Measuring Homophily in Social Network: Identification of Flow of Inspiring Influence under New Vistas of Evolutionary Dynamics

Hameed Al-Qaheri

Department of Quantitative Methods and
Information Systems
Kuwait Universit, Safat, Kuwait

Soumya Banerjee

Department of Computer   Science
Birla Institute of Technology, Mesra,
India

Abstract—**Interaction with different person leads to different kinds of ideas and sharing or some nourishing effects which might influence others to believe or trust or even join some association and subsequently become the member of that community. This will facilitate to enjoy all kinds of social privileges. These concepts of grouping similar objects can be experienced as well as could be implemented on any Social Networks. The concept of homophily could assist to design the affiliation graph (of similar and close similar entities) of every member of any social network thus identifying the most popular community. In this paper we propose and discuss three tier data-mining algorithms) of a social network and evolutionary dynamics from graph properties perspective (embeddedness, betweenness and graph occupancy). A novel contribution is made in the proposal incorporating the principle of evolutionary dynamics to investigate the graph properties. The work also has been extended towards certain specific introspection about the distribution of the impact, and incentives of evolutionary algorithm for social network based events. The experiments demonstrate the interplay between on-line strategies and social network occupancy to maximize their individual profit levels.**

Keywords—*Homophily; Affiliation; Embeddedness; Betweenness; Graph occupancy; Evolutionary dynamics*

I.   INTRODUCTION

Different properties of social network have demonstrated potential interplay between events, participants and   social network itself.   There are numbers of instances, where, attribute of social network could drive the application area of the network itself [21].   Hence, certain properties of social network have an emerging impact [27] and homophily like behavior is definitely one of them.   Prior research demonstrates impressive role of such behavior on the application and analysis of social network [28] .

Homophily [1] [5] [6] [20], the tendency of individuals to form association with individuals of similar socio-cultural background, becomes the basic governing structural component of any social network and it has been the focus of many social network studies [2] [7]. Social network studies reveal that *"social networks are homogeneous with regard to many socio-demographic, behavioral and interpersonal characteristics"* [3]. In any existing social network such as Facebook or Twitter there lies some common functional attributes such as *'posting photos', 'sending messages', 'likes',*

'dislikes', etc.. These kinds of activities lead to the concept of affiliation towards a community [20]. From influence of social propagation**,** Facebook and Twitter are dedicated to disseminating the information and thus the concept of Twitter follower graph and cascading of influence also reinforces the hypothesis of different influence measurement model [4]. Considering the broader definition of the problem, this paper finds a close similarity between graph theory and a social network homophilic structure and explores the empirical significance of influence propagation or a popular community ranking and detection. The paper validates the existing graph postulates with a proposed mining algorithm and simulation applied a Facebook data set. Investigation yields certain significant results with regards to popular community structure and ranking based on different classical graph theory properties like path traversal, Betweenness and Embeddedness of social network nodes. After intial validation through graph simulation, an initiative has been solicited with a self-organizing and evolutionary principle, which could dynamically trace the variants of social network. The role of evolutionary dynamics [20] is also considerably significant as it is defined as a study of the mathematical principles according to which life has evolved and continues to evolve. The evolution is also visible in the formation of social graph. Reinforcement and validation of graph properties has been demonstrated through the algorithmic strategies coined from evolutionary dynamics [13] [14]**.**The remaining part of the paper is organized as follows: Section 2 elaborates the statement of the problem with the parameters of graph theory followed by existing methodologies, examples of social and affiliation graph in section 2A and motivation of the analysis has been discussed in section 2B. Section 3 describes mathematical treatments responsible for proposed algorithm, presented in section 4. Section 5 discusses the data set for experiments and their implication on the graph properties of social network. Section 5.1 introduced the role of evolutionary dynamics to validate the simple graph properties for social network instances, which may implicate in the mining of graph related inferences. Finally section 6 gives conclusion and mentions further scope of relevance research on the paradigm.

II.   STATEMENT OF THE PROBLEM

A community is formed in order to propagate or transfer or share different knowledge across the network. And to disperse

the utility of one's community, it is required to apply some techniques, such as voting or polling, on which knowledge about the particular community has to be shared. When forming community, certain basic parameters along with the links among the members of the community are to be considered. Parameters and links are as follows:

- Friend list
- Community links
- Graph occupancy period: Initial time and final time
- Paths and connectivity
- Path traversed
- Embeddedness
- Betweenness of nodes
- Affiliation
- Co-evaluation of social and affiliation network

A. Friend List

Friendship is developed in a social site based on some common factor, like as, members belonging to same school, working area, community and so on. Similar type of characteristic people can be blocked into a common structure and even a few from the block can also belong to some other structure based upon their choice. [8] The homophily test of friendship in a social site can be interpreted with the help of the following example. Let there be a network where *"m"* fraction of all individuals are male and *"f"* fraction of all individuals are female. Considering a given edge in this network, if we independently assign each node the gender male with probability *"m"* and the gender female with probability *"f"* then the both ends of the edge will be male with probability *"m^2"* and similarly both ends will be female with probability *"f^2"*. But if the first end of the edge is male and the second end is female or vice versa then there exist *'cross-gender edge'*. This condition will take place with a probability *"2mf"*. Thus the test for homophily according to gender can be summarized as – if the fraction of cross-gender edges is less than *"2mf"* then there is a presence of homophily [8].

B. Community Links

Community can be created by a group of members by selection and social influence method. The tendency of people to form friendships with others who are like them are termed as *selection* **[8].** The selection criteria are mainly race or ethnicity or similar characteristics. People may also modify their behaviors to bring them more closely into alignment with the behavior of their friends. This process is vividly described as *socialization* and *social influence* **[8].**

The individual similar characteristics drive the formation of links but social influence is a mechanism by which the existing links in the network serve to share people's characteristics.

C. Graph occupancy period: Initial time and final time

It is the amounts of time spend on visiting a node. The duration of remaining in a focus is calculated by checking the difference in final time and initial time. This indicates the graph occupancy period.

D. Paths and connectivity

According to the social scientists John Barnes defines graph theory as "Terminological jungle, in which any newcomer may plant a tree."[8] A *path* is defined to be as a sequence of nodes with the property that each consecutive pair in the sequence is connected by an edge. The paths can also be analyzed as not just the nodes but also the sequence of edges linking these nodes. Connectivity can be described by saying a graph is connected if for every pair of nodes, there is a path between them. If a graph is not connected, then it is separated into a set of connected pieces. Connected components of a graph are a subset of the nodes such that the following two properties hold [8].

