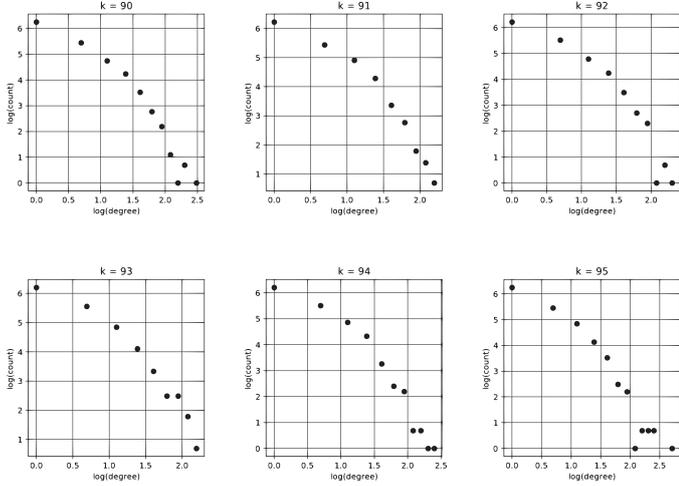
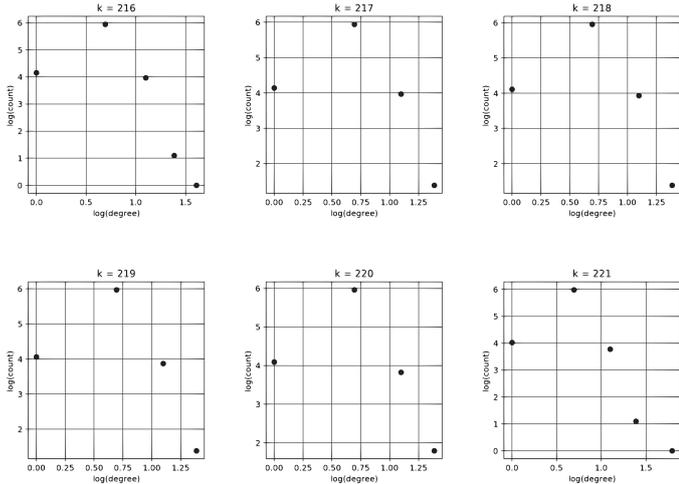
In this section we present the degree distributions that we obtain as we let our model to evolve on a set of nodes with certain initial condition. We let this process run till $n = 500$.

A. Analysis of Model A

In the following, Fig 3 and Fig 4 describe the plots that we obtain once the degree distributions are generated for *Model A*.

According to the plots that are generated we can infer the following:

- For small values of k this model gives rise to a scale free distribution when k is even and for larger values of k the degree distribution follows the *stretched exponential distribution* in the log-scale. This in turn implies that in the normal scale the plot ensures that the number of nodes with high degree is fairly large as


 Fig. 3. Degree-distributions for Model A for small k (36-41).

 Fig. 4. Degree-distributions for Model A for large k (90-95).

 Fig. 5. Degree-distributions for Model A for very large $k \approx n/2$ (216-221).

compared to the *scale free distribution* which implies that for large values of k there are several nodes with high degree.

- The intuitive reasoning for the aforementioned result is that for small values of k this model behaves basically similar to the BA model. Whereas for larger values of k we can observe that once a new node gets attached to the already existing node the pulling power of the already existing node does not increase and the pulling power of the k^{th} -neighbor increases. This phenomenon creates some sort of an oscillation on the increase in the degree of the certain number of nodes which implies that this set of nodes experience enhancement of degree simultaneously. This explains the *stretched exponential distribution* in the log-scale for larger values of k .
- We also observe in Figure 5 that for very large values of $k \approx n/2$ we obtain a degree distribution that is significantly different from the power law distribution. This distribution is also fairly different from the distribution that we get for smaller values of k .

