Proceedings of the 2021 Winter Simulation Conference
S. Kim, B. Feng, K. Smith, S. Masoud, Z. Zheng, C. Szabo, and M. Loper, eds.

AN UNCERTAINTY QUANTIFICATION APPROACH FOR AGENT-BASED MODELING OF HUMAN BEHAVIOR IN NETWORKED ANAGRAM GAMES

Zhihao Hu, Xinwei Deng

Department of Statistics
Virginia Tech
800 Washington St. SW
Blacksburg, VA 24061, USA

Chris J. Kuhlman

Biocomplexity Institute & Initiative
University of Virginia
1827 University Avenue
Charlottesville, VA 22904, USA

ABSTRACT

In a group anagram game, players are provided letters to form as many words as possible. They can also request letters from their neighbors and reply to letter requests. Currently, a single agent-based model is produced from all experimental data, with dependence only on number of neighbors. In this work, we build, exercise, and evaluate enhanced agent behavior models for networked group anagram games under an uncertainty quantification framework. Specifically, we cluster game data for players based on their skill levels (forming words, requesting letters, and replying to requests), perform multinomial logistic regression for transition probabilities, and quantify uncertainty within each cluster. The result of this process is a model where players are assigned different numbers of neighbors and different skill levels in the game. We conduct simulations of ego agents with neighbors to demonstrate the efficacy of our proposed methods.

1 INTRODUCTION

1.1 Background and Motivation

In previous work (Cedeno-Mieles et al. 2018; Ren et al. 2018), online group anagram games (GrAGs) or experiments were conducted, where players share alphabetic letters to form words. Experiments were performed with human subjects as a priming task to generate collective identity within a group. Collective identity (CI) is an individuals cognitive, moral, and emotional connection with a broader community, category, practice, or institution (Polletta and Jasper 2001). From experimental data, an agent-based model (ABM) was developed to enable simulation of games for conditions beyond those tested. The experiments and model are described below as background, and then we describe the need for model enhancements, which provides the motivation for this work.

Group Anagram Game. Figure 1 provides an illustration of four consecutive time steps in our GrAG. Communication channels are in purple, and on these channels a player may request letters and reply to letter requests. Overall, a player may take any of four actions, any number of times, and in any order: (i) request a letter from a neighbor (request sent), (ii) reply to a request with the letter (reply sent), (iii) form a word (form word), and (iv) think or idle (i.e., a no-op condition). In our online experiments, players are initially given three letters (shown in brown, in the black boxes). Players can request letters from their neighbors, and neighbors can reply with the requested letters. For example, v_1 requests a e from v_2 at time t, and v_2 sends a reply with the e at time $(t+1)$ so that e gets added to v_1's letter set, with which it forms words. Received letters are shown in black. If a player shares a letter with the requestor, then both the requestor and the player replying have a copy of the letter. A person never loses a letter, even when they share it with others. This is to encourage sharing letters and forming more words; players evenly split the total earnings from a game, where the earnings are proportional to the total number of words formed by the *team*. Also, a person may use a letter in any number of words, and any number of times within a word. For example,

978-1-6654-3311-2/21/$31.00 ©2021 IEEE

v_1 uses *s* and *e* in forming *sea* at time $(t+2)$ and in forming *seem* at $(t+3)$, and uses the single *e* twice in the latter word. Games are played for 300 seconds (5 minutes). Each player typically forms between 10 and 40 words. The setup in Figure 1 is specifically designed to illustrate the various actions. In reality, most player actions, at each second of a 300-second game, are think or idle; our models are consistent with this behavior. Also, players are motivated in a game because players split evenly a reward (sum of money) that is proportional to the total number of words formed by the team. Additional game details are in (Ren, Cedeno-Mieles, et al. 2018).

