SIGRAD 2012
A. Kerren and S. Seipel (Editors)

Towards Interactive Visual Analysis of Microscopic-Level Simulation Data

Martin Luboschik, Christian Tominski, Arne T. Bittig, Adelinde M. Uhrmacher, and Heidrun Schumann

Institute for Computer Science, University of Rostock, Germany

Abstract

In this work, we aim at facilitating the analysis of spatial simulations of particles at the microscopic level. This level poses significant challenges to interactive visual analysis tools. On the one hand, the data may contain up to 100.000 data points, and on the other hand, the data exhibit Brownian motion. As a first step to deal with these challenges, we apply well-accepted techniques to visualize the data and to allow analysts to interact with the data and their visual representation. Preliminary results from a spatial simulation of protein–lipid-raft interaction indicate that interactive visual solutions are indeed a useful addition to the modeling and simulation toolbox.

Categories and Subject Descriptors (according to ACM CCS): I.3.6 [Computer Graphics]: Miscellaneous—Visualization of simulation data

1. Introduction

Interactive visual analysis tools have the potential to significantly aid in modeling and simulation of biological phenomena. Current research investigates spatial simulation at the *microscopic* level, a level that imposes significant challenges to be addressed by a visualization solution. A primary concern is the visualization of thousands of moving entities associated with multiple attributes evolving in space and time. On top of that, the visualization has to deal with chaotic movement, because entities moving at the microscopic level are governed by Brownian motion.

We present our ongoing work to facilitate the simulation of spatial phenomena at the microscopic level under consideration of the above challenges. We show how common visualization approaches can be applied as a first step to support analyzing microscopic-level simulation data.

In the next section, we will provide some more details about the application background, the involved simulation methods, and the generated simulation data. Based on that we describe major challenges to be tackled by interactive visual analysis tools. Then we illustrate how the current simulation practice can be improved by means of interactive visualization of the simulated particles. The results achieved with our preliminary solution attest to the advantages of interactive visual analysis.

2. Simulation of Proteins and Lipid Rafts

Spatial simulation of cell biological phenomena at the microscopic level can reveal previously unknown behavior, even for well-studied components such as the MAPK pathway [TTNtW10]. New modeling and simulation approaches follow this promising avenue. In this work, we deal with membrane surface dynamics of receptor proteins and lipid rafts [NBPH06], which are relevant to cellular signaling (e.g. in the cancer-related Wnt pathway [KYS09]). To gain new insights into the interplay of proteins and lipid rafts, the simulation experts apply a rule-based model description and hybrid simulation approach, called ML-Space, which combines (mesoscopic) reaction-diffusion system and (microscopic) particle-based simulation [BHMU11].

Here the focus is on the particle-based simulation. The simulated models comprise up to two thousand proteins with a diameter of approximately 2 nm and up to a few hundred lipid rafts with diameters varying between 6 and 50 nm. The particles are simulated in continuous 2D space.

During the simulation, the following rules are applied. Based on a diffusion factor, the positions of the particles are updated randomly such that the particles exhibit Brownian motion [CCFV05]. Potential collisions are handled differently depending on the involved particles. When a protein and a lipid raft collide, the protein is placed just inside the

raft's boundary and its diffusion factor is reduced. If two entities of the same type are about to collide (for which no applicable reaction is defined in the model), a new random position is drawn until the collision is resolved. Moving a raft implies taking along all the proteins it includes. That is, the protein positions are updated by their own and their encompassing raft's update vector. When a protein inside a raft moves against the raft's boundary, it is placed right outside of the raft and its original diffusion factor is restored.

The simulation is started with all proteins outside rafts and stopped when the number of proteins inside rafts has reached a steady state. The simulated time frame then is between fractions of a second and a few minutes .

3. Visualization Challenges

Developing interactive visual tools to assist in the simulation is challenging for reasons related to the generated data and for reasons related to the tasks and goals of the analysts.

The data can become quite large with thousands of particles evolving over hundreds of time steps, resulting in data sets that can easily exceed 100,000 data points. On top of that, the simulated particles follow Brownian motion due the physical laws in force at the microscopic level. Therefore, any movement analysis will have to face undirected high frequency motions as the fundamental components of the data and heavily self-intersecting movement trajectories. So visual clutter will be a serious concern.

Within this chaotic motion of many particles, our collaborators seek to gain insight into the complex spatial interplay of different types of particles. In the first place, the scientists need to confirm certain characteristics to validate the simulated model and the simulator itself. Secondly, the data of valid simulation runs need to be analyzed for new findings regarding the simulated biological phenomenon. This analysis involves several facets.