1) *Every node in the subset has a path to every other [8]*
2) *The subset is not part of some larger set with the property that every node can reach every other [8].*

E. Path Traversed

The path traversal is important in spread of important information. It is required to examine whether something flowing through a network has to travel just a few hops or more. The *"Length of a path"* is the number of steps it contains from beginning to end that is the number of edges in the sequence that comprises it. The path traversal technique used over here is the *"Breadth - First - Search"*. The method of the traversal is one just need to keep discovering nodes layer-by-layer, building each new layer from the nodes that are connected to at least one node in the previous layer. Since it searches the graph outward from a starting node, reaching the closest node first, it is named as breadth-first-search [8].

F. Embeddedness

The number of common neighbors the two end points in a network has referred to as embeddedness of an edge. This is illustrated with the help of a schematic diagram in **schema1**.
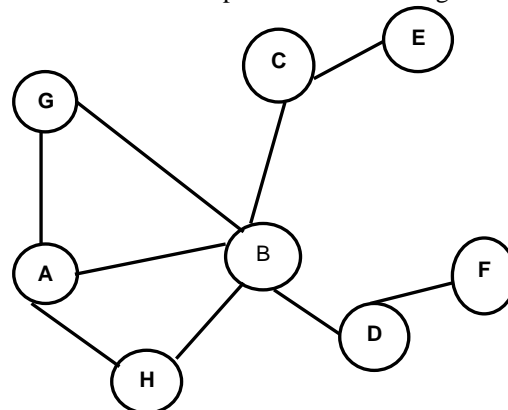

Schema 1 An affiliation network Case

Here, embeddedness for two node A and node B has two common neighbors' node G and node H. Thus the concept of embeddedness provides information that if two individuals are

connected by an embedded edge then this makes it easier for them to develop a trust level and generate confidence for transferring vital information or interacting with each other [8].

G. Betweenness

Betweenness of a node is explained as the total amount of flow that it carries when there exists a unit of flow between each pair of nodes is divided up evenly over shortest path. Nodes with high betweenness occupy critical roles in the network structure. To compute betweenness efficiently we use the notation of breadth-first-search. For a give graphical structure the calculation of betweenness is done on the perspective of time. For each given node the total flow from that node to all other is distributed over the edges. This technique is applied on every node in order to simply add up the flow from all of them to get the betwenness on every edge [8](shown in schema 2).
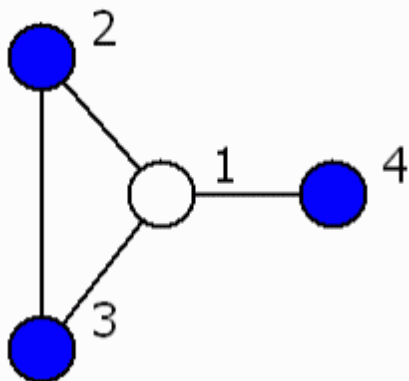
Schema 2: Local betweenness(the local betweenness of actor 1 is 2

H. Affiliation

Affiliation, a concept that is associated with homophily graph, **[8] [9]** can be used to represent the participants, i.e. a set of people, in a set of foci (representing some kind of community). For example, node A, representing a person could participate in focus X through an edge. These kinds of graph are said to be affiliation network, since it represents the affiliation of people (on left) with foci (on right). Affiliation network is one of the examples of the bipartite graph. **Bipartite graph:** A graph is said to be bipartite if its nodes can be divided into two sets in such a way that every edge connects a node in one set to a node in the other set [8].

Scheme 3 is an example showing nodes A and B representing people participating in   and Literature Club and Soft Computing) foci.

I. Co-evolution of social and affiliation networks

New friend links are formed and people become associated with new foci over the period of time.
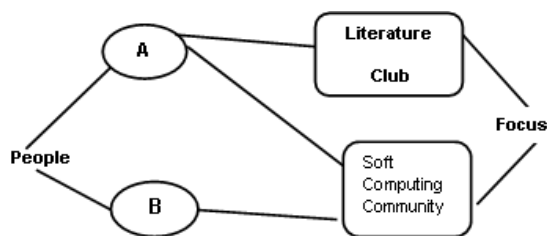
Schema 3 Affiliation through bipartite graph.

This kind of formation can lead to a kind of co-evaluation which might indicate the selection choice of each individual and there social influence. For example if two people belong to a same focus then there is a probability that they become friends and can influence each other with their community they belong. According to the graph theoretic representation nodes are used as both people and foci but the difference is created by distinct type of edges. Firstly, an edge in a social network, it connects two people and indicates friendship. Secondly, an edge in an affiliation network, usually known as 'social-affiliation network'. This edge connects a person to a focus and designates the operation of the person in the focus. These two parameters can be resembled in the following **schema 3**.
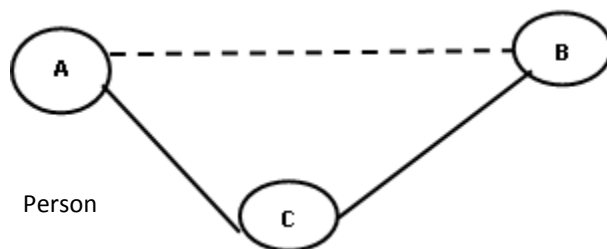
Schema 3 (a) When A, B, C are three different persons.
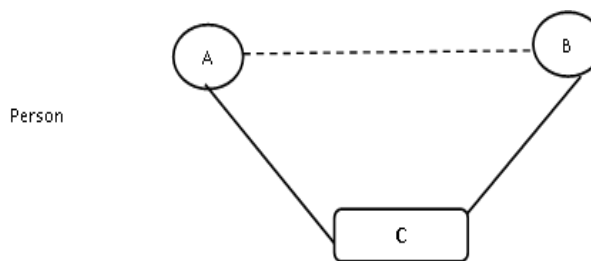
Schema 3 (b) A and B represent people but C denotes focus.

III.   RECENT TRENDS AND MOTIVATION

While exploring the incentive generation, the basic economics model of incentive distribution has become crucial trend to be studied. Research already revealed the impact of incentives on worker self-selection in a controlled and restricted laboratory experiment. Subjects face the choice between a fixed and a variable payment scheme. Considering the status of the treatment, the variable payment is a piece rate, a tournament, or a revenue-sharing scheme [22] [23].

