Evolving Social Networks via Friend Recommendations

Amit Kumar Verma
Department of Computer Science and Engineering
National Institute of Technology Meghalaya, Shillong
India - 793003
Email: mt4descentis@gmail.com

Manjish Pal
Department of Computer Science and Engineering
National Institute of Technology Meghalaya, Shillong
India - 793003
Email: manjishster@gmail.com

***Abstract*—A social network grows over a period of time with the formation of new connections and relations. In recent years we have witnessed a massive growth of online social networks like Facebook, Twitter etc. So it has become a problem of extreme importance to know the destiny of these networks. Thus predicting the evolution of a social network is a question of extreme importance. A good model for evolution of a social network can help in understanding the properties responsible for the changes occurring in a network structure. In this paper we propose such a model for evolution of social networks. We model the social network as an undirected graph where nodes represent people and edges represent the *friendship* between them. We define the evolution process as a set of rules which resembles very closely to how a social network grows in real life. We simulate the evolution process and show, how starting from an initial network, a network evolves using this model. We also discuss how our model can be used to model various complex social networks other than online social networks like political networks, various organizations etc..**

***Keywords*—Social Networks, Friend Recommendations, Graphs, Communities.**

I. Introduction

The analysis of online Social Networks has allowed us to answer many questions regarding the characteristics of network and how the network changes. Modeling the evolution of a social network and predicting the structure of the future network is a complex problem. Social networks grow and change quickly over time with the addition of new edges, signifying the appearance of new interactions/relations in the underlying social structure. In this work, we consider a social network as a network comprising of members which are connected in the network by the "friendship" relation. We try to understand the mechanisms by which the social network evolves over time and using this information we design a model which allows us to predict the structure of the future network.

In order to model the evolution of a Social Network we have to know the network characteristics, as in many evolution studies, the underlying process for network change is assumed to be centered at the behavioral characteristics of the network members [3]. In a social network, network members tend to 'choose' their friends by comparing relevant individual characteristics of the others with their own. A fundamental finding in many choice networks is that social members with similar characteristics are more often connected with one another than with more dissimilar ones. This is known as the "similarity effect" in social networks [8]. So we can infer that for any relationship between the network members of a social network there must be some common characteristics between those members which lead to that relation. This basic idea can be used to design the evolution model. In our work we call these common characteristics as *factors*. For example in online social networks like Facebook we can observe certain *factors* namely frequency with which friends tag each other in their posts, place where they live, work place (school, office, university etc.), common interests (movies, songs, books) etc. These *factors* signify how similar the members of the social network are. Considering more *factors* we can come up with a model that can provide a more precise measure of the similarity amongst the individuals. It has also been shown in [5] that gender plays an important role in deciding the level of friendship in people thus gender can also be considered as a *factor*. The strength of the friendship depends on the number of common *factors* as well as the weight of these common *factors*. It means that we can associate the term "quality" with these relationships which shows how strong is the relationship among the members.

In the following section we will discuss some prior work regarding evolution models. Subsequently, we will describe the model of evolution proposed by us.

A. Related Work

Several researchers have turned their attention to the evolution of social networks at a global scale. For example networks become denser over time, in the sense that the number of edges grows super-linearly with number of nodes [7]. In this paper they reported that the network diameter often shrinks over time, in contrast to the conventional concept that such distance measures should increase slowly as a function of the number of nodes. Some efforts has also been made in the direction to find the properties responsible for the network evolution. A variety of network formation strategies were investigated showing that *edge locality* plays a critical role in network evolution [6]. Many models have been designed to predict the links in social networks for example in [2] they introduced the notion of graph evolution rules in which they developed Graph Evolution Rule Miner(GERM) software to extract the rules responsible for network evolution and applied these rules to predict the future network. In the direction of basic principles responsible for social network evolution researchers have shown that the most important characteristic of social evolution is that the outcome of evolution process is not the result of central authority but are the consequences of the simultaneous choices of persons [9][12][13]. Social actors try to realize their own goals by choosing between behavioral alternatives that are available to them under certain restriction [11]. In [10] they have reviewed, classified and compared different models of social networks. They have classified these models into two main categories, first in which the addition of new links depends on the local network structure(*Network Evolution Models*), and second in which probability of each link existing depends only on nodal attributes (*Nodal Attribute Models*). In their work, they have shown that *Nodal Attribute Models* produce very clear community structure. In next section we will discuss about our work and will describe our evolution model.