Understanding spatio-temporal characteristics is at the heart of any spatial simulation. This requires appropriate communication of the spatial and temporal dimensions of the simulation as well as the particles' location and movement in space and time. Although this is nothing uncommon, it becomes difficult when several attributes associated with each entity evolve during the movement and have to be visualized accordingly. While the visualization of a single attribute along self-intersecting trajectories is complex for its own, it becomes even more challenging when two or more related attributes need to be shown together.

Besides these primary challenges, there are further aspects concerning the exploration of data. For example, flexible interaction techniques are required to enable the user to focus on specific parts of the data. In order to extract characteristic behavior, it should be possible to compare potentially interesting regions.

Developing tools that address these challenges is our ongoing research. Next we present first steps towards this goal.

4. Visualization of Protein–Lipid-Raft Interaction

Our initial solutions consists of a preprocessing step to enhance the data, a dynamic animation of moving particles and a static visualization of movement trajectories, as well as interaction techniques to filter the data. The following paragraphs will provide the details.

Data Enhancement The raw data hold only sparse information about the simulated proteins and lipid rafts. There are tuples (i,t,p,e), where i is an identifier, t is a time-stamp, p is the particle's position, and e is an optional event flag. Events occur when proteins entered or exited a lipid raft, and when the simulator resolved a potential collision.

Because the time domain is continuous, the simulator reports particle positions and events separately for individual particles at arbitrary time-stamps. In order to ease the data enhancement process, we first transfer the data to a unified time line, where missing data points are computed via linear interpolation. In a second step, we derive data attributes $a_1,\ldots,a_n$ providing further details on the particles movement and thus form tuples $(i,t,p,e,a_1,\ldots a_n)$. Following the suggestions for potentially interesting attributes by Andrienko et al. [AAH*11], we compute speed, direction, and curvature as derived attributes. Furthermore, we take the special simulation background into account and compute for each protein the distance and the relative difference in direction to the closest lipid raft. The enter/exit events are translated to a *containment* attribute. Beside storing attributes on a per-tuple basis, we additionally compute cumulative values per particle (e.g., average speed). This comes in handy when searching for interesting particles through dynamic queries.

Visualization Design Due to the multi-faceted character of the simulation data and the analysis goals, it makes sense to investigate different visualization techniques. Our initial solution is a two-fold design: a dynamic representation of the particle movement and a static representation of derived movement attributes.

The dynamic representation shows animated particles for live monitoring of the simulation. As illustrated in Fig. 1, each particle is represented by a circle whose size corresponds to the particle's diameter. Color is used to encode the type of the particle: lipid rafts are green, proteins inside lipid rafts are red, and proteins outside lipid rafts are blue. To include a bit of the history of the particle movement, old positions are indicated as dimmed circles.

The static representation in Fig. 2 shows the particles' movement as trajectories which encompass all positions a particle has visited during the simulation. We use a straightforward polyline visualization of the trajectories. Along the polylines we color-code a user-selected attribute. That cod-

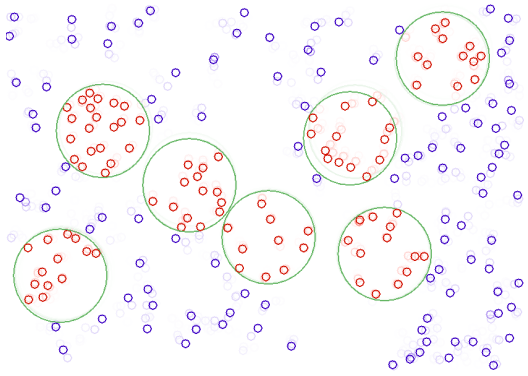


Figure 1: *Snapshot of the particle animation.*

ing is based on an appropriate classification of the attribute's value range and on adapted color schemes from [HB03].

Interaction To address the chaotic character of the Brownian motion, we provide Gaussian kernels of different size. These can be applied and adjusted interactively depending on the concrete data and the desired smoothing results. They have to be used carefully, since smoothing implies a loss of high frequencies in the data. On the other hand it is beneficial for example when examining the overall movement direction of lipid rafts and interior proteins. To further address the serious clutter of trajectories, we provide several interactive filter sliders. These can be used to reduce the time range to be shown or to focus on particles of different type. Filtering out class intervals of the visualized attribute via an interactive legend further helps in focusing on those parts of the trajectories relevant to the task at hand. Common zoom and pan operations and smooth viewport transitions facilitate overview and detail exploration.