The extension of applying the graph properties of social network is broadly inspired by Julia Poncela Casasnovas's research **[24].** She addressed the study of the evolution of cooperation on complex networks, using among the different social dilemmas. Emphasizing mainly on the Prisoner's Dilemma game as a metaphor of the problem, her research analysed possible outcomes of the dynamics, depending on the underlying topology. Very recently, in 2013, it has been pointed out that the topology not only highlights homophily or other associated properties but also it leads toward role discovery problem. role discovery problem **[25]** finds groups of nodes that share similar topological structure in the graph. But is it only topology or economies of incentive to quantify the distribution of pay off under network? We investigated from significant work on evolutionary games of Nowak et. al. **[26]** and found that even payoff determines reproductive rate and successful individuals have a higher payoff and produce more offspring. Still it cannot be assured that the payoff also could be signified by the carrying capacity of individual participants in networked games. Finally, this work adopts the emerging strategies of evolutionary game based incentive distribution for any instance of social network under test.

IV. EXPLORING MATHEMATICAL TREATMENTS

The most influential community (focus) can be determined by the frequency of the clicks made by the individual nodes. However, visiting a focus and being a follower and then a member of that focus are two different aspects. Visiting a node could mean only collect information while being a member means making the community well know. As such, to find the most influential focus we need to find all possible paths from the source to the destination and finding the shortest path by computing the Betweenness values. Considering the graph one node at a time, the distribution of the total flow over the edge from that node to the other nodes could be computed. And hence, the betweenness of the every node could be calculated as follows:

Betweenness of every node = $\sum$ flows from different nodes

(1)

The shortest path traversed from the source to the destination can be found by the *Breadth-First* Search algorithm. Hence the number of shortest path to each node should be the sum of the number of shortest path to all nodes directly above it in the *breadth-first* search [8]. Valuation of the members of the community or focus can be denoted by the time spent on it by a particular node. Here lies the concept of Graph Occupancy, which can be calculated by calculating the time difference between Final time of leaving the focus and Initial time of entering the focus. Mathematically, this can be expressed as follows:

$$f_i = b^t$$ /* frequency is non-linear relation with

betweenness.   (2)

Where   t={1,2,……..m}

Since we concentrate on each node for calculating the betweenness and thus finding their frequency of participants in making a focus famous thus we have the following set of Betweenness (B_i) and Frequency (F_i).This is expressed as:

$$B_i = \{b_1, b_2, \ldots b_n\} \text{ and } F_i = \{f_1, f_2, \ldots f_n\}$$

Equation 2, describes the frequency of acceptance of a focus by different nodes. Based on a number of choices made by each individual, we can select out the most beneficial or famous (leader) communities, in terms of the population focus. This entire logic can be expressed mathematically as:

$$g(p) = \Pi\left(\sum_{i=1}^{n} B_i + F_i \sin(\Pi p) + C\right) \text{ (3)}$$

where, p is the popularity of a node and C is a constant for each delay unit process. In equation [3], the number of paths to be followed to reach the destination is obtained by the betweenness calculation denoted as B_i and frequency of the particular node can also be obtained and denoted as F_i. The summation of these two values yields the minimum possible path to be traversed to point the popularity of the community. The maximum popularity of a community could be found by the product of this result, in other words, by applying the concept of *Max-Min* function, the maximum popularity of a focus can be obtained for the minimum path travelled from the source to the destination.

V. PROPOSED ALGORITHM -I

Finding the most influential community could be achieved through the following algorithm which has distinct blocks to evaluate the popularity and influence of node(s). The algorithm contains certain unconventional graph properties such as betweenness and embeddedness.

1. Begin

2. Initialization of links present in between the nodes and focus

3. Initialization of variables

Betweeness=b_i,   b= {b_1, b_2, ….b_n }

Graph frequency= (f_i), f_i= {f_1 , f_2,….f_n }

Popularity P

4. Finding value $f_i = b^t$ {t=1, 2, 3….n}

5. Calculating the betweenness:

Betweenness of every node= $\sum$ flows from different nodes

6. Values form the betweenness leads to find the shortest path

7. Calculating the shortest path based on Breadth first search

8. Select a random node n_i which has visited a particular community node at least once $n_i \epsilon N$

9. w*hile* $b_i \leq b_n$
do

       Calculate the Graph frequency f_i

      if*f*$_i \geq f_{i+1}$

then

        $F = f_i$

end if

      Calculate $g(p) = \boldsymbol{\pi} (\sum_{i=1}^{n} B_i + F_i \sin(\pi P) + C)$

10. end while

11.   **end begin**

VI. IMPLEMENTATION AND ANALYSIS

The data set for validating the proposed model was collected from Facebook community network. The reason to choose Facebook was its close resemblance, in terms of nodes, betweenness and edges, to the classical graph theory models. A set of nine homophilic nodes, at a particular time, were selected for initial validation. In these sets, each node is connected to its influential nodes, which is denoted by the links and their weights. There are communities that belong to the nodes and each of these nodes tries to promote their own communities.

To denote the most influential focus or community, we need to find the shortest path (which also motivates the other nodes to join the community and increase the popularity). *Embeddedness* of the nodes are also considered to identify the common neighbors among them. **Table 1** shows the parameters to be considered for the proposed algorithm. The output value in the table was computed after implementing the algorithm (using MATLAB version 7.10.0.499 (R2010 a)). The detailed explanations of the parameters are as follows: **Friend List:** 9 different Nodes represent 9 different friends with their communities

Link Present: edges among the nodes with weights.

Embeddness: A→B edge having common neighbors. There may be present or might not be present. If not present, then denoted by NIL. If present then node number is given.

Betweenness: involves reasoning about the set of all shortest paths, between pairs of nodes.

Graph occupancy: amount of time spent on a node. Leading to the further popularity of a community and increasing the node linkage. The corresponding outputs are given in the following **Table 1.**

