Utilizing Dynamic Molecular Modelling Technique for Predicting Changes in Complex Social Networks

Krzysztof Juszczyszyn[1], Anna Musiał[2], Katarzyna Musial[3-], Piotr Bródka[1]
[1] *Institute of Informatics, Wrocław University of Technology, Poland*
[2] *Institute of Physics, Wroclaw University of Technology, Poland*
[3] *School of Design, Engineering & Computing, Bournemouth University, United Kingdom*
krzysztof@pwr.wroc.pl, anna.musial@pwr.wroc.pl, kmusial@bournemouth.ac.uk, piotr.brodka@pwr.wroc.pl,

Abstract

We present a method that utilises dynamic molecular modelling technique in order to track the changes within complex social network. The users forming a social network are interpreted as large sets of interacting particles. The data for the conducted research was obtained from e-mail communication within Enron company. Based on this information the social network of employees was extracted and then used to evaluate the methodology of social network dynamics modelling proposed by authors.

1. Introduction

There are many types of complex networked systems (CNS). One of the classifications distinguishes infrastructural (Internet, WWW, energy and transportation networks) and natural complex systems (biological networks, social systems and ecosystems) [20]. Complex structure consists of multiple interacting components whose global behaviour cannot be simply inferred from the behaviour of the components [2][20]. Such networks consist of thousand and millions highly interconnected entities that influence one another [10].

Due to the scale and complexity of such systems, computer simulations became increasingly popular and useful tool utilized in analysis of such systems. Simulations supplement the traditional approaches – formal theories and empirical studies, as well as they serve as analytical models. Simulations seem to be the only way to insight into global system dynamics [2][3] as the CNS and their features are hardly definable in terms of any analytical model. Physics has provided several methodological approaches to tackle this issue. One of the approaches used in tracking system dynamics are cellular automata [5], starting from famous Game of Life artificial life model of Conway. Dirk Helbing was one of the first researchers who applied the particle-based approach to social dynamics. In [18] he created a dynamic model that includes many established models as special cases, the gravity model, some diffusion models, the evolutionary game theory and the social field theory, but it also implies numerous new results. Spatial mobility and concentration of interacting particles may be modelled with molecular dynamics paradigm. It allows many applicable extensions based on the reinterpretation of potentials and distance in the given space [4]. The proposition of such modification is described in this article and is utilised to track changes in complex social networks with the experiments performed on network derived from the data about email communication. The first attempt to this problem we made by authors in [17]. The rest of the paper is structured as follows: in section 2 we present the idea of dynamic molecular modelling and simulation. In section 3 we present our approach in details and finally, in section 4 the description of experiments and discussion is provided.

2. Dynamic Molecular Modelling and Simulation

Dynamic molecular modelling is one of simulation methods. Whenever we are interested in a dynamic of physical system that contains many interacting objects we are dealing with a many-body problem.

There are some simulation parameters that need to be defined: the value and number of time steps. The interaction between particles should be also described. The interaction is usually given by a two-particle potential (meaning that each two particles are interacting), which depends on the distance between the particles. Our approach assumes that the potential

reflects the tendency of two particles (nodes) to change their distance in social space. The potential is commonly known as an analytical formula obtained from either theory or experiment, e.g. Lenard-Jones's potential. Based on the potential, the force acting between particles can be calculated using the following formula: $\vec{F} = -\nabla V$, where V is the interaction potential. If the force is known, the time evolution of the system can be obtained by solving for each particle separately the classical equation of motion. For objects with constant mass the formula describing the 2nd Newton's principle of dynamics takes the form: $\vec{F} = m\vec{a} = m\frac{\overrightarrow{d^2 r}}{dt^2}$ by using the definition of the momentum and acceleration a (which is defined as a second derivative of the position vector r). During the simulation above equation must be solved for each particle in each simulation step.