Although being rather straight-forward solutions, the insights that could be gained with the help of them are already quite promising.

5. Preliminary Results

The interactive visual approach was successfully applied to confirm that the simulated models are mostly valid. Additionally, new and unexpected findings could be revealed using the interactive visualization.

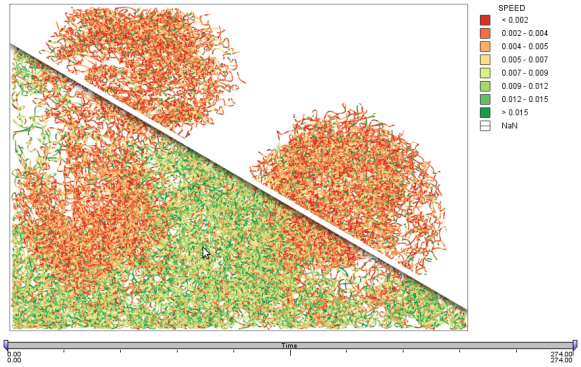
Dynamic Representation The particle animation (Fig. 1) communicates fundamental characteristics of the spatial simulation. When looking at the animation as a whole, one can confirm that the Brownian motion mechanism performs well: All particles move constantly resulting in a fully dynamic animation without any suspicious behavior such as halting particles or preferred moving directions. Another important result is the correctness of the inclusion mechanism. At the beginning, the animation shows empty lipid rafts, which continuously collect proteins during the simulation. The color coding helps in distinguishing outside and inside proteins, with the latter constantly growing in numbers. Finally the lipid rafts reach a kind of steady state, dropping and collecting proteins in equal measure. When looking at individual animation frames, the current particle positions and the particle distribution can be seen without clutter. Since no conspicuous features like overlapping proteins appear, it was concluded that the basic position updates work correctly.

On the other hand, the animation suffers from intense flickering due to Brownian motion. It is difficult to judge particle characteristics like speed, general direction, or collisions. The same applies for the particles' evolution in terms of their movement and their associated attributes. Showing dimmed circles for previous positions and providing interaction techniques to navigate in time are only first approaches to overcome this drawback. Still this issue requires further investigations.

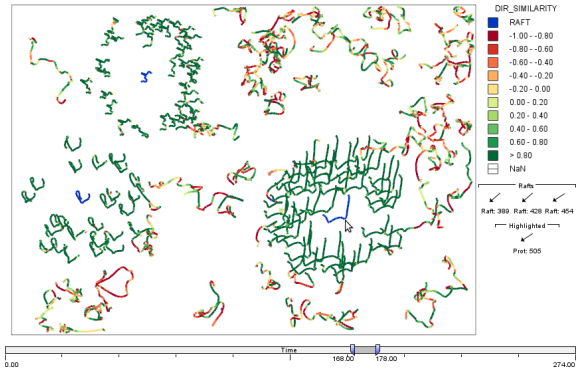
Static Representation The static representation primarily focuses on the evolution of the particle movement. All particle trajectories are shown condensed in one view (Fig. 2). Alternately selecting a derived attribute to be color-coded along the trajectories helps in confirming further aspects of the simulation. For example, Fig. 2(a) (bottom-left) shows the evolution of the particle speed. Although individual particles are difficult to extract, it is overall clearly visible that proteins slow down (red) in certain regions. To find out why, an interactive filter (operating on the containment attribute) is used to focus on protein's trajectory segments that lie inside of lipid rafts. The filtered visual representation in Fig. 2(a) (top-right) reveals that the slow speeds occur exclusively inside lipid rafts. This is the expected behavior as modeled by the rule to reduce the diffusion factor when proteins enter a lipid raft.

The trajectory based visualization is also useful to confirm the synced movement of lipid rafts and interior proteins (Fig. 2(b)). Interactive filters are applied to limit the visible time span and to create appropriately small trajectory extracts. This way, visual clutter is reduced and individual trajectories become visible allowing for the comparison of movement directions. Analyzing the synced movement is supported by a color map that encodes the derived similarity of movement directions (from green: equal direction to red: opposite direction, blue: lipid raft). The colored trajectories clearly show that the lipid rafts and their included proteins follow similar paths, as it was intended by the model's rules.