TABLE I.   LIST OF NODES WITH THEIR CONNECTIVITY, FLOW OF INFORMATION AND FURTHER INCREASE IN POPULARITY

Parameters						
Friend list	Links present	Embeddedness	Betweenness		Graph occupancy	
Node 1	(4,5,6)	Node 4	(1,5)	0.2100	(1,1)	0.0541
					(1,2)	0.0462
					(1,3)	0.0474
			(1,6)	1.5300	(1,4)	0.0148
					(1,5)	0.0151
					(1,6)	0.1478
Node 2	(3,5,6)	Node 3	(2,3)	0.5100	(2,2)	0.0591
					(2,3)	0.0793
					(2,4)	0.0131
			(2,5)	0.3200	(2,5)	0.0213
					(2,6)	0.0106
Node 3	(4,5,6,7)	Node 4	(3,4)	0.1500	(3,1)	0.0046
					(3,3)	0.0266
					(3,4)	0.0145
					(3,5)	0.0135
					(3,6)	0.0081
Node 4	(1,3,5,6,8)	NIL	(4,1)	0.4500	(4,1)	0.0435
					(4,2)	0.0221
					(4,3)	0.0446
			(4,6)	0.7900	(4,4)	0.0591
					(4,5)	0.0065
					(4,6)	0.1234
Node 5	(1,2,3,4)	NIL	(5,3)	0.3200	(5,1)	0.0111
					(5,2)	0.0135
					(5,3)	0.0827
			(5,4)	0.3600	(5,4)	0.0381
					(5,5)	0.0375
					(5,6)	0.0194
Node 6	(1,2,3,6)	Node 4	(6,2)	0.4100	(6,1)	0.0028
					(6,2)	0.0377
					(6,3)	0.0352
			(6,3)	0.2900	(6,4)	0.0030
					(6,5)	0.0119
					(6,6)	0.0591
Node 7	(3)	NIL	(7,3)	0.4500	(7,2)	0.0191
					(7,3)	0.0267
					(7,4)	0.0046
					(7,5)	0.0070
Node 8	(9)	NIL	(8,9)	0.5500	(8,2)	0.0466
					(8,3)	0.0652
					(8,4)	0.0113
					(8,5)	0.0171
Node 9	(4,8)	NIL	(9,4)	0.3000	(9,1)	0.0092
					(9,3)	0.0531
					(9,4)	0.0290
					(9,5)	0.0271
					(9,6)	0.0162

Fig. 1. Nodes with their links and weights

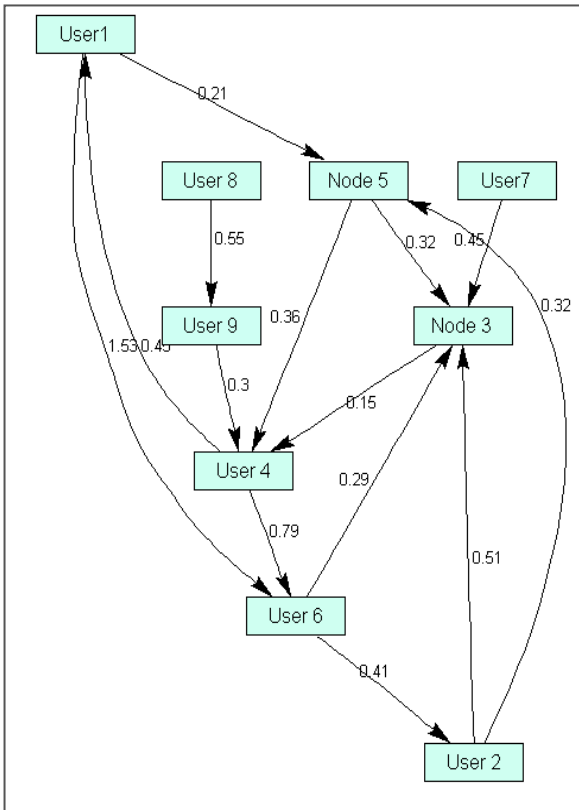Fig. 2. Enhancements in new connections with the existing invoking nodes

Figure 1 represents nine different influential nodes with their links represented in the directed graph format. Each of the edges has their weights marked on it. Here, the shortest path has been considered from a source node say Node1 and the destination node, Node 6.Time duration, that is the amount of time spent by the nodes in their communities, can be an invoking factor to join their community. Using equation (3), **Figure 2**demonstrates that the most influencing or inspiring node enhances the number of functionally active node. In this figure, it is shown that the inspired nodes have added on more of the connectivity with other different nodes.

To crawl Facebook, we implemented a distributed, multithreaded crawler using Python with support for remote method invocation (RMI) **[11].** Facebook provides a feature to show 10 randomly selected users from a given regional network; we performed repeated queries to this service to gather 50 user IDs to "seed" our breadth-first searches of social links on each network[1]**[11].**

A. Role of Evolutionary Dynamics

Based on the intial simulation, it has been demonstrated that there exist a strong cohesive directions with graph theory and social network in the context of homophilic community detection.
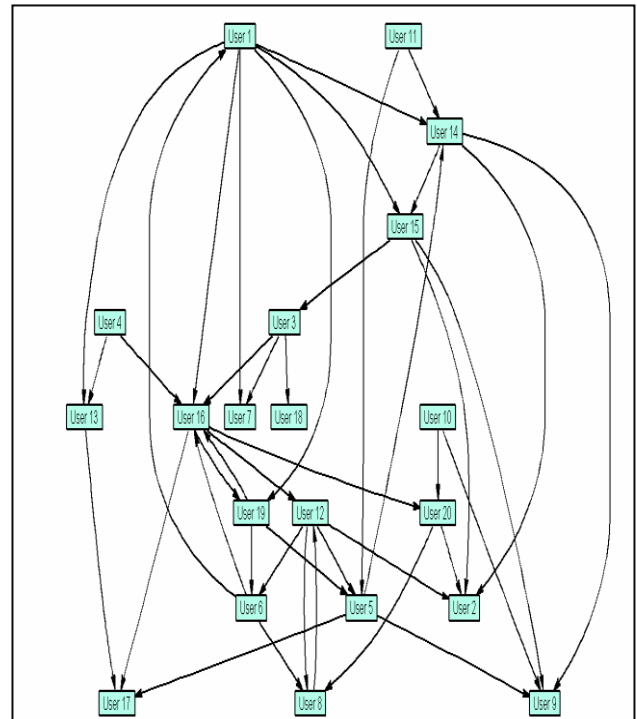
We need further investigation for the specific attributes in the context of more homophily identification from the perspective of either directed or undirected graph. The extension of the algorithm will be significant to quantify the application specific investigation of homophily. The extension also could revalidate the correlation of graph properties under social network.

