# A Neighbor of Initial Contact system for Academic Collaboration

Sreedhar Bhukya
Department of Computer and Information Sciences
University of Hyderabad
Hyderabad- 500046
India
{sr2naik@gmail.com}

***ABSTRACT*: *It has been observed in the literature that social networks have characteristics such as assortative mixing, high clustering, short average path lengths, broad degree distributions and the existence of community structure. In our study we designed a model in academic collaboration' which satisfies all the above characteristics, based on some existing social network models. In addition, this model facilitates interaction between various communities (academic/ research groups). This model gives very high clustering coefficient by retaining the asymptotically scale-free degree distribution. Here the community structure is raised from a mixture of random attachment and implicit preferential attachment. In addition to earlier works which only considered Neighbor of Initial Contact (NIC) as implicit preferential contact, we have considered Neighbor of Neighbor of Initial Contact (NNIC) also. This model supports the occurrence of a contact between two Initial contacts if the new vertex chooses more than one initial contacts. This ultimately will develop a complex social network rather than the one that was taken as basic reference.*

**Keywords**: Social networks, Novel model,Random initial contact,Neighbor of neighbor initial contact, Tertiary contact, Academic collaboration

**Received:** 23 March 2011, Revised 23 April 2011, Accepted 30 April 2011

© 2011 DLINE. All rights reserved

## 1. Introduction

Now a days research in collaborations becoming domain independent. For example stock market analyst is taking the help of physics simulator for future predictions. Thus there is a necessity of collaboration between people in different domains (different communities, in the language of social networking.) Here we develop a novel model for collaborations in academic communities which gives a possibility of interacting with a person in a different community, yet retaining the community structure. Social networks are made of nodes that are tied by one or more specific types of relationships. The vertex represents individuals or organizations. Social networks have been intensively studied by Social scientists [3-5], for several decades in order to understand local phenomena such as local formation and their dynamics, as well as network wide process, like transmission of information, spreading disease, spreading rumor, sharing ideas etc. Various types of social networks, such as those related to professional collaboration [6-8], Internet dating [9], and opinion formation among people have been studied. Social networks involve Financial, Cultural, Educational, Families, Relations and so on. Social networks create relationship between vertices; Social networks include Sociology, basic Mathematics and graph theory. The basic mathematics structure for a social network is a graph. The main social network properties includes hierarchical community structure [10], small world property [11], power law distribution of nodes degree [19] and the most basic is Barabasi Albert model of scale free networks [12]. The more online social network gains popularity, the more scientific community is attracted by the research opportunities that these new fields give. Most popular online social networks is Facebook, where user can add friends, send them messages, and update their personal profiles to notify friends about themselves. Essential characteristics for social networks are believed to include assortative mixing [13, 14], high clustering, short average path lengths, broad degree distributions [15, 16]The existence of community structure, growing community can be roughly speaking set of vertices with dense internal connection, such that the inter community connection are relatively sparse [2].

Sousa et al. [20] developed a project for social networking system for educational professionals, this paper consider what kind of technologies could be used to create a web application that provided, type of interaction, behavior needed, requirements, technologies and the system implementation .

A report [21] also has been submitted which is examining three specific types of collaborative behavior and assessing their impacts on knowledge creation, Drawing on the toolkits of Social and Dynamic Network Analysis and a dataset of computer science department tenure and research track of faculty members of a major U.S. university.

The advantage of our model can be understood from the fallowing example. Let us consider a person contacting a person in a research group for his own purpose and suppose that he/she didn’t get adequate support from that person or from his neighbors, but he may get required support from some friend of friend for his/her initial contact. Then the only way a new person could get help is that his primary contact has to be updated or create a contact with his friend of friend for supporting his new contact and introduce his new contact to his friend of friend. The same thing will happen in our day to day life also. If a person contacts us for some purpose and we are unable to help him, we will try to help him by some contacts of our friends. The extreme case of this nature is that we may try to contact our friend of friend for this purpose. We have implemented the same thing in our new model. In the old model [2], information about friends only used to be updated, where as in our model information about friend of friend also has been updated. Of course this model creates a complex social network model but, sharing of information or data will be very fast. This fulfills the actual purpose of social networking in an efficient way with a faster growth rate by keeping the community structure as it is.

