AN ARTIFICIAL MOLECULAR MODEL TO FOSTER COMMUNITIES

Christoph Schommer
Department of Computer Science, University Luxembourg, 6 Coudenhove-Kalergi, 1359 Luxembourg, Luxembourg

Keywords: Bio-inspired modeling, Graph mining, Bibliographic communities.

Abstract: This paper introduces in extracts a bio-inspired model that understands graphs as artificial chemical constructs. The main objective is to identify this model as an autonomous and adaptive system that performs internal tasks, for example a communication with its environment. The model itself focus on artificial atomicity of nodes, artificial molecular connections in between, and functional proteins, which are self-concentrated constructs. The model implicates a solid fundament, but fosters an artificial vitality through catalysts: these merge attacked atomic nodes – in case of common "interests" (inside the molecular model) – to functional proteins and therefore consequently contribute to a vivid shape of communities. As an application example, the theoretical model is clarified with bibliographic entries to form bibliographic communities dynamically while having a bibliographic stream entries as input.

1 A SHAPE OF MOLECULES

Given an *actor node*, which represents a node within a network, then we call this node *atomic* (α_{A_i}) in a sense that it is not divisible. Additionally, each node α_{A_i} shares an activation σ_{A_i} and owns a set of items Γ_{A_i} (which will be described later on). An association between two actor nodes A_i and A_j is then represented by a *molecular bond*, which describes a directed *relationship* on a lower level. Each set of items Γ_{A_i} contains atomic entities, which exist in form of a hierarchy.

$$\textbf{sBond}(\alpha_{A_i}, \alpha_{A_j}) = \begin{cases} \omega_{A_i \to A_j}: & \alpha_{A_j} \to \alpha_{A_i} \\ & \vee \\ & \alpha_{A_i} \to \alpha_{A_j} \\ 0: & else \end{cases} \quad (1)$$

A single molecular bond has a weight of $\omega_{A_i \to A_j}$, which is expressed by the conditional probability P(A_j—A_i). If the relationship between the two actor nodes is directed in both directions, then we call this bi-relationship a *double molecular bond*:

$$\textbf{dBond}(\alpha_{A_i}, \alpha_{A_j}) = \begin{cases} \omega_{A_i \to A_j} \times \omega_{A_j \to A_i}: & \alpha_{A_j} \to \alpha_{A_i} \\ & \wedge \\ & \alpha_{A_i} \to \alpha_{A_j} \\ 0: & else \end{cases} \quad (2)$$

The relationship is then characterized by the common connection weight ω_{A_i,A_j}, which is the product of the individual values. Both *atomic actor nodes* share an activation as well (σ_{A_i}, σ_{A_j}). With respect to this, both structures then own a *molecular structure* in between, meaning that any combination of an atomic structure results in a *molecule*. Such a *molecule* might be of different granularity and size, and being expressive in respect to its arity. As presented in Figure 1, a single molecular bonds (left) and a double molecular bonds (right) is shown (each consisting of two atomic actor nucleus).

Figure 1: Single molecular bonds (top) and double molecular bonds (bottom) between two atomic actor nucleus.

With Γ_{A_i}, we allow each actor to own a number of *items*, for example interests that are organized within a hierarchical system. We then receive levels of different granularity with for example $\gamma^0 = \{Root\}$, $\gamma^1 = \{A, \ldots, Z\}$, $\gamma^2 = \{A_1, \ldots, Z_n\}$, and $\gamma^3 = \{A_1, \ldots, Z_{nm}\}$, etc. The concept is that each time a association is performed, the actor's interest $\gamma^j_{A_i}$ may be extended by another interest. The crucial idea is that interests may be substituted by the superordinate hierarchy $\gamma^{j-1}_{A_i}$ in case that a minimum number of interests (= min^k_γ) on the specified level exists:

Schommer C. (2009).
AN ARTIFICIAL MOLECULAR MODEL TO FOSTER COMMUNITIES.
In *Proceedings of the International Conference on Knowledge Discovery and Information Retrieval*, pages 219-222
DOI: 10.5220/0002326902190222

$$\gamma^{j}_{A_i} \rightarrow \gamma^{j-1}_{A_i} \tag{3}$$

if $|(\gamma^{j-1}_{A_i}| \geq min^{k}_{\gamma}$ for an actor node A_i. For example, if an actor A_i is interested in $\gamma^{j}_{A_i} = \{A_{11}, A_{12}, A_{13}, A_{14}\}$, then the interest level may be replaced to $\gamma^{j-1}_{A_i} = \{A_1\}$ in case that the threshold min^{k}_{γ} is achieved.