It is possible to calculate the positions of all particles after a given time using the Verlet's algorithm [21]. Input data indispensable for starting the algorithm are: initial positions of all particles, number of particles, formula for interaction force between each two particles. The essential part of the procedure, is the formula used to describe interactions between the particles. This force is identical for all two interacting particles within the system. To model the described system the Lennard-Jones potential is utilized:

$$V(r) = 4\varepsilon\left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6}\right] \qquad (1)$$

where: r is the distance between particles. The character of the potential is shown below (Fig. 1).

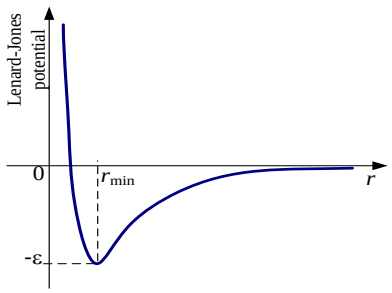

Fig. 1 Lennard-Jones potential.

Above potential is fully described by two parameters: the depth of the potential – ε responsible for the strength of interactions between particles and σ – related to the minimum distance between to particles. As it can be seen the potential has a global minimum equal ε for $r_{min} = 2^{1/6}\sigma = 1.12\sigma$. Other characteristic features of this potential are: (i) when the distance between the particles grows to infinity the value of the potential aims zero asymptotically, (ii) for smaller distances the particles are attracting each other. For distances smaller then σ the character of the potential changes to strongly repulsive. To obtain the formula for the force between two particles the partial derivative of potential should be taken.

The next step is to transform the forces into dimensionless form. This way the magnitude of used values, normally extremely small, are better represented in a computer and the operations on them can be more precise. To do this the distances are introduced in [σ] units and time in $\left[\sqrt{\frac{m\sigma^2}{48\varepsilon}}\right]$, where m is the mass of interacting particles. As a first approximation the masses of all particles are equal to 1. Knowing the forces, the Verlet's algorithm may be used to obtain the position and the velocity of each particle in following time steps, as described in [17]. To compare the experimental data with model, the positions of the particles should be generated and the value of the parameters of the Lennard-Jones potential should be tuned.

3. Description of Experiments

Application of the molecular dynamics to complex social network modelling requires completing several tasks, the overall methodology assumes four main stages: (i) Data preparation – extracting social network and creating temporal images of social network, (ii) Embedding social network graph in 2D space, (iii) Estimating the character of social force potential between network nodes, (iv) Using dynamic molecular modelling to predict the evolution of the network.

3.1. Data Preparation

The social network that has been chosen for the experiments was extracted from the data about Enron email logs (http://www.cs.cmu.edu/~enron/). The Enron dataset consists of the employees' mail boxes. The data used in our experiment contains information about email exchange between 151 users from the period of 48 months (01.01.1999 – 31.12.2002). The relationship from user x to y exists if there was at least one email sent from x to y.

First, the data has to be cleansed by the removal of bad email addresses and unification of duplicates. The

strength of relationship S from user x to y is calculated by dividing the number of emails sent from x to y by the number of emails sent by x in total. Note that every email with more than one recipient was treated as $1/n$ of a regular email, where n is the number of its recipients. The strength of the relationship between two users can be seen as the distance between two particles – the greater the strength the closer the users (particles) to each other. The resulting social network $SN = \langle N, S \rangle$ is defined as a tuple consisting of the set of network nodes N and a set of weighted relationships $S : N \times N \rightarrow [0,1]$.

To prepare the data for simulation we divided it into 74 time windows using the approach of sliding time frame. The whole data were divided into frames covering 60 days each, and starting in consecutive moments which differ by 20 days. For each created time window the strength of relationships were calculated separately and as a result 74 social networks that can be seen as temporal images of evolving social structure were built. From now on we denote to them as: $SN(t_0)$, $SN(t_1),\ldots,SN(t_n)$, e.g. $SN(t_n)$ is a social network image in time t_n, where t_0 is the first time period. We have also used this approach in [17] where the emails from our university server logs were investigated.