B. Analysis of Model B

In the following, Fig 6 and Fig 7 describe the plots that we obtain once the degree distributions of *Model B* are generated and based on these we infer the following

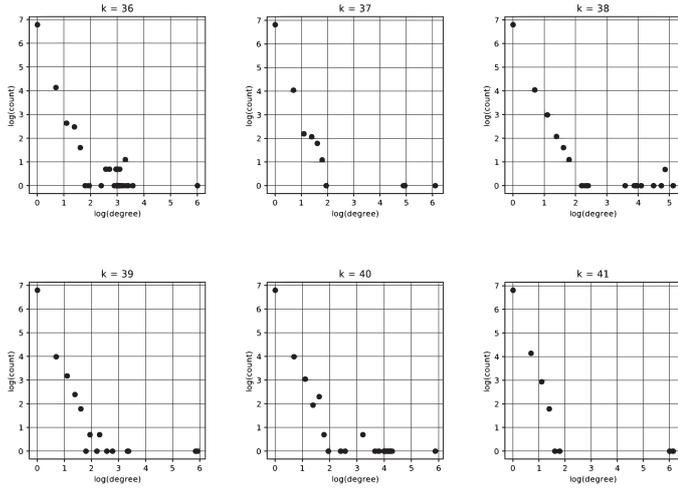
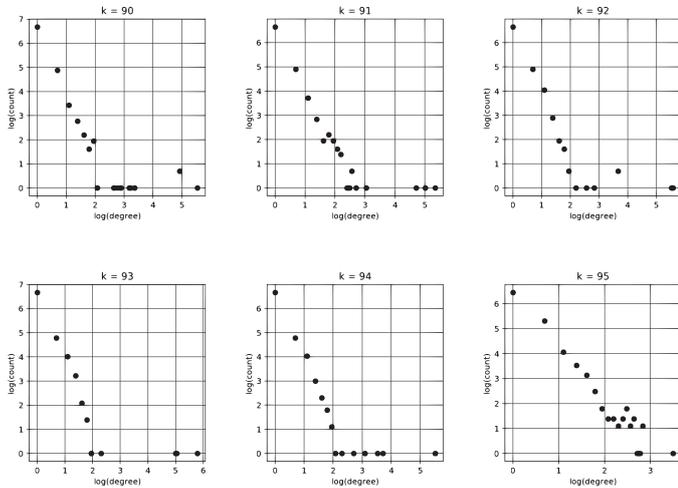
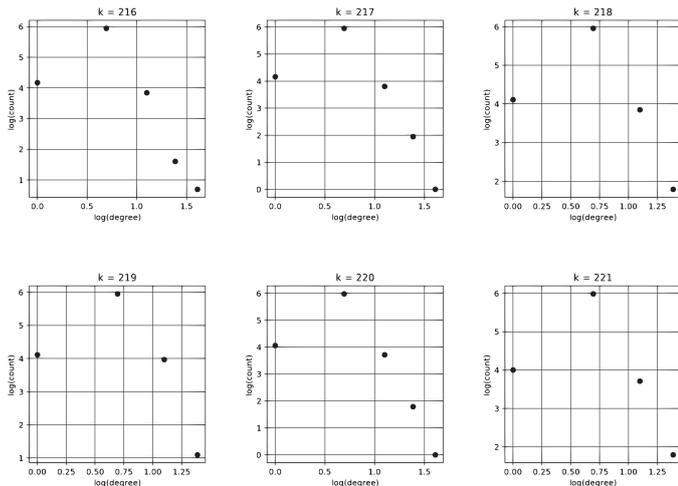
- As said earlier in this model the incoming node is attached to that node which is on the boundary of the k -neighborhood of the node that maximizes the objective function.
- From the plots, it is clear that for both small and large values of k we get a *power law* distribution. This observation can be explained intuitively as follows: since the incoming node x is getting attached to a node which is on the boundary of the k -neighborhood of a node y , the pulling power of node y increases with time because it is defined as the number of edges on the boundary of the k -neighborhood of y and hence in further iterations the node y gets enhancement in its power.
- This makes the model a generalization of the BA model which is quite similar to it.
- We also observe in Figure 8 that similar to the case of Model A, for very large values of k we obtain a degree distribution that is very different from the power law distribution. This behavior is sort of expected in both models because the process of finding the power of a particular node starts after a large number of nodes had already been added. This leads the degree distribution to concentrate on a few nodes only.

VIII. COMPARISON WITH PRIOR MODELS

In this section we compare our models with the prior generalizations of the BA model and also discuss how close to reality these models can be considered.

