Modeling Dynamics of Social Networks: A Survey

Saoussen Aouay*, Salma Jamoussi*, Faiez Gargouri*, Ajith Abraham[†‡]

* MIRACL, Multimedia InfoRmation system and Advanced Computing Laboratory
Higher Institute of Computer Science and Multimedia, Sfax, BP 1030 Tunisia
aouay.saoussen@gmail.com, jamoussi@gmail.com, faiez.gargouri@isimsf.rnu.tn
[†]Machine Intelligence Research Labs (MIR Labs), Washington 98071, USA
[‡]IT4Innovations - Center of excellence, VSB - Technical University of Ostrava, Czech Republic
ajith.abraham@ieee.org

Abstract—In the last decade, social networks have increasingly attracting the attention of several researchers in various fields. By tradition, social network analysis (SNA) is performed on static graphs but this representation is very limited for a sound, network analysis. Thus, thanks to the availability of large social networks data sets, the interest in modeling how these networks evolve dynamically has increased steadily. Many studies have been oriented toward understanding the behavior and the evolution of network structures over time. This paper does the survey of complex networks models and methods which are proposed to reproduce structural changes of these graphs.

Index Terms—social network, modeling graphs, dynamic structure, evolution, link prediction

I. INTRODUCTION

Complex networks are present in several scientific areas [1][2]. They have been widely investigated due to their importance in different systems such as the World Wide Web, power grids, transportation systems, ecosystems, genetic networks, social networks, biological networks, etc [3].

Social networks illustrate social interactions among people; they occupy an increasingly important role in our daily life. This success is due to the diversity of services platforms of web 2.0 include marking, web discussions, suggested contacts, chat, etc... With the growth in the online social media, social network analysis has attracted the interest of several researchers and it is instead of a wide range of applications ranging from modeling social interactions, community detection through recommendation systems.

With the increasing availability of large collections of data from social networks, the interest shown in the analysis of how these networks change over time has steadily increased. Traditionally, social network analysis is done on a single snapshot of the network, called the analysis of static networks, but in recent years, several researchers have studied the dynamics of networks and they have proposed different models to monitor structural change over time.

This research does the survey of modeling the dynamics and structural evolution of social networks. The remainder of the paper is structured as follows: In the next Section, we describe the principal characteristics of social networks. In Section 3 we present the well-known networks models. Then we provide an overview of existing work in the area of link prediction and modeling dynamics of social network respectively. Finally, the last part includes conclusions and some possible future directions.

II. BASIC PROPERTIES OF SOCIAL NETWORKS

Generally, a social network is modeled by a graph $G = (V, E)$ where V presents the set of nodes and E is the edges between entities. In several studies concerning complex networks, a set of common properties has been observed in order to characterize the behavior of the network. In the following part, definitions of some features are provided which are summarized in Table 1.

The diameter: is the longest of the shortest paths between all possible pairs of nodes through their connections. Different studies have shown that large network of interactions having a small diameter, which is also known as the small-world phenomenon or the six degrees of separation [4].

The density: of a graph measuring the extent to which it is a complete graph. The real networks often have a density very weak. We often observe that the number of links is proportional to the number of nodes.

The Distribution degrees: of a node is known by the number of its direct neighbors. In real networks is often observed that there is some nodes that have very high degrees, and many others have very low degrees. Distribution of degrees is often described by a distribution power law [5].

The Clustering coefficient: quantifies how likely two vertex with a common neighbor may be neighbors themselves. This corresponds to the local density of a vertex. Clustering coefficient of the whole network is the average of all local coefficients [4].

The community structure: have been presented in different studies which have demonstrated that most social networks having a modular structure [6][7]. In such structure, the nodes form communities where vertex in the same groups are very connected to each other than to others out-sided.

III. COMPLEX NETWORK MODELS

Many models have been proposed and studied in order to produce the structures of complex networks and the relationships between their different components. In the following, we will illustrate the four most known types of these models.

