A Network Analytic Approach to Investigating a Land-Use Change Agent-Based Model

Ju-Sung Lee and Tatiana Filatova

Abstract Precise analysis of agent-based model (ABM) outputs can be a challenging and even onerous endeavor. Multiple runs or Monte Carlo sampling of one's model (for the purposes of calibration, sensitivity, or parameter-outcome analysis) often yields a large set of trajectories or state transitions which may, under certain measurements, characterize the model's behavior. These temporal state transitions can be represented as a directed graph (or network) which is then amenable to network analytic and graph theoretic measurements. Building on strategies of aggregating model outputs from multiple runs into graphs, we devise a temporally constrained graph aggregating state changes from runs and examine its properties in order to characterize the behavior of a land-use change ABM, the RHEA model. Features of these graphs are transformed into measures of complexity which in turn vary with different parameter or experimental conditions. This approach provides insights into the model behavior beyond traditional statistical analysis. We find that increasing the complexity in our experimental conditions can ironically decrease the complexity in the model behavior.

Keywords Agent-based model analysis • Graph representation • Network analysis • Complexity metrics • Land-use change

Paper presentation for the Social Simulation Conference 2015 (SSC, ESSA), Groningen, The Netherlands, Sept. 16–18, 2015.

J.-S. Lee (✉)
University of Twente, Enschede, The Netherlands
Erasmus University Rotterdam, Rotterdam, The Netherlands
e-mail: lee@eshcc.eur.nl

T. Filatova
University of Twente, Enschede, The Netherlands
e-mail: t.filatova@utwente.nl

W. Jager et al. (eds.), *Advances in Social Simulation 2015*, Advances in Intelligent Systems and Computing 528, DOI 10.1007/978-3-319-47253-9_20

1 Introduction

Agent-based models (ABMs) are capable of producing a plethora of complex output which require sophisticated techniques for analysis and visualization. An overview of current ABM output analysis and visualization techniques [8] highlights the breadth of techniques necessary for ABM output analysis. Still, there remain many advanced approaches that have yet to be commonly employed by ABM researchers. In this paper, we explore an network analytic approach for investigating ABM behavior and demonstrate the utility of this approach by applying it to an agent-based land-use change (LUC) model called RHEA (**R**isks and **H**edonics in an **E**mpirical **A**gent-based land market) [3].

1.1 Networked ABM Output

While the interaction of agents in ABMs often constitute networks of varying modalities (i.e., multiple classes of entities) and topologies (i.e., the shape of the overall structure), the model's state transitions from one discrete time period to the next can also be represented as network, specifically a *weighted directed graph*, and then subjected to network analytic methods particularly drawn from graph theory and *social network analysis* (SNA). The weights of the edges in such graphs would represent the number of times the two connected states were traversed in one or more runs of the simulation. Crouser et al. [2] propose the use of *aggregated temporal graphs* (ATGs), the vertices of which encapsulate a unique state configuration of the model. Multiple model runs are then aggregated into a single graph.

Examination of a smaller, more focused set of unique states would yield fully connected graphs (or cliques) in which only the edge weights vary, and the aggregation would lose much of the model's complexity. For example, a model that oscillates between two states in a staggered pattern would appear as a small graph with only two vertices and a set of bidirectional edges while the pattern of transitions would be lost. To address this limitation, we apply a variant of this approach in which time is disaggregated in the graph portrayal; we call this variant *temporally constrained aggregated graphs* (TCAGs). Each vertex under this approach is identified uniquely by both its state and time signature. The multiple edges (or edge weight) between two vertices would then indicate the number of simulation runs that traverse the states represented by those two vertices at that specific time interval. While a plot of lines may also be used to portray these state changes, its visual limitations make it unsuited for our analyses.

While the two aforementioned network analytic approaches bear some resemblance to *time aggregated graphs* [6, 9], their techniques are quite different from the latter, which entails an aggregation of the edges for a fixed vertex set (of dynamic social and transportation networks) rather than an aggregation of vertices for graphs of varying vertex sets.