Inspired by the phenomenal contribution in evolutionary game theory by Nowak and his colleagues, several non linear characteristics have been started adopting the concept of evolutionary dynamics [12] [13]. It is evident that evolutionary dynamics are defined by nonlinear differential equations and therefore can be imported for revalidating complex graph and network with the growth of a social graph as mentioned in the first part of the algorithm. An evolutionary dynamic assigns each population game F an ordinary differential equation [14] $\dot{x} = V^F(x)$ on the simplex x. One simple and general way to define an evolutionary dynamic is via a growth rate function:

$$g: R^n \times x \rightarrow R^n \ (4)$$

Here, g represents the (absolute) growth rate of strategy i and it will as a function of the current payoff toreward the strategy. The previous algorithm only considered the shortest path, breadth first search and graph frequency. As Evolutionary Dynamics can also retrospect the growth aspects, propagation and mutation of message under any state of graph, hence therefore the conventional graph properties have been revalidated using the extended algorithm. It should be mentioned that proposed algorithm tries to incorporate potential strength of evolutionary dynamics for simulating

---

[1]http://code.google.com/p/crawl-e/

the behavior and growth of social graph structure. Series of more precise plots have been accomplished in post simulation of extended algorithm **(Figure 4-8)**. Persistence across the network also indicates the out degree and also any average number of distinct tags of groups and of tag assignments of users having k_{out} neighbours could be evaluated from social network as shown. The users, who have more contacts in the social network, tend also to be more active in terms of tags and groups. Average number of distributive tags denoted as n_t, group tags on the specific message as n_g and subsequently n_w represents the list of predefined choices. Correlations between the activity of participants and their number of declared friends and neighbours can be identified: here also the k_{out} neighbours have been taken into consideration. The data has been log-binned **(Figure 3)**: by definition a bin of constant logarithmic width signifies that the logarithm of the upper edge of a bin (x_i+1) is equal to the logarithm of the lower edge of that bin (x_i) plus the bin width (b) **[19]**. Here, the symbols indicate the average, and the error bars with near optimal 25 and 75 percentiles for each bin. The algorithm deploys the Python 2.7.3, which was released on April 2012. The present implementation also allows the feature of Automatic numbering of fields in the str.format() method, which have been reflected in the post implementation stage of algorithm.

Algorithm II: Graph Pattern $(G, P(t), \mu)$

---

1. Define the initial state of the graph , i.e. define $G_i = \delta_i$

   /* δ_i is the [2]Kronecker symbol*/

2. Solve for Pi(t), which provides functions P(t) and ⋯→i(t)
   /*P(t) : Probability of sample occupancy time for t_m

   Evaluate: $\dfrac{dP}{dt} = -\dfrac{dY}{dt} = -\sum_i \mu_i P_i$

   (5)

   /* The probability Yi(t) that at least one mutation has occurred while the system was at state i before time t*/

3. Initialize the system with N classes of social network instances at time t = 0, with its intialstart of connection, k_{out} :out-degree of graph, occupancy on graph, and termination instance:

   evaluate: $\rho_i(t_m) = \dfrac{\mu_i P_i(t_m)}{\sum_i \mu_i P_i(t_m)}$ (6)

   /* where: probability $\rho_i(t_m)$ that the social network is at state i, P(t): sample occupancy, μ represents probable mutation rate of messages across participants according to the proportion of occupation*/

4. Sample the next mutation time according to the cumulative probability P(t). This can be done via the

   inversion method, such that the next time $t_m = P^{-1}(r)$, where r is a uniform random variable between 0 and 1.

5. Add t_m to the current time of occupancy of participants and betweenness.

6. Choose the specific score and plot according to their respective nore transition of the network and update the state of the system as per Step 4.

7. Remove extinct and redundant classes from the list and reduce the number N of classes accordingly.

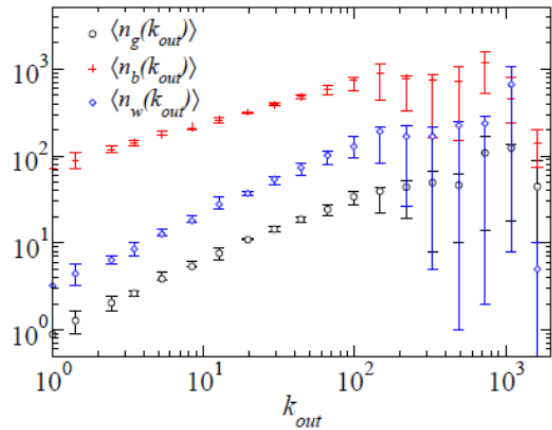8. Return to Step 1 until finished.

Fig. 3. Distribution of tagging with out degree (k_{out})
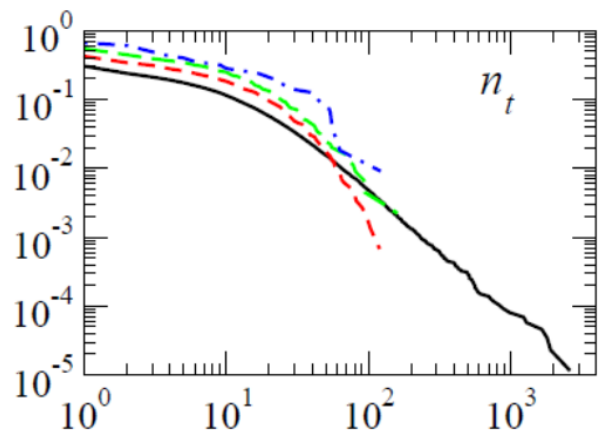

Fig. 4. Degree Correlation and distribution of Message tagging

In case of large scale network like Facebook, the most conventionally investigated mixing pattern involves the degree (number of neighbours) of nodes. This type of mixing improvises the likelihood; leading users with a given number of neighbours connect with users of similar degree. This property is emphasized by computing multi-point degree correlation functions. Complementary cumulative conditional distributions as mentioned with group tagging, specific message tagging and pre-defined choice tagging, compared with the global cumulative distributions denoted by black lines (Figure 4). Even among the subset of users with a given k_{out}, a strong disparity is still observed in the amount of activity and also around a specific community. Subsequently, the

---

[2]It is simply a function on two variables, i and j which are integers, when for each social network instances cardinality of the variables is large, then it could assist to approximate inference based with a constrained, lower complexity, adaptively sized sum for the target cardinal value **[10].**

occupancy on a specific graph instance results in the following plot: Log-log plot of the distribution of the contact durations and of the cumulated duration of all the contacts two individuals' m and n have over a day (w_{mn}). An interesting inference could be drawn that out of 88% of the total contacts sustained less than 1 minute on a specific tag or comments, but more than 0.2% persisted more than 5 minutes against a specific topic of interest. For the cumulated durations, 64% of the total duration of contacts between two individuals during one day last less than 2 minutes, but 9% last more than 10 minutes and 0.38% more than 1 hour. The small symbols reciprocate to the actual distributions, and the large symbols to the log-binned distributions [18] (Figure 5).
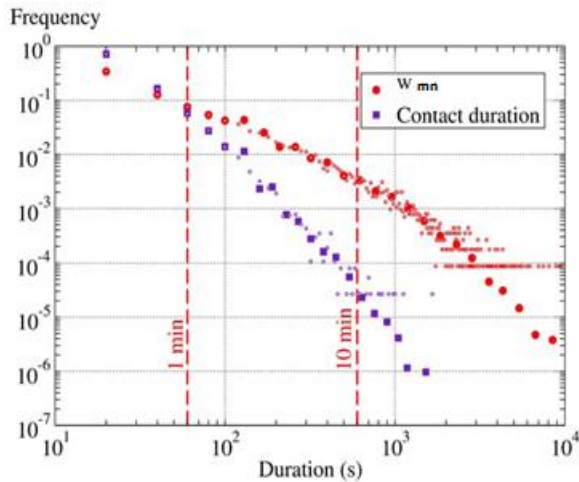