II. Our Contribution

In this section we discuss about the evolution model that we have proposed. The basic idea that we have used to design this evolution model is the *transitive* property among the relationships i.e if a person A is connected to a person B who in turn is connected to a person C, then there is high possibility that person A and C will be connected. We have designed the evolution model which uses this transitive property to evolve the network. In our model we represent the network as undirected graph such that nodes represent people and edges represent connections between them. These connections can be of any form of friendship relation like "friends", "acquaintance", "co-workers" etc.

A. Preliminaries

We have modeled the social network as a graph $G = (V, E)$ where the nodes(V) represent the people present in the network and an edge $e = (u, v) \in E$ represents the friendship relation between them. As we have already discussed that for any social relation there must be some common characteristics (*factors*) which decides these relations. For this purpose we have associated *factors* with each edge. For simplicity, in our model we have assumed that the number of *factors* is finite. With each edge we associate certain *factors* which represent the common attributes between the nodes which are connected by that edge. Let F be a finite set of *factors* s.t $F \subset \mathbb{N}$ and $F_j \subseteq F$ be the *factors* of an edge e_j for $j = 1, 2, 3, ..., m$ where m is the total number of edges. To decide the "quality" of *friendship* we have associated *score* values with these *factors*. These values will decide at what degree the *factors* between the people are similar and the cumulative *score* value will be the measure of "quality" of *friendship*. For each factor in set F_j a *score* value s is given to it. For example if $f_1, f_2, f_3, f_4, ..., f_k$ is a set of *factors* for an edge e_j then $s_1, s_2, s_3, s_4, ..., s_k$ are the *score* values associated with these *factors*. The cumulative *score* value for each set of *factors* F_j on edge e_j is denoted by S_j which is equal to the $\sum_{i=1}^{k} s_i$. Now we have defined the structure of our graph $G = (V, E)$ in which V is the set of nodes and E is the set of edges and each edge e_j is associated with a set of *factors* F_j and a cumulative *score* value S_j for $j = 0, 1, 2, 3, ..., m$ where m is total number of edges in graph G. In next section we will describe the evolution process for our model.

B. Process of Evolution

Let at time $i = 0$, initial graph be G_0 and set of score values be $score_0$ which is the union of the all the scores of the edges present in G_0. At every time step we will try to add edges between the nodes according to the evolution rules. Suppose at time stamp i, $G_i = (V, E_i)$ is the graph evolved from the initial graph then for all $i \geq 0$, $G_i = (V, E_i)$ is subgraph of $G_{i+1} = (V, E_{i+1})$ i.e. $E_i \subseteq E_{i+1}$. According to our rules of evolution E_{i+1} is obtained from E_i. The set of scores of edges of G_i is denoted by $S_i = \{s_{uv} | (u, v) \in E_i\}$. In this model, once an edge is added to the graph during the evolution process, it is never deleted in any future time-stamp. Also the number of nodes in the graph is fixed. We assume that the set of factors F_e for an edge e will never change during the evolution process, i.e if F_e is the set of *factors* for an edge e at time step i then at any time step $k > i$, in G_k, F_e will remain to be the set of *factors* for the edge e. We have also considered a model in which we are allowed to add some nodes in the graph, but once added it will never be removed. In our next section we will describe the evolution rules on the basis of which the evolution of the network is done.