A. Comparison with DMS model

As mentioned before, the DMS model starts with an initial state and the degree of the node is dependent on the initial state of the evolution process, a feature which our models also


 Fig. 6. Degree-distributions for Model B for small k (36-41)

 Fig. 7. Degree-distributions for Model B for large k (90-95)

 Fig. 8. Degree-distributions for Model B for very large $k \approx n/2$ (216-221).

possess. The DMS model achieves the power law for large values of the degree whereas in our case the Model B achieves power law even for small values of the degree.

B. Comparison with AKR model

In the AKR model the authors come up with a notion of a balance of the values present in a triad which is a collection of three nodes and the directed links among them. The dynamics of the model tries to maintain the balance of the triads. This model, although a generalization of the BA model uses only *local* modifications whereas our models are *global* in the sense that each node looks around a k -neighborhood where k can be fairly large. This allows our model to look at the global influence of the nodes which is not present in the AKR model.

C. Comparison with SPSK model

In the SPSK Model, the evolution of the graph is based on some operations done on the original graph called *duplicate*, *divergence* and *mutate*. This make the model very restrictive because a node can only get attached to the existing graph by performing a duplicate operation which is basically the replication of the connection of an already existing node. Our models are not restrictive in that sense and a new node is not forced to follow the topology of an already existing node and is attached purely based on the pulling power of a particular node. Despite the difference in nature of SPSK and our models we observe that the degree distribution of this model and Model A proposed by us is quite similar and follows the stretched exponential distribution.

IX. REAL LIFE IMPLICATIONS OF EVOCUT

One of the important questions regarding the models of evolution of complex networks is how realistic are they. The BA model although satisfactory is not considered to be very realistic. Thus there is definitely a need to come up with models which not only give rise to the power law distribution but are also realistic. In this section we justify how our models can explain the realities of certain complex networks. In the case of Model A, as we had observed that this model leads to a degree distribution [10] in which there are lots of nodes with large degrees. Although this doesn't follow the scale free property but explains the nature of some real political networks. It is well known that [11] the stretched exponential distribution describe very well the distributions of radio and light emissions from galaxies, of country population sizes, of daily Forex US-Mark and Franc-Mark price variations, of Vostok temperature variations and of citations of the most cited physicists in the world. This type of distribution may not be considered to be explainable by the BA model and its generalizations which tend to result in a power law distribution, but can be explained by Model A.

In the case of Model B we notice that since the model looks at a k -neighborhood of a particular node to compute its pulling power this allows us to look at a larger influence of a node as compared to the standard BA model. In fact in realistic scenarios one can conceive of several situations in which when a new node comes it doesn't get attached to the node with maximum pulling power but gets attached to a node which is at a certain distance from the node. Consider the example of celebrities in real life networks, when a person considers himself as a follower of a particular celebrity then he gets himself linked to another follower of the same celebrity rather than directly getting linked with the celebrity, and this follower might as well not be one of the closest to

the celebrity i.e. one gets attached to a *fan club* rather than the celebrity himself.

We can also consider the example of any hierarchical ecosystem [12] present in an organization like judiciary, police, banking system or a corporation in which any customer is not allowed to connect with the highest authority in that system but gets connected to a node that is lowest in that hierarchy. This notion is being captured in the case of Model B, in the sense that the k -neighborhood of a node basically creates a hierarchy around that node considering that node as most powerful. As the distance from the node increases we get down in the hierarchy, and hence when a new node comes under the influence of the k -neighborhood of a particular node at the k^{th} level of the hierarchy.