Fig. 5. Occupancy Graph for Social Network with Mutation Probability of Messages

Degree distributions of evolutionary dynamics and the empirical data: 'o' represents the data points from the simulated data and 'x' represents the data points from the empirical data new project The results of the degree distributions of the extended algorithm with dynamic values of mutation are shown in Figure 6. The upper figures are the comparison of developer and project degree distributions in linear coordinates.

The lower figures are the comparison of developer and project degree distributions in log-log coordinates. The R_2 of developer degree distribution from the simulated data in lower figure is 0.959 and the R_2 of project degree distribution from the simulated data in lower figure is 0.7657. Also the largest project size of the simulated data is just 1500. We can further lower this value by tuning the mutation parameter μ [17] and P (t),and graph function of the extended algorithm. H log represents homophily log evaluated from message exchanged towards any specific and common interest [17].
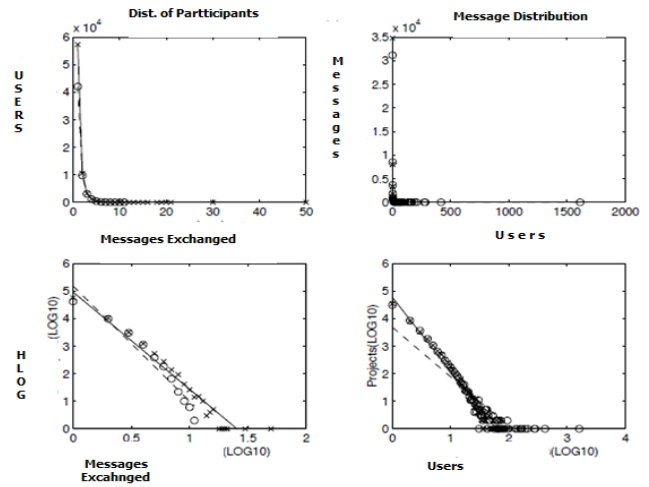

Fig. 6. Evolutionary Dynamics plot with varying mutation and Probability of occupancy
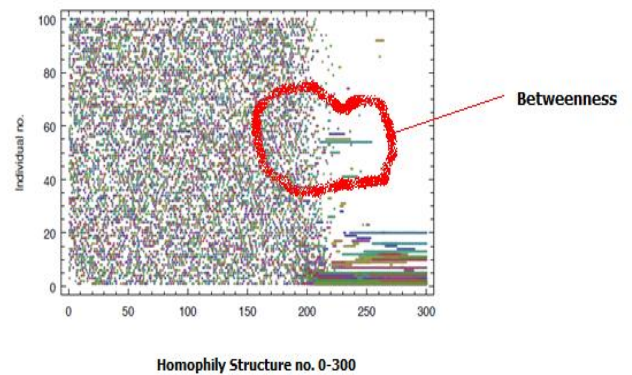

Fig. 7. Recursive simulation on test function towards Homophily from Betweenness

In order to quantify the complex large cardinal network, we incorporate classical test function like DeJong[3](The conventional De Jong's functions is the so-called sphere function, this function is unimodal and convex by nature)[16].Dimension is represented in the formulas by variable D, so as can be observed that it becomes simple to calculate selected functions for an arbitrary dimension for arbitrary number of participants. Similarly, the result of homphily structure from intial betweeness and other asscoited properties is safely turned out from local mimima due to the size of its popluation and avaiable best solution.The extended part of the algorithm demonstrates the frquencey of interplay and incentive distribution betwwen different communities played in social network.

$$1 + \sum_{i=1}^{D} \frac{x_i^2}{4000} - \prod_{i=1}^{D} \cos\left(\frac{x_i}{\sqrt{i}}\right)$$

represents Dejong Function

Algorithm III   Distribution in *Population graph*

The working strategy

Variables: N./*Homogeneous population of size*/

t /* Time Step t*/

r /* relative fitness */

/*temporal event could be Binomial time step[4] **(equation 8)** of propagation of base event under social network , if randomly chosen individual from SN then fixation of probability of new mutant μ_1 represents probable mutation rate of messages across participants according to the proportion of occupation*/

$$\mu_1 = \frac{1-1/r}{1-1/r^N} \qquad (7)$$

$$\mu_1 = [\frac{1-1/r}{1-1/r^N}][t_i + t^2_{i+1} + t^4_{2(i+1)} + \dots + t^n_{2i!/N} \qquad (8)$$

The distribution of trend on population graph of the new attributes concerning the people refers category of homophily structures (shown in red line) where the majority of people in the population have lower levels of the attribute. Gradually, as time elapses, the most influential message/ broadcast configuration propagates across the population (indicated by green line) but   subsequently; evolution dynamics pushes the higher level of the attribute with certain incentive. This is natural recipient of incentives under social network for those persons who are participating and interacting. In addition to, there are drifts of incentive distribution under different settings. Hence, eventually after a considerable period of evolutionary dynamic, drift increases at the high end of the attribute for homophilic structure and the distribution approaches reverts back again to normal.
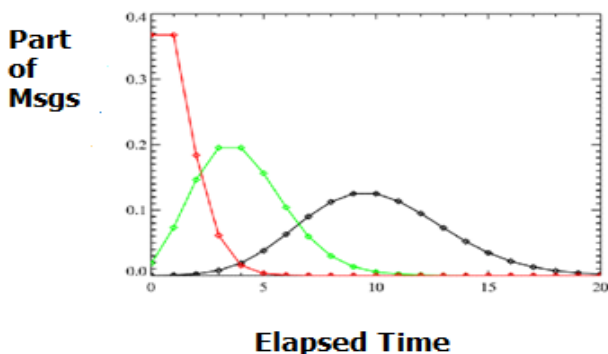