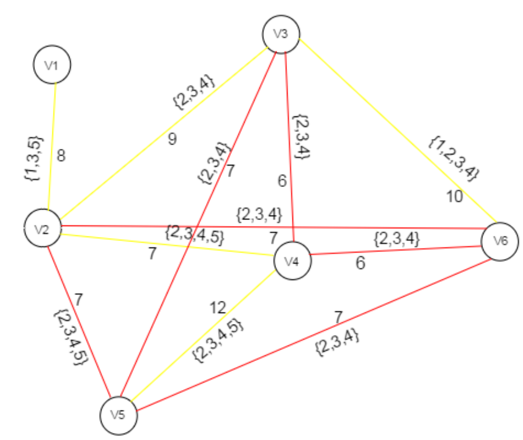

Fig. 1. An example with six nodes

C. Rules of Evolution

In this section we present the rules of evolution of $(G_i, score_i)$ which lead to some interesting properties of the social network graph. Starting from an initial social network these rules help in predicting the state of the social network after several steps. Our main assumption in the evolution process is that to add a new edge (u, v) say at the i^{th} step there has to be another vertex w such that (u, w) and (w, v) are already existing edges in the graph. This modelling is based on a general real life observation that in order for two strangers to become *friends* there is usually a common friend. We also ensure that this edge is added if the *cumulative score* is more than a particular threshold t. We define this cumulative score by first taking the intersection of the sets $F_{(u,w)}$ and $F_{(w,v)}$ and then computing the maximum of $\frac{|F_{(u,w)} \bigcap F_{(w,v)}|}{2} \cdot \left(\frac{S_{(u,w)}}{|F_{(u,w)}|} + \frac{S_{(w,v)}}{|F_{(w,v)}|} \right)$ over all possible $w \in V$. Notice that in order to define the cumulative score we have considered the *arithmetic mean* of the values $\frac{S_{(u,w)}}{|F_{(u,w)}|}$ and

$\frac{S_{(w,v)}}{|F_{(w,v)}|}$ and scaled them by the cardinality of the set of common *factors*. The intuitive reason for this score function is as follows:

- The quantity $\frac{S_{(u,w)}}{|F_{(u,w)}|}$ represents the score of a *factor* for a particular edge and thus by taking an arithmetic mean of the two quantities we get the average score of a *factor* that is common in both $F_{(u,w)}$ and $F_{(w,v)}$.
- Multiplying the arithmetic mean by $|F_{(u,w)} \bigcap F_{(w,v)}|$ scales this weight by the number of *factors* common to both $F_{(u,w)}$ and $F_{(w,v)}$.

Notice that instead of taking arithmetic mean of the two quantities we can also take the geometric mean or harmonic mean to get an average score. Also despite the fact we can recommend an edge that has a cumulative score of an more than the threshold there is still a chance that two people might decide to not be *friends* because of some random uncorrelated event. Thus in order to take care of this issue we make our evolution rules randomized i.e. even though an edge has a cumulative score more than the threshold it is added in the evolution process with probability p and rejected with probability $1 - p$. In the following we formally describe the rules of evolution of the network.

Algorithm 1 Evolution Process

$i \leftarrow 0$
while no more edges can be added **do**
for each (u, v) is not added, and $\forall$ w such that $(u, w), (w, v) \in E_i$ find,

$$z = \arg\max_w \left(\frac{k}{2} \cdot \left(\frac{s_{uw}}{k_{uw}} + \frac{s_{wv}}{k_{wv}}\right)\right) \text{ where } k = F_{uw} \bigcap F_{wv}.$$

if $(\frac{k}{2} \cdot (\frac{s_{uz}}{k_{uz}} + \frac{s_{zv}}{k_{zv}}))_w > t$ **then**
add (u, v) with probability p and discard with probability $1 - p$.
$i \leftarrow i + 1$

In figure 1 we can see the example of evolution. In this figure the network is made up of six nodes with starting connections marked as yellow, and after evolving this network we get the connections which are marked in red. For example V_2 is connected to V_3 with the set of *factors* F_{V_2,V_3} as {2,3,4} and *score* value s_{V_2,V_3} as 9 and V_3 is connected to V_6 with the set of *factors* F_{V_3,V_6} as {1,2,3,4} and *score* value s_{V_3,V_3} as 10. So after applying the score function we get the intersection of *factors* as {2,3,4} and score value as 7. Since this calculated score value is greater than the threshold value (which is 6), this new link V_2 to V_6 is added as a new connection. In our evolution model we add the edge with some probability. This process will repeat until there are no new edges that can be added.

III. Experiments and Result

To implement our evolution model we have considered some assumptions. The input to the evolution process is a graph which we called as "initial graph", so to produce these initial graphs we designed a program which produces random graph by putting random edges between the nodes. So each time we require a graph we run this program to get an initial graph which will act as input to our evolution model. As we have described that the set of *factors* which are responsible for addition of edges can vary from person to person, we have assigned *factors* on each edge randomly. So on each edge e_j we have assigned a set of *factors* F_j randomly from the set of total *factors* F. To decide the quality of "Friendship" we have assigned score values on each edge randomly as well.