49

TABLE I
BASIC MEASURES OF SOCIAL NETWORKS

number of nodes	$n = |V|$
number of edges	$m = |E|$
Degree	$d_i = \sum_{j \in V} a_{ij}$
Density	$\rho = \frac{(2m)}{(n(n-1))}$
Diameter	$D = \max_{ij} d_{ij}$
Local clustering coefficient	$C_i = \frac{2m_i}{n_i(n_i-1)}$ m_i nb of edges connected to node i n_i nb of neighbors of node i
Global clustering coefficient	$C_g = \frac{1}{n} \sum_i C_i$

A. Regular graphs

Regular networks are characterized by a structure in which the connections between vertices follow an usual model. A regular network with vertices of degree k is called a k-regular network. Other examples include regular graphs lattices, rings, mesh, n-ary trees, stars and complete graphs [8]. This type of models has the lowest heterogeneity (the number of connections of each node is more or less the same) and the probability of any two randomly chosen nodes to be wired to each other is very low or zero.

B. Random graphs

Random networks have been studied by Erds and Rnyi (ER) [9] and Gilbert [10]. These graphs are generated by starting with a disconnected set of nodes that are then paired randomly. The Gilberts model is donated by $G_{(n,p)}$ graph in which each edge is included independently with a probability $p \in [0,1]$ from the possible edges. In contrast, ERs model is presented by $G_{(n,m)}$ graph which is chosen randomly from all potentials graphs containing n nodes and m edges. Generally, the average degree of the nodes is calculated $<k> = \frac{n(n-1)p}{n} \approx np$. So, random graphs are characterized by a Poisson node degree distribution and small clustering coefficient. It is provided in literature that random graph is not appropriate for representing real networks.

C. Small-World models

As mentioned above, regular graphs are grouped, but do not show the effect small-world. Moreover, random graphs demonstrate the small world propriety, but do not show clustering. Therefore, these networks models failed to reproduce some important features of many real networks. For this reason, Watts and Strogatz in 1998 [4] have proposed Small-world models for describing social networks. This model can be produced in the following manner: Starting from a $k-$regular ring graph, each edge of the network will be rewiring randomly according to a certain probability $p \in [0,1]$. An alternative of this model was proposed by Newman, Watts [11] in which one can start from a k-regular graph and instead redirecting edges, new ones, are added between pairs of nodes selected randomly. In both models, varying the parameter p may occur different structures, starting from regular graph if $p = 0$ until random graph when $p = 1$ (see figure 1).
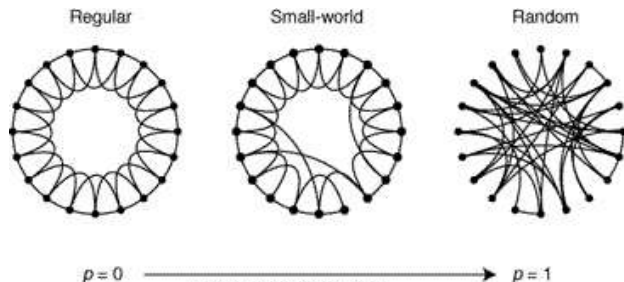

Fig. 1. Network structure from regular to random network

Small networks are very similar structurally to various social networks as meaning that they have a higher clustering coefficient and shortest paths between nodes. However, this model does not reflect other characteristics presented by real networks, such as the aspect of degree distribution.

D. Scale-free networks

Barabasi and Albert have recommended a dynamic mechanism of network growth, so-called scale-free network [12][13]. It is based on the idea that newcomers nodes attach preferentially to the most connected existing nodes. This concept is named as preferential attachment. In practice, BA model is very easy to implement: Originally, the graph has a small number of nodes. In every iteration, new added vertices are linked to a fixed number of existing nodes selected according to a probability proportional to her degree. This process can generate a degree distribution follows a power law distribution. These structures combine heterogeneity and randomness, they can have low or high modularity and many Small world networks are also scale-free. This model has attracted much attention in the, literature. Indeed, many authors have suggested modifications or extensions [14][15] to improve the behavior of BA model and bring it closer to real networks.
The table II summarizes the basic proprieties of the discussed models.