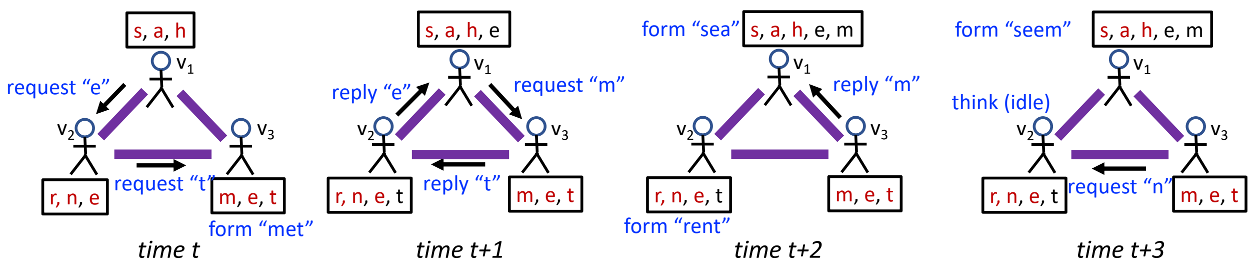


Figure 1: Illustrative play in a group anagram game (GrAG) among three players v_i, $1 \leq i \leq 3$, each with three initial letters in brown, within a black box. Received letters from neighbors are in black. In the model of these games, a player executes one action per time step where a time step is one second. The game is described in the text.

Agent-based model (Ren et al. 2018). A multinomial logistic regression model was developed to predict a player's action at time $(t+1)$ based on the player's action at time t and the values of a temporal four-dimensional vector $\mathbf{z}(t)$ that accounts for the number of letters a player has in her hand, the number of letter requests that the player has not yet replied to, the number of words formed, and the number of consecutive seconds that a player has engaged in the current action (Ren et al. 2018). This latter element applies mostly to the idle or thinking action. The resultant ABM model provides a representation of *mean behavior* over all experiments, with an explicit dependence on the number of neighbors of a player.

Motivation for work. In GrAGs, behaviors can vary significantly among players. The mean model does not capture this heterogeneity. A baseline ABM that only captures mean behavior is problematic because it implies that all agents will, over time and/or over many simulation instances, tend to the mean behavior. Thus nullifies some of the prime motivators for using an ABM approach; most importantly, incorporating heterogeneous agent behaviors.

Moreover, there are limitations to the amount of experimental data that can be collected. We used Amazon Mechanical Turk (AMT) to recruit game players, and it is well-known that AMT does not provide an unlimited pool of candidate players, and some candidates do not show up for experiments (Mason and Suri 2018). We also constrained our experiments so that a person could only play the GrAG one time, to obviate learning from past experience. Consequently, we encountered limitations in the size of our candidate pool, resulting in fewer completed games than we desired. This produced two problems to overcome in building ABMs of game player behavior: data sparsity and variability.

These challenges motivate the development of a general uncertainty quantification approach for building ABMs of human behavior in the face of data sparsity and data uncertainty, for the networked anagram game. The proposed methodology enables agents in agent-based simulations (ABSs) to behave heterogenously (we use "simulation" for computations of a simulation; we use "modeling" for the process of constructing the model used in simulation). We believe that our uncertainty quantification methodology for ABM is also applicable to other types of experiments that involve human behavior, such as (Mason and Watts 2012). Specifically, our methods center on actions performed by players in games. In our game, each player can take multiple types of actions and each action can be taken multiple times; our model predicts when each agent takes each action. Our methods can be applied to other games that have similar features.

1.2 Novelty of Our Work

The primary objective of this work is a general uncertainty quantification (UQ) approach for building ABMs of human behavior in the networked GrAG, such that different agents can have heterogenous behaviors supported by uncertainty quantification of the experimental data. The novelty of this work is in the following aspects. First, we conduct rigorous hypothesis testing to examine the homogeneity of players with different numbers k of neighbors ($k = 2$, 4, 6, and 8) in the networked anagram game. This enables data aggregation into a smaller number of homogeneous groups with more meaningful interpretations of the groups in terms of players' activities (e.g., sending requests, replying to requests, forming words). Second, a clustering technique is adopted to further partition the players from each homogenous group into several clusters, such that the players in different clusters reflect the variability of players' abilities, from low performers to high performers, in a principled way. Third, we construct a distinct multi-logit model for each cluster to estimate the transition probabilities among four action states (i.e., idle, requesting letters, replying to requests, and forming words). The transition probability is the probability that, given action a_i at time t, the next action, at time $(t+1)$ will be a_j. Moreover, we quantify the probabilistic uncertainty of the estimated transition probabilities via the asymptotic distribution of the estimated parameters. Our methods are not restricted to experiments with network structure. Specifically, any experiment that has a parameter which is naturally used to partition data can be analyzed with this approach for UQ.