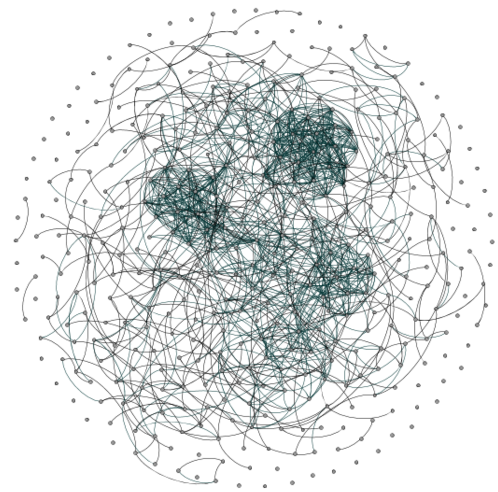

Fig. 2. Evolution process on 400 nodes using arithmetic mean

Consider an experiment in which the initial graph has 400 nodes and randomly placed edges. On each edge e_j the set of *factors* F_j that has been assigned is taken randomly from the set of all *factors* $F = \{1, 2, 3, ..., 8\}$. The score value on each edge e_j has been also taken randomly from the set of all score values $S = \{1, 2, 3, 4, ..., 16\}$. On this initial graph we apply our evolution process to get the evolved graph. To visualize the graph we have used Gephi tool. In figure 2 we can see the evolved graph.

A. Experiment with different mean

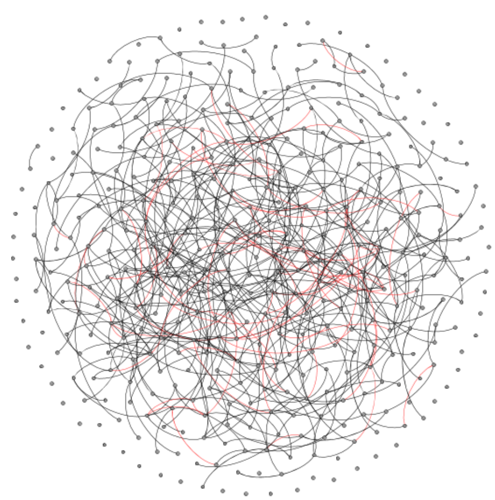

Fig. 3. Evolution process on 400 nodes using geometric mean

In the previous experiment we have calculated the cumulative *score* value by taking the *arithmetic mean* of the individual *score* values. The new connections are made by calculating the cumulative *score* values of the common *factors* of a transitive relation. We consider another experiment on 400 nodes again with a random initial graph, but instead of taking the arithmetic mean of the score values we take their geometric mean to compute the cumulative *score* values. In figure 3 the evolution does not show the formation of communities as was established by the previous experiment.

B. Experiments with different number of nodes

In this section we will show the experiment results for different number of nodes.