3.2. Embedding networks in Euclidean space

Following the in-depth discussion presented in [6] we cannot expect the social space to be metric i.e. the triangle inequality between any three nodes may not be fulfilled. On the other hand, molecular modelling assumes the interaction between the particles moving in the Euclidean space. For this reason, in order to apply the molecular modelling we must conduct *embedding* of the social network graph in metric, Euclidean 2D space.

In this step we embed the social network graphs SN (created time windows) in Euclidean 2D space, where each node is represented by a point with given coordinates. The resulting sets of points SN_0, $SN_1,\ldots,SN_n$ represent the temporal network images. We obtain the representation of social system in which the network is seen as an assembly of N particles (the nodes of social network). Numerous methods for embedding high-dimensional structures into 2D ones were developed so far, e.g. Big-Bang embedding algorithm [8] or multidimensional scaling (MDS) algorithm [13][14]. In order to embed the social network graph in Euclidean 2D space the Minimum Volume Embedding (MVE) procedure was used. MVE [18][19] is an algorithm for non-linear dimensionality reduction that uses semi-definite programming (SDP) and matrix factorization to find a low-dimensional embedding that preserves local distances between points while representing the dataset in many fewer dimensions. Authors of MVE emphasise that in all cases MVE in comparison with Semi-definite Embedding and Kernel Principal Component Analysis is able to capture more of the variance of the data in the first two eigenvectors, providing a more accurate 2-dimensional embedding [18][19]. The main features of minimum volume embedding approach are:

- MVE used for a given dataset returns always the same set of coordinates.
- Isolated nodes are neglected in the embedding process.
- MVE is stable, i.e. adding one node with a very weak connections does not influence significantly the positions of the rest nodes.

3.3. Setting up the dynamic molecular model

Because the sets of network nodes in $SN(t_0)$, $SN(t_1),\ldots SN(t_n)$ are equal, each point (node) is represented in any of the sets SN_0, $SN_1,\ldots,SN_n$.

At this point we use the formalism of molecular dynamics to associate a force potential V with any particle (network node). The actual characteristic of this potential depends on the behaviour of the particles changing their positions in time instants $t_0, t_1, \ldots, t_n$. First experiments were carried on our university emails using standard Lennard-Jones potential function and the outcomes were presented in [17].

The major changes that we have introduced in the experiments performed in this research are:

1. We have changed the embedding procedure (as discussed in sec. 3.2).

2. We do not exclude users without activities in a given time frame – this allows to dynamically change the number of users in the model which reflects the real situation in internet-based social networks.

3. The force potential function was tuned to conform with the dynamics of change of relationship strength in the network and its shape was confirmed with the statistical analysis of the data. The following parameters of the model were (see Eq.1):

- n and m exponents (in Eq.1 we see n=12, m=6) of the potential function. It was estimated that the character of social interactions in the network requires relatively smaller values: n=1.80 and m=1.75 were used in our experiments.
- ε value, which defines the minimum value of the

potential function (ε=0.5 in our case).

- σ - the location of the minimum (σ=0.16), responsible for the point of equilibrium between attracting and repelling forces in our model. It was estimated on the basis of dynamics relationship strength analysis for all users (network nodes) in our model.

The force potentials associated with the nodes reflect their abilities and tendency to establish future connections with the neighbours – the nodes which are close in terms of social space.

4. Experiments and Discussion

The experiments were conducted in order to track the temporal changes in the network structure. First of all the analysis of our dataset had shown that there is a general relationship between the number of messages exchanged in a given time window, and the change in communication during the consecutive time window. Bars on the Fig.2 represent the measured changes in the number of messages related to the former communication intensity. Aggregated data are shown for all communicating user pairs and the entire period covered by our analysis (48 months). We can see that large number of exchanged messages typically results in negative change – the users do not sustain intense message exchange for a long time. In terms of our model the decreasing number of messages is related to decreasing social distance. This effect confirms the shape of the force potential assumed for experiments (Fig.1) which results in the emergence of repelling force when the objects are close to each other. From the other hand – weak relationship strength (i.e. small number of messages sent) may be associated in the increased intensity communication in the future.