1.3 Our Contributions

The contributions of this work are in both methodology and application in the context of modeling and simulation of the anagram game.

First, we make a methodological contribution. The proposed approach for uncertainty quantification, using hypothesis testing and clustering, partitions human subjects into skill levels and characterizes uncertainty within each level. It enables uncertainty quantification for each level, rather than having the uncertainty aggregated over the entire data set. The proposed methodology makes model building more general and meaningful in reflecting the heterogeneity of players' behaviors in the anagram game. Through uncertainty quantification, we can categorize human performance relative to the modeled behavior. Explicitly, these approaches enable us to specify and quantify what it means to be a poor- or good-performing agent in a game, and to quantify behavior at multiple levels. The combination of a multi-logit model for transition probabilities with UQ produces a meaningful probabilistic framework, sufficiently complex and flexible. It provides a framework to investigate the impact of these modeling choices on the behavior of an individual *within an interacting team of players* and the overall group/team performance. See Figure 2.

Our second contribution is to make a UQ framework for ABM. The proposed UQ approach for ABM strikes a good compromise (the meaning of which is problem dependent) among model explainability, model flexibility, and model complexity, e.g., (Baker 2016; Pearl and Mackenzie 2018). An ABM model of human behavior is expected to satisfy one or more criteria such as: (1) explainability of human behavior, (2) sufficient model complexity of human behavior, (3) a favorable framework for UQ with consideration of computational cost, and (4) other considerations based on specific problems. It is known that there are various UQ tools in the literature, such as variational inference, conditional random fields, and sensitivity analysis. However, some of these methods only address some of the aforementioned aspects. For example, conditional random fields has an explainable model structure, but may not have sufficient model complexity for human behaviors that are more readily extended to behaviors beyond those studied in this work. The developed methodology can also be characterized for use within ABMs.

Our third contribution is the evaluation of the UQ method in the context of ABM for human group behavior. We demonstrate the merit of the proposed UQ method on a network of game players with a star structure in ABSs; see Figure 5 for a star configuration, with a single hub (center) node linked to leaf (degree-1) nodes. We evaluate the models and methods by locating an *ego* agent within a context, which for social networks here, means specifying the ego node's neighbors. We assign properties to the ego node and each of its neighbors. Using the "ego networks," we can study several factors: the number of neighbors, the skill levels of players, and the structure of networks. Owing to space limitations, we compare model output for two skill levels, and leave further analyses for an extended version of this work. We confirm, for example, that players with lower skill levels form fewer words.

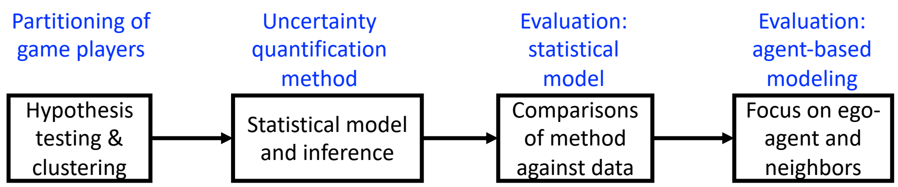


Figure 2: The approach and contributions of this paper in developing a UQ method to capture game player behavior in agent-based models. The first two boxes represent work in Section 4, and the last two boxes represent work on evaluations in Sections 5 and 6.

Paper organization. The reminder of the paper is organized as follows. Related work comprises Section 2. We first present the formalism for the mean model (Section 3), taken from (Cedeno, Hu, et al. 2020), which provides a point of departure for our new work. Then we present uncertainty quantification methods (Section 4), which includes clustering methods for players and quantifying uncertainty for model parameters. Next, we present the model evaluation (Section 5). In Section 6, we provide simulation studies of ego networks, and we conclude in Section 7.