1.2 The RHEA Model

The RHEA model simulates a bilateral, heterogeneous housing market in an empirical landscape. Buyers search for properties in a seaside town where both coastal amenities and environmental risks of flooding are spatially correlated. A realtor agent serves as a mediating agent that learns about the current market trend and willingness to pay for various spatial attributes of properties [3]. At each time step, sellers of the properties advertise their ask prices while buyers select and offer bids on properties. Then, the two groups engage in dyadic (pair-wise) negotiations, which may or may not result in a successful transaction (or trade). The ask prices are primarily market-driven: sellers (1) consult with a realtor on the appropriate price given the current trends as estimated by a hedonic analysis of the recent simulated transactions and (2) adjust the ask price if a property remains too long on a market. Buyers bid on the single property that brings them maximum utility and is affordable for their budgets. A buyer's budget constraints include their income, preference for amenities, a necessity to pay insurance premiums for flood-prone properties, and the possibility to activate an annual income growth. Spatial patterns of price dynamics and intensity of trade are the emergent outputs of the model. When studying them under various parameter settings, we noticed that the relative market power of buyers and sellers plays a role. The explanation is twofold. Firstly, a parcel that is very attractive will most likely receive several bid offers, out of which its seller chooses the highest, thus driving prices for the most desirable properties up. Secondly, sellers of the properties that are less desirable may receive only one or even no bid offers, which can result in their accepting a bid that is below their ask prices or reducing their ask prices after a number of unsuccessful trades. Thus, excess demand drives prices up while excess supply pushes them down.

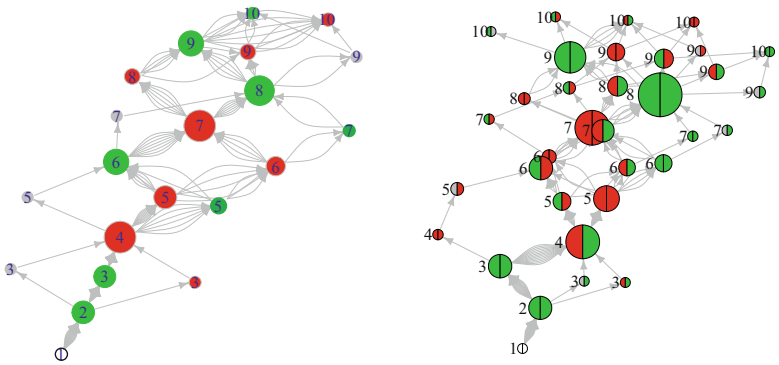
For our analyses, which includes an application of the TCAG on RHEA output, we focus on exploring the dynamics of buyers and sellers count under several key primary parameter settings. The *realtor hedonic* parameter is a binary indicator of whether the realtor agents update their formula based on the evolving market prices (*adaptive*) or retain the empirically informed *static* formula [1]. The *insurance* parameter is a binary indicator for whether or not buyers consider flood insurance in their utility purchase calculation. Engaging the *growth income* indicator parameter will allow agents' incomes, their travel costs, and the insurance premium to rise over time. Therefore, insurance and income growth directly impact buyers' utilities of the properties they consider for purchase.

2 Methods and Results

2.1 *Temporally Constrained Aggregated Graphs*

Under TCAG, ABM states are graph vertices, and the edges indicate temporal transitions and possibly state transitions. For these analyses, we focus on the buyer and seller population sizes which rise and fall throughout the simulation. In Fig. 1, we employ TCAG on the first ten time periods (t) across 15 simulation runs.[1] These runs were executed with activation of adaptive *realtor hedonics* and no activation for *insurance* and *income growth*. In Fig. 1a, we display the TCAG of the changes in the buyer population, and in Fig. 1b, we jointly track changes in the sizes of both the buyer and seller pools. Hence, each state represents a positive, negative, or no change in the population size from the previous time step. When the state consists of changes in only one measure (buyer count), there are three distinct states. When both measures are considered, this space grows to $3 \times 3 = 9$ distinct states.