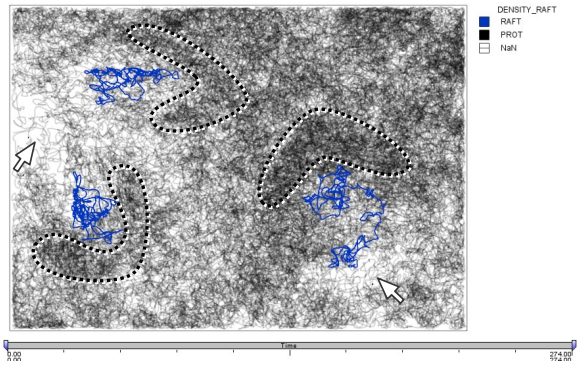
Besides confirming characteristics defined by the model, the trajectory approach also unveiled new findings. Although the dynamic visualization approach revealed no suspicious behavior concerning the distribution of proteins at the level of individual time points, the trajectory visualization does so for the *whole* simulation. In Fig. 2(c) only the particle trajectories (black: proteins, blue: lipid rafts) are shown. What we can see is an uneven distribution of particles: Dense (marked areas) and sparse (arrows) regions emerge near the



(a) Speed of particles.



(b) Similarity of direction.



(c) Dense and sparse regions.

Figure 2: *Color-coded trajectory visualization.*

lipid rafts. This pointed the simulation experts to reconsider their assumptions. A peculiarity of the model was that a moving lipid raft takes up proteins it collides with (given free space inside), but does not leave a "trail". It was reasoned that a possible gap "behind" a moving raft would be filled rather quickly by other particles diffusing into that sparse space. The lighter regions with few trajectories passing through strongly suggest that this is not the case given the relatively low particle density of the test models. This finding prompted investigations of models with different diffusion factors and protein numbers.

But as with the animation approach, the trajectory-based approach has some drawbacks. Some simulation characteristics are not visible at a glance. This includes for example the position of particles at a certain time, the particles' extent, as well as collisions or the inclusion of proteins. Therefore, dedicated attributes and interactive filters have to be used.

6. Conclusion

The challenges identified in this work are significant. Our initial approach to address them is to apply basic visualization techniques combined with appropriate interactive tools. This straight-forward solution already helped simulation experts in their work.

However, more research is required to come up with full-fledged approaches dealing with the Brownian motion, multiple evolving attributes and with larger simulation data. In future work, one could investigate mechanisms to abstract from the Brownian motion in order to allow for a more abstract visual representation. This would also open up possibilities to visualize multiple attributes and to handle larger datasets. Another promising avenue is to follow a feature-based approach to automatically extract interesting regions in the data (like those we marked manually in Fig. 2(c)).

References

[AAH*11] ANDRIENKO G., ANDRIENKO N., HURTER C., RINZIVILLO S., WROBEL S.: From Movement Tracks through Events to Places: Extracting and Characterizing Significant Places from Mobility Data. In *Proceedings of the IEEE Symposium on Visual Analytics Science and Technology (VAST)* (2011), IEEE Computer Society, pp. 161–170. 92

[BHMU11] BITTIG A. T., HAACK F., MAUS C., UHRMACHER A. M.: Adapting Rule-based Model Descriptions for Simulating in Continuous and Hybrid Space. In *Proceedings of the International Conference on Computational Methods in Systems Biology* (2011), ACM, pp. 161–170. 91

[CCFV05] CECCONI F., CENCINI M., FALCIONI M., VULPIANI A.: Brownian Motion and Diffusion: From Stochastic Processes to Chaos and Beyond. *Chaos: An Interdisciplinary Journal of Nonlinear Science 15*, 2 (2005), 026102–1–026102–9. 91

[HB03] HARROWER M. A., BREWER C. A.: ColorBrewer.org: An Online Tool for Selecting Color Schemes for Maps. *The Cartographic Journal 40*, 1 (2003), 27–37. 93

[KYS09] KIKUCHI A., YAMAMOTO H., SATO A.: Selective Activation Mechanisms of Wnt Signaling Pathways. *Trends in Cell Biology 19*, 3 (2009), 119–129. 91

[NBPH06] NICOLAU D. V., BURRAGE K., PARTON R. G., HANCOCK J. F.: Identifying Optimal Lipid Raft Characteristics Required To Promote Nanoscale Protein-Protein Interactions on the Plasma Membrane. *Molecular and Cellular Biology 26*, 1 (2006), 313–323. 91

[TTNtW10] TAKAHASHI K., TĂNASE-NICOLA S., TEN WOLDE P. R.: Spatio-temporal Correlations Can Drastically Change the Response of a MAPK Pathway. *Proceedings of the National Academy of Sciences 107*, 6 (2010), 2473–2478. 91