IV. LINK PREDICTION

Link prediction is a fundamental task in social network analysis. It consists estimating the probability of a link appear or disappear in the graph network. The issue has a broad range of applications as recommendation systems, identification of

TABLE II
COMPLEX NETWORKS MODELS

Features	Degree Distribution $P < k >$	Clustering Coefficient CC	Average Path Ln
Regular	Power-Law	High	Long $\frac{N}{2k}$
Random	Poisson $e^{-k}\frac{<k>^k}{k!}$	Low $\frac{<k>}{N}$	Short $\frac{\ln N}{\ln k}$
Small-world	Poisson (similar to random graph)	High $(1-p)^3 \frac{3(k-1)}{2(2k-1)}$	Short $= \frac{\ln Nkp}{k^2 p}$
Scale-free	Power-Law $k^{-\gamma}$	High $\frac{\ln N^2}{N}$	Short $\frac{\ln N}{\ln(\ln N)}$

professional relationship in collaboration networks, structural analysis in microbiology field, etc. Proposed methods in literature can be classified into three categories:

A. Similarity based Methods:

Principle embodied in these approaches is to compute a score connection for each pair (x, y) in the graph. Then all unobserved links are sorted based on their similarity values in descending order. So, ties having best results are selected as predicted links. In general, score functions used are generated from neighborhoods or paths between nodes:

- Neighborhood based features [16]: such as Preferential attachment (PA) feature which calculate the product of nodes degrees, Common neighbors (CN) number which presents the same neighbors of nodes, Jaccards coefficient (JC) which is calculated as the common neighbors divided by the total combined neighbors nodes,...

- Path based features [17]: like Katz feature which sums over the collection of paths and is exponentially damped by length to give the shorter paths more weights, Random Walk (RW) which consider a random walker starting from node, who will iteratively moves to a random neighbor with probability, the Shortest Path Distance (SPD) suggests that the shorter distance between nodes, the higher chance of link formation between them.

- Clustering based features: Some researchers have considered the clustering (sub-graph mining approaches, Hierarchical structure..[18]) for predicting link. In these works, the community information, reveals the existence of a high density of ties within groups of nodes, while a low concentration of links between these communities. These procedure includes improving the quality of the link prediction algorithms by a clustering procedure firstly then the algorithm is applied to the modified sub graph.

B. Maximum Likelihood Methods

These methods model the organization of links within the network and define the possible structures of the given graph. Then it calculates the distribution of the parameters by by maximizing the likelihood of each structure.Therefore the likelihood of a missing link can be calculated on the basis of founded model. Generally, there are two most known maximum likelihood approaches which are Hierarchical Structure Model and Stochastic Block Model [19].

C. Probabilistic models

These methods make the abstraction of the underlying structure of the network, based on observed data in a compact probabilistic model. The learned model consists of a group of parameters which may better adjust to the observed data. Then, unobserved links can be predicted via a the probability estimated conditionally to these parameters. There are three probabilistic models which are probabilistic relational models (PRM Probabilistic Relational model (PRM) [20], Stochastic relational models (SMR) [21] and Probabilistic Entity Relationship Model (PERM) [22].

V. DYNAMICS OF SOCIAL NETWORKS

Various link prediction approaches do not take into account the temporal aspects of the network. They treated link prediction from a statistical graphs point of view. This representation does not permit us to explore the dynamic properties of data. So to overcome this drawback, several recent studies use the time information (such as time of establishing the link, the duration of the interaction, temporal sub-graphs, etc.) for networks temporal analysis [23][24]. In fact modeling the dynamic of networks is regarded as one of the most interesting subjects in socials networks analysis. It allows us to understand the evolution of network structures and the behavior of its users over time.

In the following section we will present some approaches that have been proposed to describe the structural changes in social networks. We classify these methods into three categories: methods using nodes attributes information (such as centrality, attachment preferential, average distance ...) , methods that extract different patterns in the graph and model theirs structural changes and finally methods with bio-inspirations sources that try to simulate the evolution of the network over time.

A. Node based methods

Kumar and al [25] studied the growth of the structure of online social networks (Flickr and Yahoo 360!). This task required, first of all, a categorization of actors into three different types: singletons, most connected (giant component) and isolated communities. A set of observations was performed over time by recording the exact moments when each node and edge has become a part of the graph. They found that the average distance between users decreases over time, and that these networks exhibit power-law degree distributions. Also they showed that a large part of the network is presented in the form of stars. This pattern has

been explained by the dynamics of invitations. Based on these results a simple model was proposed. There are three types of users described in this model such as passive, linkers, and inviters. The model process is composed of two steps: At each time, a new node arrives, is randomly designated as passive, linker, or inviter actor. Then, in the same time, a set of edges arrive. Thus, the source of each edge is selected from existing linkers and inviters via the preferential attachment. Then, if the source is an inviting, it invites a non-member to join the network and the destination is a new node. Otherwise, the destination is selected from existing linkers and inviters, also by using preferential attachment. This model takes into account the addition of new nodes and edges and not fading existing relationships.