2 RELATED WORK

Modeling group anagram games. Online GrAG games (controlled experiments) and a mean logistic regression model—the starting points for this present work—are described and evaluated in (Cedeno et al. 2020). Cedeno-Mieles et al. (2018) use this model in software pipelines for data analysis and modeling. The logistic regression model is expanded upon in (CedenoMieles et al. 2019). Another anagram model, to study the effects of network structure in GrAGs, is provided in Hu et al. (2019).

Uncertainty quantification methods and analyses. ABMs use rules at the individual level to simulate a social system, which can provide results of structured behavior at an aggregated level. Such ABMs often do not account for uncertainty quantification for model parameter estimation and prediction. In our experiments, human subjects under the same nominal conditions are shown in many instances to behave differently. Thus, it is important to incorporate uncertainty in models of human behavior in the GrAG.

In the literature, design of experiments (DoE) and Guassian process are two commonly-used tools in uncertainty quantification (Ryzhov et al. 2020). For example, Papadelis and Flamos (2019) introduce calibration and UQ techniques for ABMs. They first fit a Gaussian process emulator to ABM results. Then they calibrate the model and quantified uncertainty based on the history matching method. This approach assumes that the model output distribution is normal, conditioned on the mean and variance. Stavrakas et al. (2019) develop an agent-based technology adoption model and performed parameter estimation and uncertainty quantification based on Gaussian process. Fadikar et al. (2018) use an emulation approach within a Bayesian model calibration framework to quantify uncertainty in an epidemic ABM. Zhang et al. (2016) employ an ABM to forecast residential rooftop solar adoption in San Diego county, and they provide a quantification of uncertainty in the model's predictions based on likelihood. Methods using DoE for sensitivity analysis and UQ can be found in (Alam et al. 2015) and the references therein. Other UQ techniques are Bayesian maximum entorpy (Lee and Wentz 2008; House-Peters and Chang 2011) and Bayesian melding (Sevcikova et al. 2007), among many others.

In our framework, we perform clustering of players based on the experimental data, which enables quantifying uncertainty and estimating parameters for our model based on the data in each cluster. The conditional multi-logit model is used for parameter estimation in the transition probability matrix. Then we incorporate this uncertainty into a model that produces heterogeneous behaviors in agents. Here we adopt the multinomial logistic regression to model the transition probabilities of players actions. But the proposed framework of uncertainty quantification is also applicable to other models, such as Gaussian process modeling.

3 MEAN MODEL

Ren et al. (2018), Cedeno et al. (2020) present an ABM of the GrAG. In the model, the set V of players and the set E of their communication channels (edges) are represented as an undirected graph $G(V,E)$. The game is modeled as a discrete-time stochastic process, where at each time step, a player executes one of the actions from the action set A provided in Table 1.

Table 1: The four actions of players in the GrAG and model. The set A of actions is $A = \{a_1, a_2, a_3, a_4\}$.

Item	Variable	Name	Description	Item	Variable	Name	Description
1	a_1	idling	Thinking (a no-op).	3	a_3	request	Requesting a letter from a neighbor.
2	a_2	reply	Replying to a neighbor with a requested letter.	4	a_4	word	Forming and submitting a word.

They also introduce four variables in Equation (1) below: size $Z_B(t)$ of the buffer of letter requests that v has yet to reply to at time t; number $Z_L(t)$ of letters that v has available to use (i.e., in hand) at t to form words; number $Z_W(t)$ of valid words that v has formed up to t; and number $Z_C(t)$ of consecutive time steps that v has taken the same action. $Z_C(t)$ is used to ensure agents do not stagnate in thinking; this parameter forces agents to have a finite deliberation period before acting. Let $\mathbf{z} = (1, Z_B(t), Z_L(t), Z_W(t), Z_C(t))_{g\times 1}$. They use a multinomial logistic regression to model π_{ij}—the probability of a player taking action a_j at time $t+1$, given that the player took action a_i at time t—as

$$\pi_{ij} = \frac{\exp(\mathbf{z}'\beta_j^{(i)})}{\sum_{l=1}^{4}\exp(\mathbf{z}'\beta_l^{(i)})},\ j = 1,2,3,4 \tag{1}$$