Fig. 8.   Time versus messages
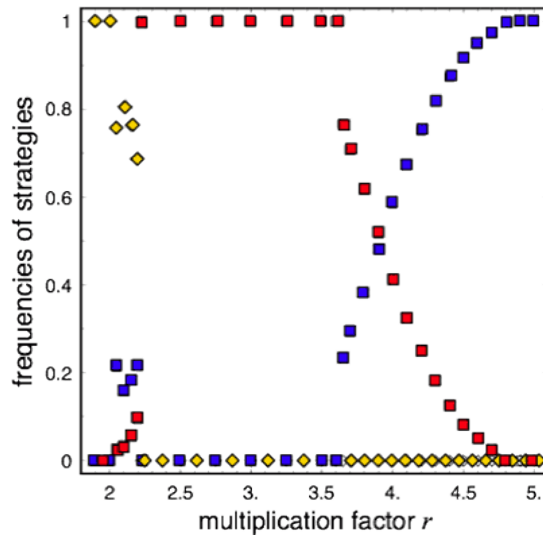
---

[4]   Refer to the Appendix for Proof

Fig. 9.   Simulation of Pay off distribution

At this stage of simulation, it was not evident   how the appropriate interaction could entail higher range of incentive among the participants. Therefore, a specific data set was chosen to exhibit different interplay among the social participants. The observations show that the **f**requency of cooperators (indicated by blue), deviators (indicated byred) and loners (represented as yellow) under smooth participation. This is also measured as the multiplication factor according to trust of discussion and interaction. Individuals are arranged on random regular graphs where each node has eight neighbors and they interact in randomly formed groups of size $N = 5$. For small multiplication factors,individuals dominate. The reason is simple: in this case even in a group of cooperators, the payoffs do not exceed the incentive of individual.   All of the proposed components of three algorithms improvise dynamic equilibrium considering the population of social network. The range of multiplication with trust increases the pay off a very small discrimination of$r$.And above the threshold deviators exist.. Only for much larger value of $r{\sim}4.056$ cooperators reappear and co-exist with deviators. Since individuals are absent, the dynamics again retain voluntary participation into compulsory interaction. Finally, for r= 4.6 cooperators take over and manage to displace deviators **(see Figure 9 and table 1).**

	Interaction (Session 1)		Interaction (Session 2)		.......Interaction (Session n)	
	Incentive		Incentive		Incentive	
	Intial	Final	Intial	Final	Intial	Final
Group 1	10 Units	Ω	10 units	Ω	10 units	Ω+ donated value
Group 2	10 Units	Ω	5/7/12 units	Ω+ intial value	6/7/15 units	Ω+ intial value
Group n	10 Units	Ω	10 units	Ω	Ω	Not defined

Table 1. Configuring design of Incentives on SN

The proposed design involves 4 stages and 4 groups described below and summarized in the following Table:

- Group 1: This group comprises of more than 65 subjects. The default initial value of the participating reward is 10 units and the final value could be measured as Ω.

- Group 2: This group represents almost 50 subjects with initial incentive but in subsequent sessions the experience differs with different fractions of incentives depending on the frequency of interaction as shown in Figure 8.

- Group n: The group enhances its subject line > 100, but in addition to the normal incentive distribution, there will be different treatments for incentive either there should be donation or additional amount the participants will push into it.

B. Comment on final Homophily Sturucure

The homphily evlalution from the social graph is completely based on the increase of posts, tags and other social network action artifacts. Keeping in mind about the non-linear aspects and growth strategy of socail network, evolutionary dynamics, comparatively better visualization of density of attributes has become possible. Especially, betweenness and graph occupancey and mutation could be the one of those key factors. Recent research of Tom A.B. Snijders and his colleagues [15] demonstrated the relation between evolutionary dynamics and homophily of social network in some common aspects of friendship, recommendation, group study and selection etc. **Figure 7** exhibits trend of precise visualization of homophily structure as guided by the extended algorithm with mutation as prime parameter. The following table 2 should be considered to understand the relationship shown in Figure 8: the red circle signifies higher cluster density and therefore providing the remaining count of homophily over the population.

Population variables	Mean	Std. Dev.	AutoCorr. St.No.	Homophily	Record Set tracking From snap.stanford.edu/data/
Male and Female	0.53	0.51	0.075	55	240000
Age	28.76	8.59	0.365	0-160	197000
Density of Propagation	24.60	13.21	0.35	160-240	230615
Network Variables					
Out Degree	*301.4*	*177.21*	0.15	58- 90	210043
Local Betweeness	*34.74*	*91.76*	0.017	170>	228456

VII. CONCLUSION AND FURTHER SCOPE OF RESEARCH

Affiliation being an important factor of homophily graph has a relevant role in promoting the focus through the nodes. The idea is how a friend in the Facebook or any social network can influence his friends to join the community he or she belongs to. This paper investigates such possibilities by exploring the homophily community in social network by using graph property based algorithm and simulation. Further, incorporation of evolutionary dynamics also contributed for investigating homophily property with better approximation. Subsequently, the growth of social network, temporal behavior and trend of it, can also be investigated by augmenting the existing evolutionary dynamics algorithm. On the other hand, the ranking algorithm or conventional classifier model can also be extended using initial attributes of embeddedness and graph traversal with graph occupancy time. Soft computing based (fuzzy and rough set) homophily identification from graph properties could be an emerging research on computational social network. Finally, in the extended analysis part, certain significant observations are made in terms of maximizing benefit or profit, based on their role at specific instances under social network. Experiments have incorporated certain public data sources to demonstrate mutual interplay of social network participants, their specific contribution towards the network and share of incentive if any.

REFERENCES

[1] Vojtek, P., &Bieliková, M. Homophily of neighborhood in Graph Relational Classifier. Pages 721–730 of: Geffert, V.,Karhumäki, J., Bertoni, A., Preneel, B., Návrat, P., &Bieliková, M. (eds), SOFSEM 2010: 36th Conf. on Current Trends in Theory and Practice of Computer Science, SpindleruvMlyn, Czech Republic, 2010, Proc. LNCS, vol. 5901. Springer.

[2] Halil Bisgin, Nitin Agarwal and Xia oweiXu, Investigating Homophily in Online Social Network, 2010,IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Web Technology, pp. 533-536.

[3] M. McPherson, L Smith Lovin and J.M. Cook, Birds of a feather: Homophily in Social Network, Annual Review of Sociology, 27(1): pp. 415-444, 2001.