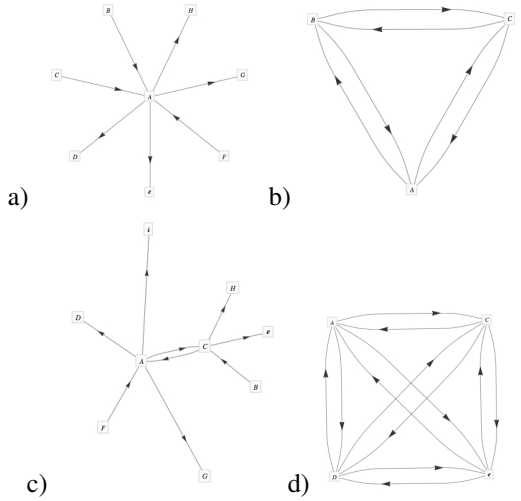

Figure 2: Selected molecular forms with a) a molecular star, b) a collection of molecular bridges (diamond), c) a molecular bridge with single/double molecular bonds (bottom), and d) a collection of diamonds.

Some examples of molecule structures are shown in Figure 2. In Figure 2a), an atomic node α_{A_i} is shown, which is being arranged as a centre of k adjacent nodes. We call this unary structure a *molecule star*, because a certain number of actor nodes are exclusively connected by a single, but centric actor node:

$$\mathbf{mStar}(\alpha_{A_i}) = \begin{cases} 1: & \nexists sbond(\alpha_{A_j}, \alpha_{A_x}) \wedge \\ & (\bigwedge sBond(\alpha_{A_i}, \alpha_{A_j}) \vee \\ & \bigwedge dBond(\alpha_{A_i}, \alpha_{A_j})) \\ 0: & else \end{cases} \tag{4}$$

$\forall 1 \leq j \leq k$ with $i \notin \{1, \ldots, k\}$ and $x \in N$. With that, the situation inside a molecular star is that each non-centric actor is only associated with the centric actor node. The connection is generally a single bond, either from the centric actor node α_{A_i} to each neighbors α_{A_j} or vice versa. In case that two *actor nodes* α_{A_i} and α_{A_j} share a *double molecular bond*, we call this a *molecular bridge*. The molecule structure is of a binary type since exactly two *atomic nodes* are involved (Figure 2b) and Figure 2c)).

$$\mathbf{mBridge}(\alpha_{A_i}, \alpha_{A_j}) = dBond(\alpha_{A_i}, \alpha_{A_j}) \tag{5}$$

Besides, *atomic actor nodes* may play a role inside each molecule, being either a node that actively stimulates another or a node that is passively stimulated by another node, for example *atomic trigger*s and *atomic reactor*s (Figure 1a)). A decomposition of a molecule concerning a semantic assignment may then be as follows: let α_{A_i}, α_{A_j} and α_{A_k} disjunctive *atomic actor nodes*, $1 \leq i \neq j \neq k \leq n$ natural numbers, and n the total number of nodes, then an *atomic actor node* α_{A_i} is a *atomic reactor*:

$$\mathbf{mReactor}(\alpha_{A_i}) = \exists \alpha_{A_j} : sBond(\alpha_{A_j}, \alpha_{A_i}) \tag{6}$$

On the other hand, an *atomic actor node* α_{A_i} is an *atomic trigger*:

$$\mathbf{mTrigger}(\alpha_{A_i}) = \exists \alpha_{A_j} : sBond(\alpha_{A_i}, \alpha_{A_j}) \tag{7}$$

An *actor node* α_{A_i} is a *atomic trigger* if it influences another actor α_{A_j} ($\alpha_{A_i} \Rightarrow \alpha_{A_j}$) with $\omega_{A_i \rightarrow A_j}$ exceeding min_{ω}. An *actor node* α_{A_i} is a *atomic reactor* if α_{A_i} is influenced by another actor α_{A_j} ($\alpha_{A_j} \Rightarrow \alpha_{A_i}$) with with $\omega_{A_j \rightarrow A_i}$ exceeding min_{ω}. Third, we may describe an *actor node* α_{A_i} as a *atomic center* if it is the central point inside a *molecular star*.