In the TCAGs, we can visually observe several features of the model's behavior. For example, the initial stages of the model appear to incur fewer distinct states especially when examining both measures jointly. This feature has implications on further development of our model. That is, if we wish to further ground the model events in reality and in real-time, then its behaviors should reflect the fact that the



(a) (b)

Fig. 1 Temporally constrained aggregated graphs of RHEA. *Green* indicates an increase from the previous time period; *red* indicates a decrease; and a *gray* node indicates no change. In the *right* graph, the *left* and *right* halves of each vertex, respectively, indicate changes to the buyer and seller pool counts. The *numerical labels* indicate the time step (t). The vertices have been sized by betweenness centrality (explained below). (**a**) Buyer counts only, $t_{\max} = 10$. (**b**) Buyers and sellers, $t_{\max} = 10$

[1] Our analyses here employ a subset of the output as longer and more runs render the current visualization less effective.

events occur in a pre-existing market, and at least the count of initial states should not be so distinct barring the occurrence of exogenous forces such as a major flood prior to model execution.

We also observe states that serve as *gateways* through which all or many trajectories pass. A pure, isolated gateway occurs when there exists a single vertex for a time step; all trajectories pass through such a vertex. In Fig. 1a, the sole vertices corresponding to times $t = 2$ and 4 are pure gateways. That is, all of the examined simulation runs incur an initial increase in buyer count and a decrease two periods later. These pure gateways reveal an almost deterministic aspect of the model's behavior. In this case, the buyer population will always initially grow and, more often than not, continue to grow until $t = 3$ after which point the market always reacts with a decrease in the buyer pool as indicated by the red, gateway vertex at $t = 4$. In Fig. 1b, the only pure gateway also occurs at $t = 2$ and corresponds to an initial growth in both the buyer and seller populations. We surmise that there exists a potential or gradient in these agent populations set forth by the model's initial conditions leading to the inevitable growth.

Alternatively, a non-isolated gateway is defined by the number and pattern of simulation trajectories that pass through the vertex. One might simply maintain a tally of passing trajectories, but this measure would fail to capture the bridging role a vertex plays in the overall set of state changes. A more robust measure is *betweenness centrality*, a measure widely used in social network analysis. Betweenness centrality (C_B) measures the extent to which a vertex lies in the pathways between all pairs of vertices, and thus captures the extent to which a vertex is a gateway while accounting for the global graph structure.[2] In order to account for higher betweenness for those vertices that receive or emit many edges, we substitute multiple edges with a single edge having a weight of the count of edges. These weights are inverted (i.e., 1/weight) so that the weighted edge between two vertices consecutively traversed in many simulation runs will be considered a shorter path than if they were traversed by few simulations. Given that C_B of vertices near the extremities will be biased downward, we scale the betweenness score by the temporal positions of the vertices such that those in the middle of the time span are penalized:

$$C_\mathrm{B}^\mathrm{t}(i) = \frac{C_\mathrm{B}(i)}{(t_i - 1)(t_\mathrm{max} - t_i)}. \tag{2}$$

[2]More precisely, betweenness centrality is the sum of the proportions of shortest paths a vertex lies on between each pair (out of all shortest paths for each pair) [4, 5]. For betweenness centrality, we identify all the shortest paths in a graph, such that $\sigma_{s,t}$ is the number of shortest paths between vertices s and t. Betweenness centrality for vertex i then:

$$C_\mathrm{B}(i) = \sum_{s \neq i \neq t} \frac{\sigma_{s,t}(i)}{\sigma_{s,t}} \tag{1}$$

.

Under this scaling, a simple TCAG with one vertex per time step would exhibit a C_B^t of 1 for all of the non-terminal vertices. We have sized the vertices in Fig. 1 proportional to C_B^t.

Not surprisingly, the gateway in Fig. 1a at $t = 4$ has high betweenness partly due to the expansion of states in the previous periods. Its betweenness would be lower had there been only one state at $t = 3$. Thus, the state change at that time plays a more significant role in the model's behavior. In Fig. 1b, we observe at $t = 8$ a green/green vertex (increase in both buyer and seller counts) as having the highest relative C_B, despite the existence of other states at that time. Multiple vertices at any time step potentially dilute the betweenness of the vertices in that step so a high betweenness here is particularly salient. In TCAG, a high C_B for a vertex indicates not only a relatively large number of passing trajectories but also high complexity in vertices and trajectories before and after itself. Thus, the gateway vertex at $t = 8$ exhibits high betweenness not simply because many trajectories pass through it but also because the structure of the subgraphs in the time steps prior to and after itself has significant heterogeneity. If, on the other hand, the trajectory structure (either before or after this gateway) was a single sequence, the gateway's betweenness centrality would be much lower.