## 2. Network growth algorithm

The algorithm includes three processes: (1) Random attachment (2) Implicit preferential contact with the neighbors of initial contact (3) In addition to the above we are proposing a contact between the initial contact to its Neighbor of Neighbor contact (tertiary ). The algorithm of the model is as follows [1]:

1) Start with a seed network of N vertices

2) Pick on average $m_r \geq 1$ random vertex as initial contacts

3) Pick on average $m_s \geq 0$ neighbors of each initial contact as secondary contact

4) Pick on average $m_t \geq 1$ neighbors of each secondary contact as tertiary contact

5) Connect the new vertex to the initial, secondary and tertiary contacts

6) Repeat the steps 2-5 until the network has grown to desired size.

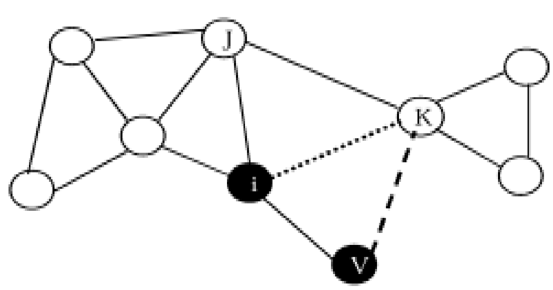


Figure.1. Growing process of community network, the new vertex ‘V’ initially connects through some one as initial contact (say i). Now i, updates its neighbor of neighbor contact list and hence connects to k. V’ connects to ms number of neighbors (say k) and mt number of neighbor of neighbors of i (say k). In this model we tried 50 sample vertices and prepared a growing network.

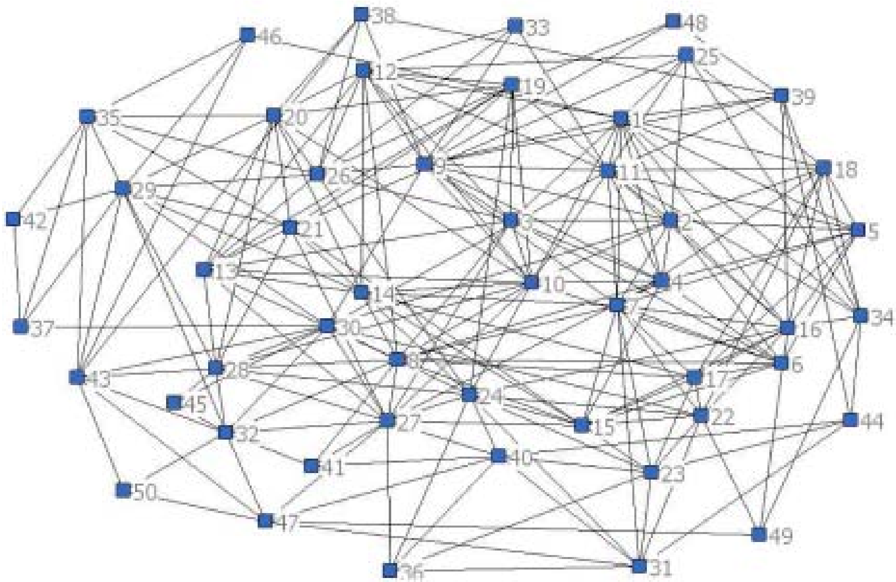


Figure 2. Showing Social Network Graph with 50 Vertices

### 3. Vertex degree distribution

We derive approximate value for the vertex degree distribution for growing network model mixing random initial contactneighbor of neighbor initial contact and neighbor of initial contacts. Power law degree distribution with p (k) ~ $k^i$ with exponent 2< i < i have derived [17, 19]. In this model also the lower bound to the degree exponent is found to be 3, which is same as in the earlier model.

The rate equation which describes how the degree of a vertex changes on average during one time step of the network growth is constructed. The degree of vertex $v_i$ grows in 3 processes:

1) When a new vertex directly links to $v_i$ at any time t, there will be on average ~ t vertices. Here we are selecting mr out of them with a probability $m_r$ /t.