where $\boldsymbol{\beta}_j^{(i)} = (\beta_{j1}^{(i)}, \ldots, \beta_{j,g}^{(i)})'$, prime indicates vector transpose, and $\beta_{j,h}^{(i)}$ are the elements of $\boldsymbol{\beta}_j^{(i)}$, with $1 \le h \le g$ being the index of the element of the $\mathbf{z}$ vector. For a given i (the index i on action a_i), the parameter set can be expressed as

$$\mathbf{B}^{(i)} = \begin{pmatrix} \beta_{11}^{(i)} & \beta_{12}^{(i)} & \cdots & \beta_{1,g}^{(i)} \\ \beta_{21}^{(i)} & \beta_{22}^{(i)} & \cdots & \beta_{2,g}^{(i)} \\ \vdots & \vdots & \ddots & \vdots \\ \beta_{41}^{(i)} & \beta_{42}^{(i)} & \cdots & \beta_{4,g}^{(i)} \end{pmatrix}, \tag{2}$$

with matrix entries $\beta_{j,h}^{(i)}$.

Game players in the GrAG data are grouped by their degrees k in the network G, i.e., k is the number of neighbors a player has in the game. To estimate the parameter sets $\mathbf{B}^{(i)}$, they use maximum likelihood estimation across the experimental observations for each $k = 2$, 4, 6, and 8. For a given i and k, denote the corresponding observational data as $\mathscr{D}_k^{(i)}$. Then they conduct parameter estimation by

$$\hat{\mathbf{B}}^{(i)}(k) = \arg\max \log L(\mathbf{B}^{(i)} | \mathscr{D}_k^{(i)}),$$

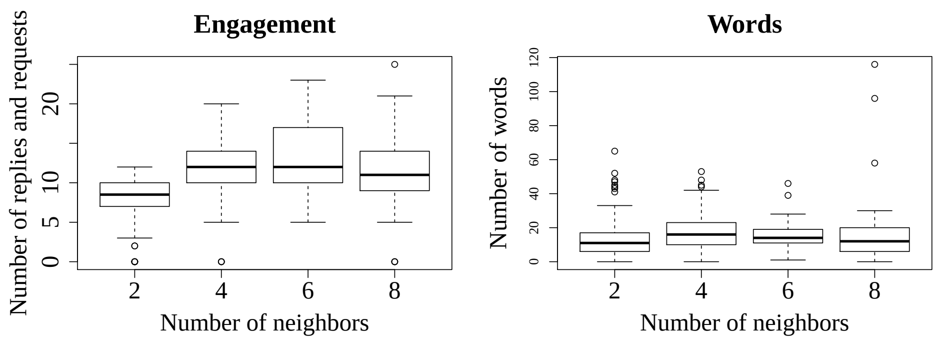


Figure 3: Boxplots of engagements and words. The left plot shows engagements of each player in games with $k = 2$, 4, 6, 8 neighbors. The right plot shows number of words of each player in games with 2, 4, 6, 8 neighbors.

where $L(\mathbf{B}^{(i)}|\mathscr{D}_k^{(i)})$ is the likelihood function with respect to the data $\mathscr{D}_k^{(i)}$ collected under the setting of k neighbors in the experiments.

By applying this model, all agents with the same degrees (i.e. number of neighbors) k in the network G are assigned the same coefficient matrix. Thus, these agents will have the same behaviors in expectation. However, we would like agents to exhibit heterogeneous behaviors. Consequently, we devise a method to produce variability in actions among agents with the same degree k. This is the subject of the next section.

4 UNCERTAINTY QUANTIFICATION METHODS

4.1 Clustering Methods for Players

Our goal is to partition players based on their activity in a game. The number of letters a player requests, the number of letter requests a player replies to, and the number of words a player forms in a game are used to quantify a player's activity in a game. We define two variables, engagements and words. Engagements is the sum of the number of requests and number of replies of a player, and words is the number of words a player forms in a game. The engagements and words are used to partition players.