There exist operational functions that are applicable to relationships inside the molecules from a practical point of view. A first measure is the distance between two actor nodes $d(\alpha_{A_i}, \alpha_{A_j})$, which leads to us the shortest path problem from α_{A_i} to α_{A_j}. But while passing inner actor nodes, the distance, however, must be influenced by the corresponding activation states σ_{A_i} as well. We therefore propose a distance measure that sums up all actor node activities being on the the shortest path from α_{A_i} to α_{A_j}. Since we can not guarantee that an actor node A_j can be reached from actor node A_i and as we do not know if all participating actor nodes share at least a single molecular bond with its successor, we suggest

$$d(\alpha_{A_i}, \alpha_{A_j}) = \begin{cases} \sum_{k=i}^{j} \sigma_{A_k}: & A_i \rightarrow A_j \\ undefined: & else \end{cases} \tag{8}$$

Depending on the relationship, the strength $s(\alpha_{A_i}, \ldots \alpha_{A_j}$ of a relationship must be calculated in a different way. For example, a *molecular star* is certainly depending on the number of associated actor nodes α_{A_j}, their activation states σ_{A_j}, and the connection weights $\omega_{A_i \rightarrow A_j}$ and $\omega_{A_j \rightarrow A_i}$, respectively, among them:

$$\mathbf{s}_{mStar}(\alpha_{A_1}, \ldots \alpha_{A_n}) = \prod_{k,l=1}^{n} \omega_{A_k \rightarrow A_l} \tag{9}$$

with $k < l$. For *molecular bridges*, however, the "harmony" between the double bonds justifies a stronger bridge (than a bridge with more varying single bonds). The more harmonic a double bonds is the stronger the bridge is.

$$\mathbf{s}_{mBridge}(\alpha_{A_i}, \alpha_{A_j}) = \omega_{A_i \rightarrow A_j} \times \omega_{A_j \rightarrow A_i} \tag{10}$$

with $k < l$.

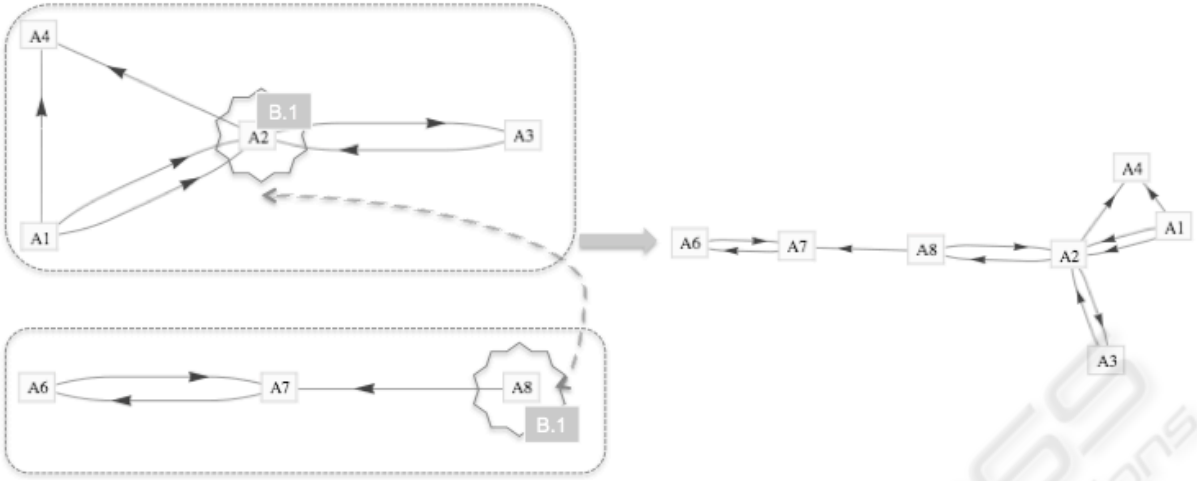

Figure 3: Reaction of two independent molecules to a functional protein. The merge is initiated by a catalyst τ that forms a catalytic bridge between α_{A_2} and α_{A_8}. The merge evolves because of a common interest B_1: in this case, it is an identical classification chapter within the ACM classification system.

2 CATALYTIC BRIDGES

So far, the described molecular model stands for a static description of the theory of graph. Therefore, to overcome such a static molecular existence, we extend our model by taking advantage of the set of items Γ_{A_i} for each α_{A_i}. It is interesting that each *atomic actor node* is allowed to *react* with another α_{A_j} through a catalyst τ: in case that an actor node α_{A_i} with a set of items ξ_i owns the same or a subset of interests than another actor node α_{A_j}, then both may react, merge and establish a *catalytic bridge* τ_{A_i,A_j} (Figure 3).

A functional protein Π_k is therefore unlike a static collection of nodes but moreover a vivid (artificial) and autonomous system. Besides, we understand these functional proteins as an operating structure that is commissioned to complete tasks: it is conceivable to send information and to describe the structure it obsesses. Functional proteins may be forced to continuously improve its own structure: such an improvement may be the update of existing single or double bonds or atomic nodes (decrease/increase).