2) When a vertex links to $v_i$ as secondary contact, the selection will give rise to preferential attachment. These will be ms,mts in number.

3) When a vertex links to $v_i$ as tertiary contact, this will also be a random preferential attachment. These will be $m_r$,$m_s$,$m_t s_t$ in number.

These three processes lead to following rate equation for the degree of vertex $v_i$ [1].

$$\frac{\partial k_i}{\partial t} = \frac{1}{t}\left( m_r + \frac{m_r m_s + 2\, m_r m_s m_t}{2(m_r + m_r m_s + 2\, m_r m_s m_t)} k_i \right) \qquad (1)$$

Based on the average initial degree of a vertex is

$$K_{init} = m_r + m_r m_s + 2\, m_r m_s m_t$$

Separating and integrating from $t_i$ to t, and from $k_{init}$ to $k_i$ , we will get the following time evaluation for the vertex degrees.

$$k_i(t) = B\left(\frac{1}{t_i}\right)^{1/A} - C \qquad (2)$$

Where

$$A = 2\left( \frac{m_r + m_r m_s + 2\, m_r m_s m_t}{m_r m_s + 2\, m_r m_s m_t} \right)$$

(4)

$$B = A(m_r + \frac{1}{2} m_r m_s - 3 m_r m_s m_t)$$
$$C = A m_r$$

From time evolution of vertex $k_i$ (t), we can calculate the degrees of distribution p(k) by forming cumulative distribution F(k) and differentiating with respect to k. Since the mean field approximation[1] the degree $k_i(t)$ of a vertex vi increases monotonously from the time $t_i$ the vertex initially added to the network, the fraction of vertices whose degree is less than $k_i(t)$ at t is equivalent to the fraction of vertices that introduced after time ti. Since t is evenly distributed, this fraction is $(t-t_i)/2$. These facts lead to the cumulative distribution [1].

$$F(k_i) = P(\tilde{k} \le t) = P(\tilde{t} \ge t_i) = \frac{1}{t}(t - t_i) \quad (3)$$

Solving for $t_i = t_i(k_i, t) = B^A(k_i, C)^{-A} t$ from (2) and inserting it into (3), differentiating $F(k_i)$ with respect to $k_i$, and replacing the notation $k_i$ by k in the equation, we get the probability density distribution for the degree k as

$$P(k) = AB^A (k + C)^{-2} / ms + 2 m_s m_t^{-3}$$

Here A, B and C are as above. In the limit of large k, the distribution becomes a power law p (k) ~ $k_i$ with 3 =3+2/$m_s$, $m_s > 0$, leading to 3< i < ∞ . Hence the lower bound to the degree exponent is 3. Although the lower bound for degree exponent is same as earlier model. The probability density distribution is larger compared to earlier model, where the denominator of the first term of degree exponent is larger compared to the earlier model.

**4. Clustering**

The clustering coefficient on vertex degree can also be found by the rate equation method [18]. Let us examine how the number of triangles $E_i$ changes with time. The triangle around vi are mainly generated by three processes.

1. Vertex $v_i$ is chosen as one of the initial contact with probability $m_r/t$ and new vertex links to some of its neighbors as secondary contact, giving raise to a triangle.

2. The vertex $v_i$ is chosen as secondary contact and the new vertex links to it as its primary or tertiary contact giving raise to a triangles.

3. The vertex $v_i$ is chosen as tertiary contact and the new vertex links to it as its primary or secondary contact, giving raise to a triangles.