2.2 Measuring Graph and ABM Complexity

Furthermore, the very count of vertices of aggregated states may describe the complexity of the model given that the states of similarly behaving runs would aggregate into a single chain of events. A perfectly random, or maximally complex, model will yield transition pathways that are less aggregable across simulation runs. The upper bound of vertices would then be $(t_{max}-1)\times n_{states}+1$ vertices where t_{max} is the maximum time in the TCAG and n_{states} is the number of distinct states. While the upper bound for Fig. 1a is $9\times3+1=28$, our TCAG contains only 21 vertices. The upper bound for Fig. 1b is $9\times9+1=82$ while the TCAG contains 34 vertices. The lowest complexity occurs when all model outputs exhibit identical states at each step though the state from one period to the next could vary[3]; the complexity for such graphs is $t_{max}-1=9$. Our *vertex complexity* score (α) then is the vertex count as a proportion of the range defined by these upper and lower bounds:

$$|V(G_l)| = t_{max} \tag{3}$$

$$|V(G_u)| = (t_{max}-1)\times n_{states}+1 \tag{4}$$

$$\alpha = \frac{|V(G)|-|V(G_l)|}{|V(G_u)|-|V(G_l)|} \tag{5}$$

[3] An alternative formulation would differentiate between a graph comprising homogeneous states and one with heterogeneous ones.

where G_l and G_u are graphs corresponding to minimal and maximal complexity, respectively, and $|V(G)|$ is the count of vertices of some graph G. A naïve inference would place our model's complexity as being neither minimally or maximally complex: the univariate measure (buyers only) yields an $\alpha = 0.61$ while a bivariate analysis (buyers and sellers) yields $\alpha = 0.33$. Hence, depending on the complexity of the analyzed output measures, the RHEA simulation's complexity is either $\frac{1}{3}$ or $\approx\frac{2}{3}$.

We also consider the complexity of the sequence of state changes across model runs using data compression as a benchmark. As a straightforward test, we use the extent of compression of the output state change data as our secondary measure of compression: the greater the data compression, the less complex the output. We base this compression factor on the performance of the Unix utility `Gzip` which employs both LZ77 (Lempel–Ziv) and Huffman coding compression schemes [7, 10]. Specifically, we select the minimum of the compression ratios afforded by the matrix of states, ordered by time step (row) by simulation run (column), and its transpose. In Fig. 2, we chart both vertex and Gzip complexities for the four combinations of activation of two key parameters that impact buyer utility: *insurance* and *income growth*.

Firstly, we notice the inclusion of insurance and income growth markedly decreases the complexity of the model in its initial phase while the complexities are less variable in the later stages. This pattern appears to hold, though somewhat diminished, in the bivariate analysis. Furthermore, the data compression (Gzip) complexity roughly produces similar trends as the vertex complexity. However, further investigation into how and why they are different is warranted.

We visually confirm the effect of the insurance and income growth parameters on simplifying the graphs in Fig. 3. While the graphs in Fig. 3 appear similar to those of Fig. 1, there are differences particularly in the initial time steps, which now incur fewer vertex states. Furthermore, this decrease in complexity is accompanied by