Leskovec and al analyzed in [26], at a more local level, the evolution of four online social networks: Flickr, Delicious, LinkedIn and Answers. They observed the power law degree distribution, growth of density over time and the decrease of average distance between nodes. They demonstrate also that edge locality has an important effect in the evolution of networks. They utilized the method of maximum likelihood to compare models according to their probability of generating the observed data. They have developed a model of network evolution, in which the penetration of new nodes in the graph is expressed by an exponential or polynomial function $N(t)$. Each node has a restricted lifetime. Each node will launch edges according to a time interval between creating an edge and another. The choice of a destination for each edge is made according to a simple triangle closure. Thus, a source node chooses an intermediate node evenly among its neighbors then the intermediate node does the same. This method also does not consider the fact that an edge can disappear from the network.

In [27] the authors focused their attention on the role of the centrality of nodes in the evolution of the structure over time . Their experiments have studied a large network of email, by sampling in sub graphs with data interval of one day. They noted that all networks are weakly correlated which is explained by the low probability of a link appears in the present, to reappear later. Therefore they tried to characterize the centrality of nodes every day, which is called dynamic centrality. They noted that the daily networks were scale-free and the degree of a node varies considerably over time and following a power law distribution. Subsequently, a dynamic scale-free model is proposed (DSF). DSF can be presented as the result of a series of vertices weighted random walks on scale-free network, accompanied by a dynamic preferential attachment rule. Starting with the global network, each daily network will be built by a random walk node reinforced.

B. Frequent Pattern based methods

Bringmann and al [28] presented an alternative approach that defines a set of rules to model the structural change in the network. The proposed to represent the data-set by collapsing all graphs into a single one undirected supplemented by the time-stamp added to every relationship when it appears in the network for the first time. Then, they tried to discover association rules between frequent patterns of interaction (sub-graphs) to illustrate the evolution of the network. So, the GERM (Graph Evolution Rule Miner) is a solution to extract the evolution rules of the graph, and it was tested on four real-world networks (Flickr, Y360, DBLP, arXiv) via differents time granularities. It allows predicting the beginning of a new relationship connected to a new node and also providing new edges between existing nodes. They dont allow following the dynamics of relationships between important users.

In [29] the authors characterize the behavior of network by examining the relationship between possible existing triads. They have showed that networks are highly dynamic in short timescales and the density of real-world networks is small and the graphs are sparse while the number of links grows only linearly. Intended for describing evolutionary patterns of the network, a probabilistic model was derived in order to calculate the probabilities of transitions among triads of nodes. So, transition triad matrix TTM was used to present the probability that, a given pattern (i) at time t will be transferred into another (j) in the future. The average values of TTM have been utilized to predicting links. Promising results have been achieved by this approach, but mining frequent sub-graphs is a very expensive task.

C. Bio-inspired methods

Authors in [30] have used the forgetting curve and swarm intelligence to model the dynamics of a coauthor network. They analyzed two types of links between nodes: social and professional ties. The first type of links is modeled using the curve of forgetting that defines the probability that a person can recall the information from the previous recall. Indeed, this can tell about the stability of relations between authors over time. Professional relations were modeled via inspired social ants behavior. Each joint publication of two authors causes an increase in the amount of pheromone in the link and strengthens the path connecting between these two authors. Experimentation were developed on a set of cooperation between the authors of DBLP network. The results present the social and professional network of two different network members.

Budka and al [31] work is based on a molecular inspired socio-dynamical model to model the evolution of an email-based social network. It takes into account creation and disappearance of the relationships. Additionally, it studied the changes of relationships strength over the time. The framework proposed is resumed in the following points: (1) Extraction of the interactions and time stamps from data-set (2) Creation of a distance matrix for each social network snapshot

(3) Embedding distance matrices into Euclidean space (4) Definition of molecular model which is based on the force potential function (determined from the changes in distances between two consecutive network snapshots). (5) Simulation of molecular model via the embedded social network windows and the potential force. (6) Recreation of network snapshots from the results of simulation according to a given threshold. Both properties have been considered during experiments, node degree distribution and clustering coefficient.