These three process are described by the rate equation [1]

$$\frac{\partial E_i}{\partial t} = \frac{k_i}{t} - \frac{1}{t}(m_r - m_s m_t - 3 m_r m_s m_t) - \frac{5 m_r m_s m_t}{2(m_r + m_s m_t + 3 m_r m_s m_t)} \quad (5)$$

where second right-hand side obtained by applying Equation. (1) integrating both sides with respect to t, and using initial condition $E_i(k_{init}, t_i) = m_r m_s(1+3m_t)$, we get the time evaluation of triangle around a vertex $v_i$ as

$$E_i(t) = (a + b_i) \ln\left(\frac{t}{t_i}\right) + \left(\frac{a + bk_i}{b}\right) \ln\left(\frac{a + bk_i}{a + bk_{init}}\right) + E_{init} \quad (6)$$

Now making use of the previously found dependent of ki on $t_i$ for finding $c_i(k)$. solving for $\ln(t/t_i)$in terms of $k_i$ from (2), inserting into it into (6) to get $E_i(k_i)$, and dividing $E_i(k_i)$ by the maximum possible number of triangles, $k_i(k_i-1)/2$, we arrive the clustering the coefficient.

$$c_i(k_i) = \frac{2E_i(k_i)}{k_i(k_i - 1)} \quad (7)$$

Where

$$E_i(k_i) = Abk_i \ln(k_i + C) + k_i \ln\left(\frac{a + bk_i}{a + bk_{init}}\right) - bAk_i \ln B + aA \ln(k_i + C) + \frac{a}{b} \ln\left(\frac{a + bk_i}{a + bk_{init}}\right) + E_{init} - aA \ln B$$

$$Dk_i \ln(k_i + C) + (F + Gk_i) - k_i \ln B + HDk_i \ln(k_i + C) I \ln(F + Gk_i) + J$$

where

For large values of degree k, the clustering coefficient thus depend on k as c(k)~ln k/k. This has very large clustering coefficient compared to the earlier work where it was c(k)~1/k.

## 5. Results

In this model we tried 50 sample vertices and prepared a growing network, where edge to vertex ratio and triangle to vertex ratio for 50 nodes has been prepared. The results are given in Table: 1 .here one can see an enormous increase in secondary contacts. In addition tertiary contacts also have been added in our model, which leads to a faster and complex growth of network.

| Data on our proposed model | Initial Contact (IC) | Secondary Contacts (SC) | Neighbor of Neighbor IC (NNIC) |
|---|---|---|---|
| Vertices | 2.8 | 5.56 | 2.78 |
| Triangles | 0.8 | 6.0 | 6.44 |

Table 1. Result of sample Vertices

### 5.1 Simulation results

The below results have been represented graphically by calculating the degree (number of contacts) of a node. This also is shows an enormous growth in degree of nodes.

Simulation results

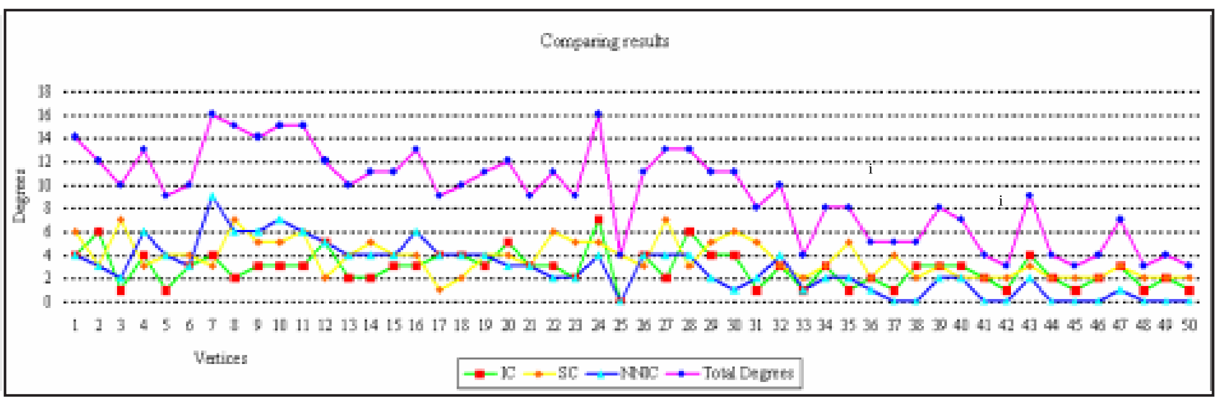


Figure 3. Comparison results of growing network community: initial contacts are growing very slow rate compared to secondary contact i.e. indicates initial contact, indicates secondary contacts, and indicates neighbor of neighbor of initial contact connects to the vertex v , Finally indicates degree of each vertices, when initial, secondary and tertiary contact connect to a vertex v. Our network community is growing very fast and complex when compared to existing model, vertices simulation results based on Table: 1