3 AN APPLICATION EXAMPLE

In case of bibliographic networks, the existence of molecular stars and bridges leads to a more detailed characterization of an author (node); for example, a protein may (autonomously) inform about experts – who form a central part (node) within a molecular star – inside a community and/or about noticeable actor/co-authorships through molecular bridges. A *functional protein* may therefore autonomously (and independently) send information to its environment, for example to natural users while providing them with information about the existence of "interests" and/or the structure of the associated community (retrieve). With the mentioned semantic roles of an actor, the activity of the protein – and with it the community as well – can be measured. The more active *atomic trigger* exist, the more active the community generally will be. On the other side, a more reactive community exists if the number of reactor nodes are lower.

If actor atomic author nodes are interconnected by a catalytic bridge, then they substantiate a common interest. For example, this may occur with respect to common research topics, possibly identified by the ACM classification system. In any case, a merge between author nodes foster a dynamic-adaptive network behavior, because novel connections between actor nodes may be established depending on their "interest" and the bibliographic stream of incoming publications.

As a consequence, a *bibliographic community* is therefore not only a set of relationships among author nodes but consequently previously independent molecules by a *catalytic bridge*. So, we understand bibliographic communities as mental systems that are physically expressed by even such functional proteins with *relationships* (author A_i is associated to A_j) and *semantic roles* (A_i is a trigger, A_j the central point of a star). These communities change over time while figuring out the protein generation process in response to a bibliographic stream, but will definitely operate

in a non-stream environment as well.

4 CONCLUSIONS

This position paper contains in extracts a bio-inspired model, which follows the natural example of a molecular world and which understands graph-related structures as molecular entities. The main objective is to define a model that autonomously and adaptively behaves while performing internal tasks like the communication with its environment, for example inside communities. In this respect, fundamental components like single/double bonds have been presented as well as simple molecular shapes.

Currently, we are working on the stability of atomic node, molecules, and proteins: a first approach towards the stability of proteins is surely to count the number of actor nodes at time points t and $t-1$, respectively, where we then get

$$\Delta(\Pi_i,t) = \frac{\alpha^t_{A_i} - \alpha^{t-1}_{A_i}}{\alpha^{t-1}_{A_i}} \tag{11}$$

The stability of a protein decreases, if $\Delta(\Pi_i,t) \leq 0$; it increases, if $\Delta(\Pi_i,t) > 0$. Even better, the corresponding activity weights $\omega_{A_i \to A_j}$ of the bonds and the activation state of the atomic actor node σ_{A_i} shall be taken into account. However, the question concerning the stability of molecular bonds and atomic actors is herewith not answered and we for example check up if a "valency" can be simulated as well and if other criteria may be taken to fulfill a merge between actor nodes: when a catalyst starts its activity, does it make a difference to start with some actor node or is it of interesting to distinguish between "begin" and "end" actor nodes? Furthermore, the semantic roles inside a protein surely plays a promising aspect towards the stability, since if all actor nodes are satisfied and "convivial" in some way then the stability surely is stronger than in another situation. Another interesting point is the communication of a protein with its environment, respective with other proteins: how can information smoothly addressed to all proteins? Here, we are currently thinking on taking into account achievements from other bio-inspired systems like artificial immune systems.

ACKNOWLEDGEMENTS

The work is currently been done at the MINE research group of the ILIAS Laboratory, Department of Computer Science and Communication, University of Luxembourg.

REFERENCES

Cai, D., Shao, Z., Hen, X., Yan, X., and Han, J. (2005). Community mining from multi-relational networks. In Lecture Notes in Computer Science Vol. 3721/2005. Springer.

Girvan, M. and Newman, M. E. J. (2002). Community structure in social and biological networks. In National Academy of Science USA, 99:8271-8276.

Inokuchi, A., Washio, T., Nishimura, K., and Motoda, H. (2002). A fast algorithm for mining frequent connected subgraphs. In IBM Research, Tokyo Research Laboratory. IBM Press.

Leach, A. (2001). Molecular Modelling - Principles and Applications. Prentice Hall, 2nd edition. Mika, P. (2007). Social Networks and the Semantic Web. Springer, 1st edition.

Schommer, C. (2008). Sieving publishing communities in dblp. In ICDIM Third International Conference on Digital Information Management. IEEE Computer Society.

Yan, X. and Han, J. (2002). gspan: Graph-based substructure pattern mining. In IEEE International Conference on Data Mining (ICDM 2002). I EEE Computer Society.