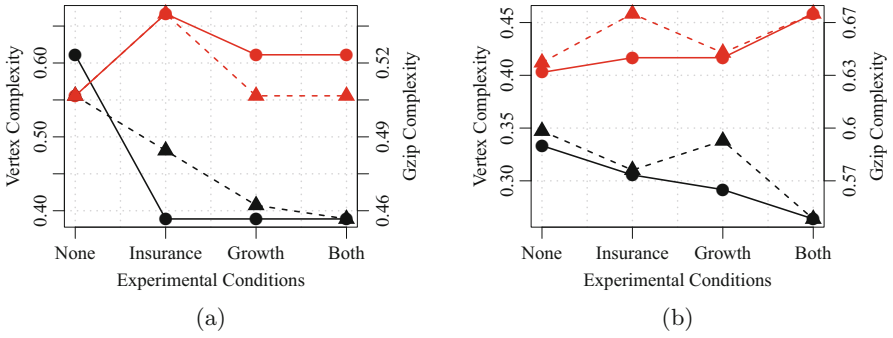
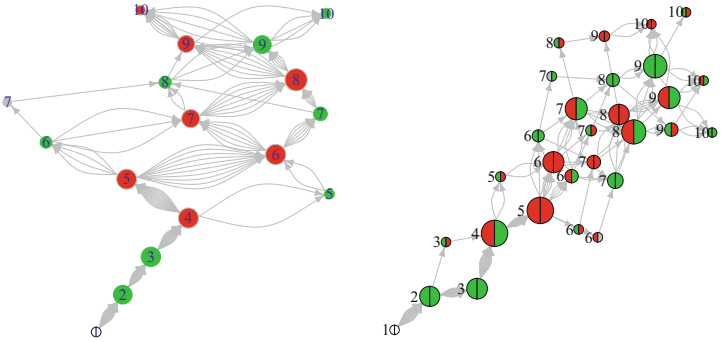


Fig. 2 Complexity of TCAGs. The *black* and *red lines*, respectively, denote the complexities for the first ten periods and the last ten periods (50–60) of the simulation run. The *solid* and *dashed lines* denote the vertex and Gzip complexities, respectively. (**a**) Univariate (buyers only). (**b**) Bivariate (buyers and sellers)



(a) (b)

Fig. 3 Temporally constrained aggregated graphs of RHEA for insurance and income growth conditions. (**a**) Univariate (buyers only). (**b**) Bivariate (buyers and sellers)

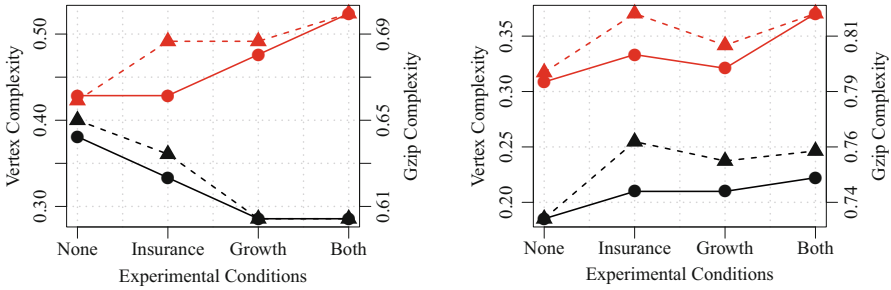


Fig. 4 Complexity of TCAGs for Static Hedonics. The *black* and *red lines*, respectively, denote the complexities for the first ten periods and the last ten periods. The *solid* and *dashed lines* denote the vertex and Gzip complexities, respectively. (**a**) Univariate (buyers only). (**b**) Bivariate (buyers and sellers)

fewer gateway vertices indicated by the lack of largely sized vertices in both Fig. 3a and b. In fact, higher betweenness centrality scores can only occur in more complex TCAGs. Given that the betweenness measure for a given vertex scales with both the trajectories that pass through it and possibly the count of vertices that occur before and after itself, the measure itself encapsulates some portion of vertex complexity.

In the outputs above, we had activated another key parameter: the use of *adaptive realtor hedonics* to determine market prices. So next, we examine the effect of a market price determination through *static* hedonics on RHEA's graph complexity. In Fig. 4, we plot the various complexities in the univariate and bivariate cases.

The impacts of the static hedonic condition are modest. While the initial TCAG complexity for the univariate graph decreases here as it does for the adaptive hedonic setting, the magnitude of the decline (due to activation of both insurance and income growth) is smaller. Secondly, the TCAG based on the final stages of the model runs

exhibits an increase in complexity through the activation of the two parameters; earlier, we observed no distinct trend in the final stages for the *adaptive* realtor hedonic condition. The bivariate situation departs from the declining trend: a mild increase is visible (Fig. 4b). All this points to the subtle interaction between the market evolution (or lack thereof) and complexity in the buyer utility function.