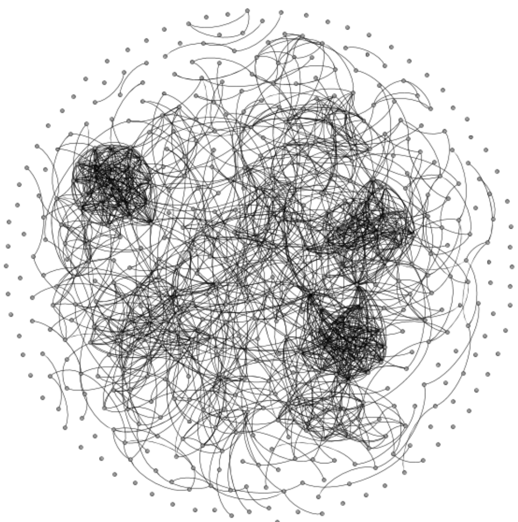

Fig. 4. Evolution process on 500 nodes

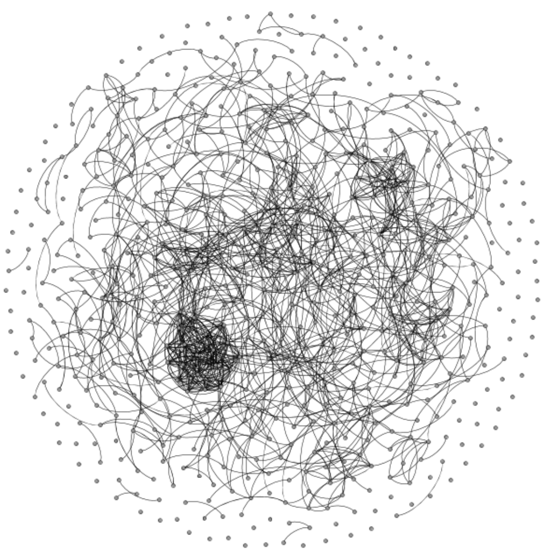

Fig. 5. Evolution process on 600 nodes

C. An iterative evolution process

We have proposed another extension of our model in which rather than applying the evolution process to fixed number of nodes we are iteratively adding more nodes to the graph and applying the evolution process. This process will not terminate as nodes will be added at every stage of evolution so we have taken the snapshot of graph at time T. In this process random number of nodes from a new set of nodes V_{new} are added at every time step T. These nodes are randomly connected to some nodes which are already present in the network and on this modified graph the evolution process is applied. So at some time step T we produce the graph which can be considered as a snapshot of the evolution process. The following algorithm shows the iterative evolution process.

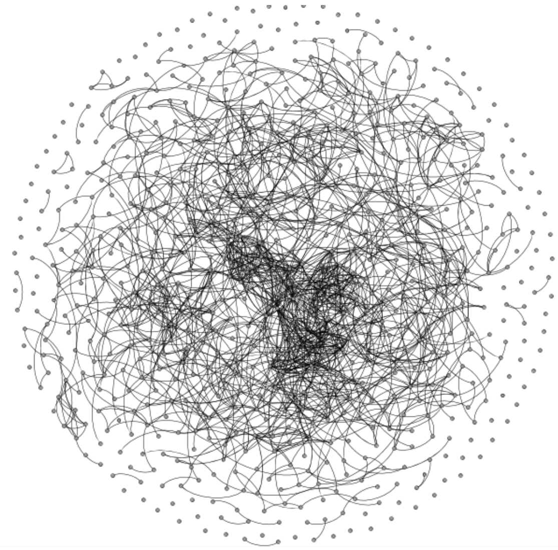

Fig. 6. Evolution process on 700 nodes

Algorithm 2 Iterative Evolution Process

while no more new nodes are added **do**
  $i \leftarrow 0$
  while no more edges can be added to G_i **do**
    for each (u, v) is not added, and $\forall\ w$ such that $(u, w), (w, v) \in E_i$ find,

$$z = \arg\max_w \left(\frac{k}{2} \cdot \left(\frac{s_{uw}}{k_{uw}} + \frac{s_{wv}}{k_{wv}}\right)\right) \text{ where } k = F_{uw} \bigcap F_{wv}.$$

    if $(\frac{k}{2} \cdot (\frac{s_{uz}}{k_{uz}} + \frac{s_{zv}}{k_{zv}}))_w > t$ **then**
      add (u, v) with probability p and discard with probability $1 - p$.
    $i \leftarrow i + 1$
  add random set of nodes V_{new} and a random set of edges connecting to G_i.

We have implemented the iterative evolution process on 100 nodes. We can see in figure 7 the graph on which the iterative evolution process has been applied. In this experiment we have taken the set of nodes V_{new} as 20 from which random number of nodes are taken and are added to the graph at every time step, and the evolution process continues with the modified graph. Figure 7 shows the initial graph to which some random nodes are connected to it. Figure 8 is the graph obtained after evolution when 10 nodes get added random to the graph in 7 at some time step. We can also observe formation of communities in iterative evolution process as shown by the networks evolved through the evolution process with fixed number of nodes.