The given table III summarizes the studied approaches.

VI. CONCLUSION

Social networks have achieved a high success over the years and have become a very important way for communication between people. They have been attracted the interest of several analysis researches.This paper presents a brief study of modeling dynamic of social network. We have presented the most known models proposed to model the structure of social networks and their proprieties. Several studies have focused in link prediction and dynamics of socials networks have been illustrated. Modeling the evolution of social networks remains a very interesting topic and a difficult task. In fact, biology presents an inexhaustible inspiration source, which has more and more success in artificial intelligence. Proposing a framework based a bio-inspired paradigm for the underling problem is one of the future direction to be emphasized.

REFERENCES

[1] S. H. Strogatz, "Exploring complex networks," *Nature*, vol. 410, pp. 268–276, 2001.
[2] X. F. Wang, "Complex networks: Topology, dynamics and synchronization," *I. J. Bifurcation and Chaos*, vol. 12, no. 5, pp. 885–916, 2002.
[3] A.-L. Barabasi, *Linked: How Everything Is Connected to Everything Else and What It Means for Business, Science, and Everyday Life*. Plume Books, April 2003.
[4] D. Watts and S. Strogatz, "Collective dynamics of 'small-world' networks," *Nature*, no. 393, pp. 440–442, 1998.
[5] M. Newman, A. L. Barabsi, and D. J. Watts, *The Structure and Dynamics of Networks*. Princeton University Press, 2006.
[6] G. W. Flake, S. Lawrence, C. L. Giles, and F. M. Coetzee, "Self-organization and identification of web communities," *IEEE Computer*, vol. 35, pp. 66–71, 2002.
[7] M. Girvan and M. E. J. Newman, "Community structure in social and biological networks," *PNAS*, vol. 99, no. 12, pp. 7821–7826, June 2002.
[8] T. G. Lewis, S. Pickl, B. Peek, and G. Xu, "Network science [guest editorial]," *IEEE Network*, vol. 24, no. 6, pp. 4–5, 2010.
[9] P. Erdös and A. Rényi, "On random graphs i," *Publicationes Mathematicae Debrecen*, vol. 6, p. 290, 1959.
[10] E. N. Gilbert, "Random graphs," *Annals of Mathematical Statistics*, vol. 30, no. 4, pp. 1141–1144, 1959.
[11] L. Adamic, "Collective dynamics of 'small-world' networks," *Lecture Notes in Computer Science*, no. 166, pp. 443–452, 1999.
[12] M. Newman and D. Watts, "Scaling and percolation in the small-world network model," *Phys. Rev. E*, vol. 60, no. 6, pp. 7332–7342, Dec. 1999.
[13] A. L. Barabasi and R. Albert, "Emergence of scaling in random networks," *Science*, vol. 286, pp. 509–512, 1999.
[14] E. Jin, M. Girvan, and M. Newman, "Structure of growing social networks," *Phys. Rev. E*, vol. 64, no. 4, p. 046132, Sep. 2001.
[15] H. Jeong, Z. Nda, and A.-L. Barabsi, "Measuring preferential attachment in evolving networks," *Europhysics Letters*, vol. 61, no. 4, pp. 567–572, 2003.
[16] D. Liben-Nowell and J. Kleinberg, "The link-prediction problem for social networks," *J. Am. Soc. Inf. Sci. Technol.*, vol. 58, no. 7, pp. 1019–1031, May 2007. [Online]. Available: http://dx.doi.org/10.1002/asi.v58:7