The number of neighbors of a player in a game could affect the engagements and words. For example, a player can request more letters and reply to more letter requests with more neighbors. Thus, we want to study whether players should be partitioned based on their numbers of neighbors. Figure 3 shows the numbers of engagements and words for different numbers of neighbors. Data were collected at even numbers of neighbors to span a wide range of neighbors; we planned to interpolate at intermediate values. Engagements increase with the number of neighbors, but becomes saturated when the number of neighbors is greater than four. The numbers of words are nearly the same for different numbers of neighbors. In summary, game data for players with $k = 4$, 6, and 8 neighbors are observed to be similar, and these data are different from those of players with $k = 2$ neighbors.

We conduct hypothesis testing on engagements and words to determine whether we can partition players on $k = 2$ neighbors and on 4, 6, 8 neighbors separately. We denote μ_r^{eng} and μ_r^{word} for $r = 2,4,6,8$ as the mean engagements and words for players with r neighbors, respectively. We also denote μ_{468}^{eng} and μ_{468}^{word} as the mean numbers of engagements and words for players with 4, 6, or 8 neighbors, respectively. The hypotheses of our two-sample t-tests are:

$$\begin{aligned}
&H_0^{r,s,eng}: \ \mu_r^{eng} = \mu_s^{eng} \quad vs. \quad H_1^{r,s,eng}: \ \mu_r^{eng} \neq \mu_s^{eng}, \\
&H_0^{r,s,word}: \ \mu_r^{word} = \mu_s^{word} \quad vs. \quad H_1^{r,s,word}: \ \mu_r^{word} \neq \mu_s^{word}, \\
&\text{where } r,s = 2,4,6,8, \ \text{ and } r < s.
\end{aligned}$$

We also perform hypothesis testing on players with 2 neighbors and players with 4, 6, or 8 neighbors, and the hypotheses are:

$$H_0^{eng}: \ \mu_2^{eng} = \mu_{468}^{eng} \quad vs. \quad H_1^{eng}: \ \mu_2^{eng} \neq \mu_{468}^{eng},$$
$$H_0^{word}: \ \mu_2^{word} = \mu_{468}^{word} \quad vs. \quad H_1^{word}: \ \mu_2^{word} \neq \mu_{468}^{word}.$$

To partition players, we use the kmeans clustering method (Hartigan and Wong 1979). Before clustering players, the engagement and words are standardized first, so no variable would dominate the clustering. To determine the number of clusters in the kmeans clustering, we use the Bayesian Information Criterion (BIC) as our criterion (Li et al. 2016). Based on the BIC and the size of data, we select four clusters, which gives the smallest BIC values.

4.2 Quantifying Uncertainty For Model Parameters

In the mean logistic regression model (Cedeno et al. 2020), the next player action depends on the $\boldsymbol{B}^{(i)}$ matrix and $\mathbf{z}$ vector input variables. The $\boldsymbol{B}^{(i)}$ matrix is kept the same for all players in a game. Thus, players will have the same probability vector given the same set of $\mathbf{z}$ vector values. In order to study player behaviors using a hetergeneous model structure, we modify the logistic regression model as below.

The probability vector is $\boldsymbol{\pi}_i = (\pi_{i1}, \pi_{i2}, \pi_{i3}, \pi_{i4})^T$, whose components are calculated using Equation (1). To produce different player behaviors, we devise an approach that uses the asymptotic normal distribution of $\hat{\boldsymbol{B}}$. Different players are assigned different $\boldsymbol{B}^{(i)}$ matrices, thus different players will have different behaviors. To generate different $\boldsymbol{B}^{(i)}$ matrices from the mean model, we use two different approaches, described next.

We first transform the $\hat{\boldsymbol{B}}^{(i)}$ matrix to the $\hat{\boldsymbol{\beta}}^{(i)}$ vector, then we can use the asymptotic normal distribution of parameter estimators $\hat{\boldsymbol{\beta}}^{(i)}$, which is the asymptotic property of maximum likelihood estimators. For simplicity of notation, let $\hat{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}}^{(i)} = (\hat{\boldsymbol{\beta}}_2^{(i)T}, \hat{\boldsymbol{\beta}}_3^{(i)T}, \hat{\boldsymbol{\beta}}_4^{(i)T})^T$, then $\hat{\boldsymbol{\beta}}$ follows a multivariate nomal distribution, $\hat{\boldsymbol{\beta}} \sim \mathrm{MN}(\hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\Sigma}})$. There are two ways to obtain a heterogeneous probability vector: one can draw a random sample from the confidence region of $\hat{\boldsymbol{\beta}}$ or directly draw a random sample from the distribution of $\hat{\boldsymbol{\beta}}$. These are addressed next.