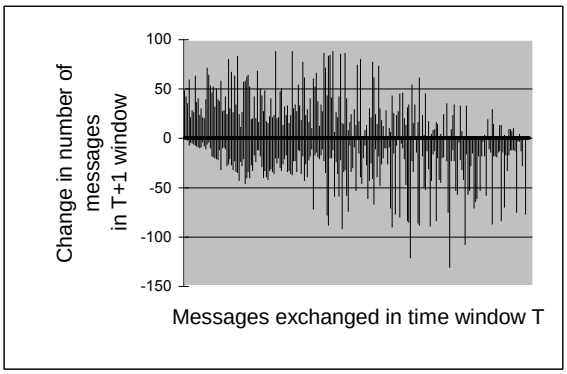

Fig. 2 The change in the number of echanged messages.

For the verification of results the MVE embedding was done for all time widows under consideration. The molecular simulation was run assuming the initial positions of particles taken from the results of MVE embedding of the first time window.

The verification of results was based on the comparison of the distances between all pairs of particles (network nodes) measured for their positions in 2D MVE embeddings and the results returned by the molecular model.

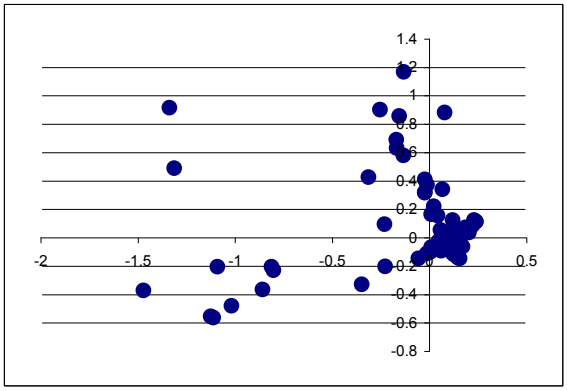

Fig. 3 MVE embedding social graph of 151 nodes in 2D space for time window no. 25.

In both cases distances between any given node and all its counterparts in the network were computed for all embeddings of social network and the positions in the 2D space of molecular model after the equal number of time steps. These vectors of distances were compared using the Pearson correlation coefficient. The average results for the members of the all considered classes are as follows:*t=0* (initial positions): **1.00** *t=10 :* **0.71**, *t=20 :* **0.62**, *t=25 :* **0.47.**

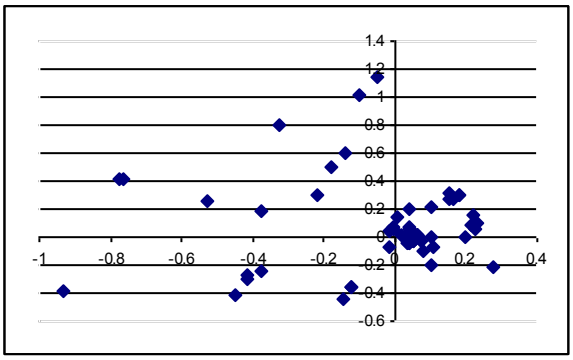

Fig. 4 Position of 151 particles after 25 simulation steps

Figs. 3 and 4 show the positions of network nodes for MVE embedding for the 25th time window and the predictions returned by the molecular model. We may see that the accuracy of predictions decrease with time; this effect may be explained by the obvious individual differences between users (network nodes) while the model assumes the same potential function for all nodes. We may also note, that in some cases (about

10% of the nodes) the model's predictions significantly differ their positions based or real data. These nodes are isolated or loosely connected ones, for which the model is not so accurate – this issue was observed also in first experiments described in [17].