[4] Eytan Bakshy, Winter A mason, Jake M.H .Ofman, Duncan J. watts, Everyone's an Influencer : Quantifying Influence on Twitter , published in Proceedings WSDM 2011, Hongkong, February 2011.

[5] Reza Zafarani, William D Cole, and Huan Liu. Sentiment Propagation in Social Networks A Case Study in LiveJournal. In Patricia Chai, Sun-Ki And Salerno, John And Mabry, editor, SBP 2010 - Advances in SocialComputing, Lecture Notes in Computer Science 6007, pp. 413–420. Springer Berlin, 2010.

[6] James H. Fowler, Jaime E. Settle, and Nicholas A. Christakis.Correlated genotypes in friendship networks.PNAS, pp. 108:1993, 2011.

[7] Cosma Rohilla Shalizi and Andrew C. Thomas. Homophily and Contagion Are Generically Confounded in Observational Social Network Studies. Sociological Methods and Research (arXiv:1004.4704v3), 2011.

[8]   *Networks, Crowds, and Markets: Reasoning about a Highly Connected World.* By David Easley and Jon Kleinberg.Cambridge University Press, 2010.

[9]   Benjamin Goluby Matthew O. Jackson, How Homophily Affects the Speed of Learning and Best-Response Dynamics, Q*uarterly Journal of Economics* Vol. 127, Iss. 3, pp 1287--1338, 2012.

[10] Chris Pal, Charles Sutton, Andrew McCallum. Constrained Kronecker Deltas for Fast Approximate Inference and Estimation.Submitted to UAI, 2005.

[11] Bryce Boe and Christo Wilson.User Interactions in Social Networks and their   Implications, accessableat http://cs.ucsb.edu/~bboe/public/pubs/interaction-eurosys09.pdfcrawl-e: Highly distributed web crawling framework written in python. Google Code, 2008.

[12] Chatterjee K, D Zufferey, MA Nowak (2012). Evolutionary game dynamics in populations with different learners.*JtheorBiol* 301: 161-173. http://dx.doi.org/10.1016/j.jtbi.2012.02.021

[13] Allen B, MA Nowak (2012). Evolutionary shift dynamics on a cycle.*JtheorBiol*   311:   28-39. http://dx.doi.org/10.1016/j.jtbi.2012.07.006.

[14] Survival of dominated strategies under evolutionary dynamics, Josef Hofbauer and William H. Sandholm,

Theoretical Economics 6 (2011), pp. 341–377.

[15] Alessandro Lomi, Tom A.B. Snijders, Christian E.G. Steglich, and Vanina Jasmine Torló (2011). Why Are Some More Peer Than Others? Evidence from a Longitudinal Study of Social Networks and Individual Academic Performance. *Social Science Research*, 40 (2011), 1506-1520.

[16] X.-S. Yang, Test problems in optimization, in: Engineering Optimization: An Introduction with Metaheuristic Applications (EdsXin-She Yang), John Wiley & Sons, 2010.

[17] Y. Gao,G. Madey, V. Freeh, Modeling and Simulation of a Complex Social   System: A   Case   Study,   accessible   at http://www3.nd.edu/~oss/Papers/SimulationSym37_gao.pdf

[18] J. Stehlé, and Et al, High-Resolution Measurements of Face-to-Face Contact Patterns in a Primary School.Published: Aug 16, 2011,DOI: 10.1371/journal.pone.0023176

[19] E. P. White, B. J. Enquist, and J. L. Green. Ecological On estimating the exponents of power-law frequency distributions.   Archives E089-052-A1.   2008.   Ecology   89:905-912.accessable   at http://www.esapubs.org/archive/ecol/E089/052/appendix-A.pdf

[20] Al-Qaheri, H, Banerjee , S., Ghosh, G. Evaluating the power of homophily and graph properties in Social Network: Measuring the flow of inspiring influence using evolutionary dynamics, Science and Information Conference (SAI), London, 2013 , IEEE XPlore, pp. 294 – 303.

[21] Himabindu Lakkaraju, Julian McAuley, Jure Leskovec What's in a name? Understanding the Interplay between Titles, Content, and

Communities in Social Media, Association for the Advancement of Artificial Intelligence (2013

[22] Dohmen, T., and Falk, A., Performance Pay and   Multi-Dimensional Sorting: Productivity, Preferences and   Gender, American Economic Review, 101(2), pp. 556-590, 2011**.**

[23] DellaVigna, S., J. A. List, and Malmendier U.,   Testing for Altruism and Social Pressure in Charitable,   Giving, Quarterly Journal of Economics, 127(1), pp.1-56. 2012.

[24] J. Poncela Casasnovas, Evolutionary Games in Complex Topologies, Springer Theses DOI: 10.1007/978-3-642-30117-9, Springer-Verlag Berlin Heidelberg 2012.

[25] Sean Gilpin, Tina Eliassi-Rad and Ian Davidson , Guided Learning for Role Discovery (GLRD): Framework, Algorithms, and Applications KDD'13, August 11–14, 2013, Chicago, Illinois, USA.

[26] Sebastian Novak, K. Chatterjee and M.Nowak, Density games Journal of Theoretical Biology 334, pp.26–34, 2013.

[27] Himabindu Lakkaraju, Indrajit Bhattacharya, Chiranjib Bhattacharyya, Dynamic Multi-Relational Chinese Restaurant Process   for Analyzing Influences on Users in Social Media .IEEE International Conference on Data Mining (ICDM), 2012.

[28] Debanjan Mahata and Nitin Agarwal. Grouping the Similar among the Disconnected Bloggers, Social Network Analysis and Social Media Mining: Emerging Research. Guandong Xu and Lin Li (Eds.). IGI Global, 2012.

APPENDIX: PROOF OF STATEMENT IN EQUATION 8:

The proposed model is to trace   the time distribution mode in between different intermediate events. Considering the normal binomial distribution function, which specifies the number of times (x) that an event occurs in n independent trials where p is the probability of the event occurring in a single trial.   Why we considered this trial? As, the exact probability distribution for any number of discrete trials may represent the number of mutant messages, therefore it becomes obvious that for any discrete time t, say for any temporal event the value of number of events could be large. Hence function becomes continuous. Thus relative fitness r also changes respective to bi-nominal distribution of mutant messages.

$$\text{Hence: } \mu_1 = \left[\frac{1-1/r}{1-1/r^N}\right][t_i + t^2_{i+1} + t^4_{2(i\ +1)} + \ldots + t^n_{2i!/N}$$

is a normal representation of r with bi-nomial time t