## 6. Conclusion

In this paper, a model which reproduces very efficient networks compared to real social networks has been developed. And also here, the lower bound to the degree exponent is the same. The probability distribution for the degree k is in agreement with the earlier result for m =0. The clustering coefficient got an enormous raise in growth rate of ln(k )/ k compared to theearlier result 1/k for large values of the degree k. This is very useful in the case of academic groups, which helps in faster information flow and an enormous growth in research. Thus here an efficient but complex model of social network has been developed which gives an enormous growth in probability distribution and clustering coefficient and edge to vertex ratio by retaining the community structure. This model can be used to develop a new kind of social networking among various research groups.

**Tool**

We have used C language, UciNet, NetDraw and Excel for creating graph and simulation.

**Notations**

| Notation | Description |
|---|---|
| $m_r$ | Initial Contact |
| $m_s$ | Secondary Contact |
| $k_i$ | Degree of vertex i |
| $E_i$ | Number of triangles at vertex i |
| P(k) | Probability density distribution of degree k |

**References**


[1] Sreedhar Bhukya (2011). A novel model for social networks, BCFIC,IEEE, February 16-18, p.no 21-24

[2] Riitta Toivonen, Jukka-Pekka Onnela, Jari Sarama¨ki, Jo¨rkki Hyvo¨nen, Kimmo Kaski (2006). A model for social networks. Physica A 371 851–860.

[3] Milgram, S., *Psychology Today* 2 (1967) 60–67.

[4] Granovetter, M. (1973). The Strength of Weak Ties, *Am. J. Soc*. 78 1360–1380.

[5] Wasserman, S., Faust, K. (1994). Social Network Analysis, Cambridge University Press, Cambridge.

[6] Duncan, J., Watts Strogatz Steven, H. (1998). Collective dynamics of ‘small -world networks, NATURE 393 -440.

[7] Newman, M. (2001). The structure of scientific collaboration networks, *PNAS* January 17, 98. 404- 409.

[8] Newman, M. (2001). Coauthorship networks and patterns of scientific collaboration, PNAS, April6. Vol. 101. 5200–5205.

[9] Holme, P., Edling, C, R., Fredrik. Liljeros, Structure and Time-Evolution of an Internet Dating Community, *Soc. Networks* (2004). 26- 155–174.

[10] Girvan, M., Newman, M. E. J. (2002). Community structure in social and biological networks, *Proc. Natl. Acad. Sci. USA* 99, 7821±7826 .

[11] Newman, M. E. J. (2003).The structure and function of complex networks, *SIAM Review* 45, 167-256.

[12] Barabási, A.-L., Albert, R. (1999). Emergence of scaling in random networks, Science, 286. 509-512.

[13] Newman, Neighbor of Neighbor of Initial Contact(2002). Assortative Mixing in Networks, *Phys. Rev. Lett*. 89- 208701.

[14] Newman, M. E. J., Juyong Park (2003). Why social networks are different from other types of networks, Phys. Rev. E 68 036122.

[15] Amaral, L,A,N., Scala, A., Barth, M., Stanley, H, E. (2000). Classes of small-world networks, PNAS.Oct 10.Vol.97, 11149-11152.

[16] Marian Boguna, Romualdo Pastor-Satorras, Albert D1´az-Guilera, Alex. Arenas (2004). Models of social networks based on social distance attachment, *Phys. Rev*. E 70, 056122.

[17] Evans, T., Sarama¨ki, J. (2005). Scale-free networks from self-organization, *Phys. Rev.* E 72, 026138..

[18] Gabor Szabo, Mikko Alava, Janos. Kertesz, Phys. Structural transitions in scale-free networks,*Phys. Rev.* E 67 .

[19] Krapivsky, P, L., Redner, S. (2001). Organization of growing random networks. *Phys. Rev.* E 63, 06613.

[20] Sousa, C., etal. (2008). Social Networking System for Academic Collaboration. LNCS 5220, p. 295–298.