5. Conclusions and future work

The above results form the basis of further experiments which are to create a general framework for modelling the evolution of large internet-based complex networks. Further experiments will include:

- Testing procedures for automatic estimation of model and force potential function parameters.
- Checking correlations between groups found in social network and the clusters of nodes in molecular simulation space.
- Addressing individual properties of the users (different potential fields are being considered).
- Checking the model for different classes of internet-based social networks.

We believe that our method has a good potential and – once its feasibility is confirmed for different social networks – may be applied for large networks, which are hardly researchable with classical approaches. compatible with known solutions from complex systems field.

References

[1] Holland J. Hidden Order: How Adaptation Builds Complexity, Helix Books, USA, 1995.

[2] Schweitzer F. Brownian Agents and Active Particles – Collective Dynamics in the Natural and Social Sciences, Sporinger Series in Synergetics, New York, 2007.

[3] Bocalettia S., et.al. Complex networks: Structure and dynamics, Physics Reports 424 (2006) 175 – 308.

[4] Weidlich W. Physics and Social Science – The Approach of Synergetics, Physical Reports, (204), 1-163, 1991.

[5] Wolfram S. Theory and Application of Cellular Automata, World Scientific, Singapore, 1986.

[6] Watts D. Small Worlds: Dynamic of Networks between Order and Randomness, Princeton University Press, 2002.

[7] Bays C. Candidates for the game of life in three dimensions. Complex Systems, 1(2):373-400, 1987.

[8] Shavitt Y., Tankel T., Big-Bang Simulation for Embedding Network Distances in Euclidean Space, IEEE/ACM Transactions on Networking, 12(6), 2004.

[9] Dangalchev C. Generation Models for Scale-free Networks, Physica A 338 (2004) 659 – 671.

[10] Watts D., Strogatz S., Collective Dynamics of Small World Networks, Nature, 393, 440-442.

[11] Carrington P., Scott J., Wasserman S. (eds.) Models and Methods in Social Networks Analysis, Cambridge University Press, 2005.

[12] J.B. Kruskal, Multidimensional scaling by optimizing goodness of-fit to a non-metric hypothesis, Psychometrika, 29, str. 1-27, 1964.

[13] Kruskal, J. B., and Wish, M., Multidimensional Scaling, Sage University Paper series on Quantitative Application in the Social Sciences, 07-011. Beverly Hills and London: Sage Publications, 1978.

[14] Bronstein, A. M, Bronstein, M.M, and Kimmel, R., Generalized multidimensional scaling: a framework for isometry-invariant partial surface matching, Proc. National Academy of Sciences (PNAS), Vol. 103/5, 2006, pp. 1168-1172.

[15] Berger-Wolf T. Y., Saia J.: A Framework for Analysis of Dynamic Social Networks, Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining, 2006, Philadelphia, PA, USA.

[16] Helbing D., Quantitative Sociodynamics: Stochastic Methods and Models of Social Interaction Processes, Springer, Theory and Decision Library Series B: Mathematical and Statistical Methods, Vol. 31.

[17] Juszczyszyn K., Musial A., Musial K., Brodka P.: Molecular Dynamics Modelling of the Temporal Changes in Complex Networks, IEEE Congress on Evolutionary Computing, Trondheim, Sweden, IEEE Computer Society Press, pp. 553-559, 2009.

[18] B. Shaw and T. Jebara. "Minimum Volume Embedding", AISTATS 2007, Proceedings of the Eleventh International Conference on Artificial Intelligence and Statistics, 2007.

[19] B. Shaw and T. Jebara. "Structure Preserving Embedding", ICML 2009, Proceedings of the 26th International Conference on Machine Learning, Montreal, Canada, 2009.

[20] Barrat A., Barthelemy M., Vespignani A. Dynamical Processes on Complex Networks, Cambridge University Press, UK, 2008.

[21] Verlet L., Computer "Experiments" on Classical Fluids. I. Thermodynamical Properties of Lennard-Jones Molecules, Phys. Rev. **159**, 98 (1967).