[17] M. A. Hasan and M. J. Zaki, "A survey of link prediction in social networks," in *Social Network Data Analytics*, 2011, pp. 243–275.
[18] E. Hoseini, S. Hashemi, and A. Hamzeh, "Link prediction in social network using co-clustering based approach." in *AINA Workshops*, L. Barolli, T. Enokido, F. Xhafa, and M. Takizawa, Eds. IEEE, 2012, pp. 795–800.
[19] L. L and T. Zhou, "Link prediction in complex networks: A survey," *Physica A*, no. 6, p. 11501170.
[20] D. K. D. Heckerman, C. Meek, "Probabilistic models for relational data," in *Tech. Rep. MSR-TR-2004-30*, 2004.
[21] K. Yu, W. Chu, S. Yu, V. Tresp, and Z. Xu, "Stochastic relational models for discriminative link prediction," in *Advances in Neural Information Processing Systems*. MIT Press, 2007, pp. 333–340.
[22] D. Heckerman and C. Meek, "Probabilistic entity-relationship models, prms, and plate models," in *In Proceedings of the ICML-2004 Workshop on Statistical Relational Learning and its Connections to Other Fields*, 2004, pp. 55–60.
[23] T. Tylenda, R. Angelova, and S. Bedathur, "Towards time-aware link prediction in evolving social networks," in *Proceedings of the 3rd Workshop on Social Network Mining and Analysis*, ser. SNA-KDD '09. New York, NY, USA: ACM, 2009, pp. 9:1–9:10.
[24] Z. Huang and D. K. J. Lin, "The time-series link prediction problem with applications in communication surveillance." *INFORMS Journal on Computing*, vol. 21, no. 2, pp. 286–303, 2009.
[25] R. Kumar, J. Novak, and A. Tomkins, "Structure and evolution of online social networks," in *KDD '06: Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining*. New York, NY, USA: ACM, 2006, pp. 611–617.
[26] J. Leskovec, L. Backstrom, R. Kumar, and A. Tomkins, "Microscopic evolution of social networks." in *KDD*, Y. Li, B. L. 0001, and S. Sarawagi, Eds. ACM, 2008, pp. 462–470.
[27] S. A. Hill and D. Braha, "Dynamic model of time-dependent complex networks," *Phys. Rev. E*, vol. 82, no. 4, pp. 046 105+, 2010.
[28] B. Bringmann, M. Berlingerio, F. Bonchi, and A. Gionis, "Learning and predicting the evolution of social networks." *IEEE Intelligent Systems*, vol. 25, no. 4, pp. 26–35, 2010.
[29] K. Juszczyszyn, M. Budka, and K. Musial, "The dynamic structural patterns of social networks based on triad transitions." in *ASONAM*. IEEE Computer Society, 2011, pp. 581–586.
[30] M. Kudelka, Z. Horak, V. Snsel, P. Krmer, J. Platos, and A. Abraham, "Social and swarm aspects of co-authorship network." *Logic Journal of the IGPL*, vol. 20, no. 3, pp. 634–643, 2012.
[31] M. Budka, K. Juszczyszyn, K. Musial, and A. Musial, "Molecular model of dynamic social network based on e-mail communication." vol. 3, no. 3, pp. 543–563, 2013.

Related Works	Data Set	Basic Properties	Observations	Model Process	Applied Techniques
[25]	Yahoo Flicker	-Density -Diameter -Component Structure	-Most component are characterized by stars	Categorization of users as passive, linkers and inviters.	-Preferential attachment
[26]	Flickr Delicious Yahoo LinkedIn	-Node age -Node Degree -Time gap between edges	-Most new edges have very short distances, typically closing triangles.	-Node arrivals -Edge initiation -Edge destination selection	-Maximum likelihood - Preferential attachment
[27]	Email network	-Nodes Centrality - Degree distribution	-Node takes into account the popularity of its neighbors to initiate new edge	Generating a series of dynamic scale free networks from the underlay network	-Vertex reinforced random walk -Preferential attachment
[28]	Flickr Yahoo DBLP arXiv	- Average degree - Growth rate - Number of frequent pattern	-Mining for such Local patterns can provide the evolving network data.	- Extracting relative time patterns - Defining Graph Evolution Rules	- Association rules - Frequent pattern mining
[29]	E-mail Network	-Triade In degree -Triade Out degree -Triade Role	-Different patterns are responsible of evolution of connections between nodes.	Deriving the probabilities of transitions between triads	-Triad Transition Matrix -Frequent pattern mining
[30]	DBLP	-Vertex stability -Vertex activity -Link stability -Link activity	- Social ties are long-term relations - Professional relations are based on common interests - Professional relations change faster than social ones.	- Calculating the retention and stability of every author and tie - Increasing the quantity of pheromones for every co-author nodes and their link.	- Forgetting curve -Ant Colony Optimization
[31]	Email Network	-Distance change -Degree distribution -Clustering coefficient	-Dynamic strength of the relationships is an important factor in social networks	-Distance Matrix Creation -Embedding procedure -Molecular Model -Reconstruction of Social Networks	-Molecular Paradigm