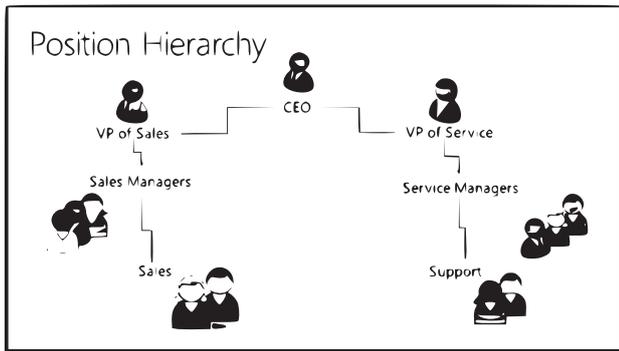


Fig. 9. Hierarchical influence in Corporate Structure

Consider the aforementioned Fig 9 in which the CEO represents the node whose pulling power is being computed and the 1-neighborhood of it consists of VP of Sales and VP of Service, the 2-neighborhood consists of Sales Managers and Service Managers and the 3-neighborhood consists of Sales and Support. Thus when a new customer comes it is likely to get in touch with the 3-neighborhood rather than the CEO. Thus in several scenarios Model B gives a better picture of the evolution of the network and hence is much more realistic than the BA model.

X. CONCLUSION AND FUTURE WORK

In this paper, we have presented a new generalization of the Barabási-Albert (BA) model that is based on the neighborhood properties of nodes in the evolving graph. This generalization is substantially different from the previous generalizations of the BA model and one of its variant gives rise to the power law distribution and can model the growth of several real life networks which don't seem to be explainable by the BA model. The other variant of our Model, despite the fact doesn't give the scale free property models the growth of some other real life networks and gives rise to a stretched exponential

distribution. We believe that our model is very powerful and further investigation into it can make it a stronger candidate in understanding the growth of several complex networks.

As part of future work it would interesting to get a mathematical proof of the distribution generated by our models and also discover new interesting variants of our model. It is well known that a master equation can be used to derive a proof of the distribution generated by any preferential attachment based model. Thus we could use this equation to mathematically prove the distribution obtained by the two variants of EvoCut. The master equation is usually formulated as

$$\frac{dN_k}{dN} = \frac{A_{k-1}N_{k-1} - A_kN_k}{A} + \delta_{k,1} \quad (4)$$

where N_k represents the number of nodes with degree k and N is time parameter and $A = \sum_k A_k N_k$ which can also be thought as the total rate. The value $\delta_{k,1}$ represents the insertion of degree 1 nodes. A_k denotes the probability of getting attached to degree k node. In order to apply this master equation in case of our models we need to obtain a function f such that $N_k = f(N_{k-1})$ which can then be used to obtain a solution to the differential equation.

REFERENCES

- [1] A. L. Barabási, Network science. Cambridge university press, 2016.
- [2] A. L. Barabási, Scale-free networks: a decade and beyond. science, 325 (5939), pp.412-413, 2009.
- [3] H. A. Simon, A behavioral model of rational choice. The quarterly journal of economics, 69(1), 99-118, 1955.
- [4] A. L. Barabási & E. Bonabeau, Scale-free networks. Scientific american, 288(5), 60-69, 2003.
- [5] S. N. Dorogovtsev, J. F. F. Mendes & A. N. Samukhin, Structure of Growing Networks: Exact Solution of the Barabási-Albert's Model, 2000.
- [6] T. Antal, P. L. Krapivsky & S. Redner, Dynamics of social balance on networks. Physical Review E, 72(3), 036121, 2005.
- [7] R. V. Sol, R. Pastor-Satorras, E. Smith & T. B. Kepler, A model of large-scale proteome evolution. Advances in Complex Systems, 5(01), 43-54, 2002.
- [8] H. Bauke, C. Moore, J. B. Rouquier & D. Sherrington, Topological phase transition in a network model with preferential attachment and node removal. The European Physical Journal B, 83(4), 519-524, 2011.
- [9] C. Moore, G. Ghoshal & M. E. Newman, Exact solutions for models of evolving networks with addition and deletion of nodes. Physical Review E, 74(3), 036121, 2006.
- [10] L. A. Adamic, R. M. Lukose, A. R. Puniyani & B. A. Huberman, Search in power-law networks. Physical review E, 64(4), 046135, 2001.
- [11] J. Laherrere & D. Sornette, Stretched exponential distributions in nature and economy: fat tails with characteristic scales. The European Physical Journal B-Condensed Matter and Complex Systems, 2(4), 525-539, 1998.
- [12] E. Ravasz, A. L. Somera, D. A. Mongru, Z. N. Oltvai & A. L. Barabási, Hierarchical organization of modularity in metabolic networks. science, 297(5586), 1551-1555, 2002.