3 Discussion

The portrayal of aggregated state transitions in the RHEA ABM offers an alternative view of its behavior from traditional methods of visualization (e.g., line graphs) and statistics (e.g., regression models); this approach may provide additional insights. The application of network analytic methods to understand the complexity of a model (here, RHEA) at various stages further develops our understanding of the model's subtle behavior particularly in reaction to varying parametric conditions. In fact, we can associate these conditions to the complexity in the model's behavior.

This approach warrants further development. Specifically, the metrics employed offer more information for subsets of the model's time span rather than in its entirety due to the limited heterogeneity in the outputs imposed by a small number of states. Naturally, we would need to next adapt the measurements to allow for a larger number of states. Furthermore, accounting for the heterogeneity of states across time (e.g., oscillation vs. homogeneous sequences) can lead alternative, possibly more robust, complexity measures. The betweenness centrality measure may be further exploited for these purposes as it accounts for some aspect of vertex complexity in exposing key, gateway states. Alternative scaling for the betweenness measure is worth investigating. For example, scaling not just by the temporal position but also by the vertex counts before and after the measured vertex would allow the measure to be strictly based on the structural features of the trajectories. Finally, the complexity induced by graphs drawn from random state sequences would be quite informative, allowing us to assess if our analyzed outputs have some equivalence to noise or not.

Acknowledgements This material is based on work supported by NWO DID MIRACLE (640-006-012) NWO VENI grant (451-11-033), and EU FP7 COMPLEX (308601).

References

1. Bin, O., Kruse, J.B., Landry, C.E.: Flood hazards, insurance rates, and amenities: evidence from the coastal housing market. J. Risk Insur. **75**(1), 63–82 (2008)
2. Crouser, R.J., Freeman, J.G., Winslow, A., Chang, R.: Exploring agent-based simulations in political science using aggregate temporal graphs. In: IEEE Pacific Visualization Symposium (PacificVis). IEEE, New York (2013)

3. Filatova, T.: Empirical agent-based land market: integrating adaptive economic behavior in urban land-use models. Comput. Environ. Urban. Syst. **54**, 397–413 (2014)
4. Freeman, L.C.: A set of measures of centrality based on betweenness. Sociometry **40**(1), 35–41 (1977)
5. Freeman, L.C.: Centrality in social networks: conceptual clarification. Soc. Netw. **1**(3), 215–239 (1979)
6. George, B., Shekhar, S.: Time-aggregated graphs for modeling spatio-temporal networks. In: Roddick, J.F., Benjamins, V.R., Si-said Cherfi, S., Chiang, R., Claramunt, C., Elmasri, R., Grandi, F., Han, H., Hepp, M., Lytras, M., Misic, V., Poels, G., Song, I.Y., Trujillo, J., Vangenot, C. (eds.) Advances in Conceptual Modeling - Theory and Practice. Lecture Notes in Computer Science, vol. 4231, pp. 85–99. Springer, Berlin (2006)
7. Huffman, D.: A method for the construction of minimum-redundancy codes. Proc. IRE **40**(9), 1098–1101 (1952)
8. Lee, J.S., Filatova, T., Ligmann-Zielinska, A., Hassani-Mahmooei, B., Stonedahl, F., Lorscheid, I., Voinov, A., Polhill, G., Sun, Z., Parker, D.: The complexities of agent-based modeling output analysis. J. Artif. Soc. Soc. Simul. **18**(4), 4 (2015)
9. Shekhar, S., Oliver, D.: Computational modeling of spatio-temporal social networks: a time-aggregated graph approach. In: 2010 Specialist Meeting – Spatio-Temporal Constraints on Social Networks (2010)
10. Ziv, J., Lempel, A.: A universal algorithm for sequential data compression. IEEE Trans. Inf. Theory **23**(3), 337–343 (1977)