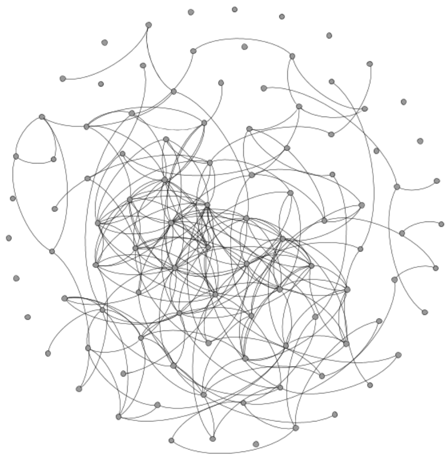

Fig. 7. Iterative Evolution process on 100 nodes

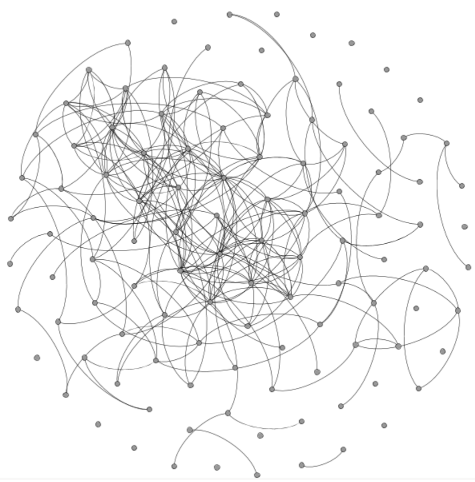

Fig. 8. Evolved graph at time step $T + k$

IV. Conclusion

In this work we propose a model of evolution of social networks based on the transitive property of the growth of friendship relation. Our model is based on predicting the similarity of two people which satisfy the transitive property. We characterize a particular relation i.e. an edge in the social network graph, by a number of sociopsychological *factors* and a score value which measure the strength of the relationship. We observe that if we use a specific formulation to derive the strength of a recommended friendship using arithmetic mean, dense communities are formed which is a characteristic of real life social networks. Changing the nature of the formulation using other means doesn't show formation communities. To the best of our knowledge such a model has not been proposed yet. We suspect that this model can be useful in understanding the evolution of complex social networks like online social networks, political networks, corporate networks etc. As part of future work, it would be interesting to see whether our model can make accurate predictions for real life social networks in terms of formations of communities. It would also be interesting to prove theoretical guarantees on the results obtained in this paper which would in turn be an in depth study of the theoretical underpinnings of this work.

References

[1] L. Backstorm, D. Huttenlocher, J. Kleinberg and X. Lan, "Group formation in large social networks: membership, growth, and evolution," *Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining*, ACM, 2006, pp. 44-54.

[2] M. Berlingerio, F. Bonchi, B. Bringmann and A. Gionis, "Mining Graph Evolution Rules", *Proc. European Conf. Machine Learning and Knowledge Discovery in Databases (ECMLPKDD)*, LNCS 5781, Springer, 2009, pp. 115-130.

[3] J. Coleman, "Foundations of Social Theory", Cambridge: The Belknap Press, 1990.

[4] D.S Goldberg and F.P Roth, "Assessing experimentally derived interactions in a small world," *Proceedings of the National Academy of Sciences*, 2003, pp. 4372-4376.

[5] R.Th.A.J. Leenders, "Structure and Influence Stastical Models for the Dynamics of Actor Attributes, Network Structure and their Independence", Amsterdam: Thesis Publishers,1995

[6] J. Leskovec, L. Backstrom, R. Kumar and A. Tomkins, "Microscopic Evolution of Social Networks," *Proc. 14th ACM SIGKDD Int'l Conf. Knowledge Discovery and Data Mining (KDD)*, ACM Press, 2008, pp. 462-470.

[7] J. Leskovec, J.M. Kleinberg and C. Faloutsos, "Graphs Over Time: Densification Laws, Shrinking Diameters and Possible Explanations," *Proc. 11th ACM SIGKDD Int'l Conf. Knowledge Discovery and Data Mining (KDD)*, ACM Press, 2005, pp. 177-187.

[8] S. Schachter, "The Psychology of Affiliation", Stanford: Stanford University Press,1995

[9] F.N. Stokman and R.V. Oosten, "The Exchange of Voting Positions: An Object-Oriented Model of Policy Networks", *European Community Decision Making Models, Applications and Comparisions*, New Haven:Yale University Press, 1994, pp. 105-127.

[10] R. Toivonen, L. Kovanen, M. Kivela, J.P Onnela, J. Saramaki and K. Kaski "A comparative study of social network models: Network evolution models and nodal attribute models," *Social Networks*, 2009, pp. 240-254.

[11] R. Wippler, "The structural-individualistic approach in Dutch sociology:toward an explanatory social science", *The Netherlands Journal of Sociology 14*, 1978, pp. 135-155.

[12] E.P.H. Zeggelink, "Strangers into friends, The Evolution of Friendship Networks Using An Individual Oriented Modeling Approach", Amsterdam: Thesis Publishers, 1993.

[13] E.P.H. Zeggelink, "Dynamics of Structure: An Individual Oriented Approach", Social Networks 16, 1994, pp. 295-333.