Sampling from the 95% confidence region of $\hat{\boldsymbol{\beta}}$. The first approach is to find the 95% confidence region of $\hat{\boldsymbol{\beta}}$ based on the asymptotic distribution of $\hat{\boldsymbol{\beta}}$. Then we uniformly draw a random point from the confidence region S_β, which is defined as

$$Pr(\hat{\boldsymbol{\beta}} \in S_\beta) = 95\% \tag{3}$$
$$(\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}_{mle})^T \hat{\Sigma}^{-1} (\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}_{mle}) < \chi_d^2(0.95) \tag{4}$$

where $\hat{\boldsymbol{\beta}}_{mle}$ is the estimation of $\boldsymbol{\beta}$, $\hat{\Sigma}^{-1}$ is the estimation of covariance of $\hat{\boldsymbol{\beta}}_{mle}$, $\chi_d^2(0.95)$ is the 0.95 quantile of the Chi-squared distribution with degree of freedom d, and d is the dimension of $\boldsymbol{\beta}$. The feasible region defined above is a ellipsoid. We can use the "uniformly" package in R to generate random points. After we obtain a randomly generated value of $\hat{\boldsymbol{\beta}}$, we are able to compute the corresponding probability vector $\boldsymbol{\pi}_i$ whose components are given by Equation (1).

Sampling from the distribution of $\hat{\boldsymbol{\beta}}$. The second approach is very similar to the first approach, but we do not need to find the confidence region S_β. Instead, we directly sample from the asymptotic normal distribution of $\hat{\boldsymbol{\beta}}$. Then we can calculate the corresponding probability vector based on the sampled value of $\hat{\boldsymbol{\beta}}$, again using Equation (1).

5 MODEL EVALUATION

We evaluate the UQ method of Section 4, and in Section 6, we use these evaluation results to reason about ABS output.

5.1 Clustering Players

Table 2 shows the p-values of two-sample t-tests. For engagements, the p-values show that 2 neighbors is significantly different from 4, 6, and 8 neighbors, while pairs of values among 4, 6, and 8 neighbors are not significantly different. For words, the p-values show that 2 neighbors is significantly different from 4 neighbors, while, again, pairs of values among 4, 6, and 8 neighbors are not significantly different. Though 2 neighbors is not significantly different from 6 and 8 neighbors, respectively, 2 neighbors is significantly different from 4, 6, and 8 neighbors together. This means that we can collect players into two groups: those players with 2 neighbors [group $g = 1$] and those players with 4, 6, and 8 neighbors [group $g = 2$].

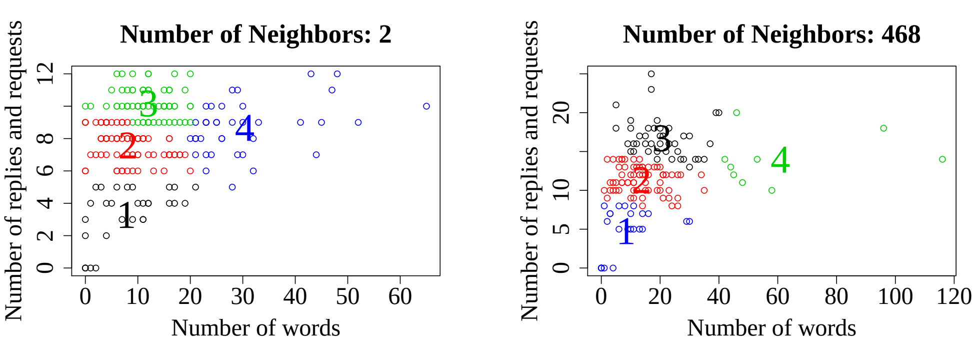
Table 2: Pairwise comparisons of engagements and pairwise comparisons of words. The numbers are p-values of two-sided two-sample t-tests.

Engagements		Words	
number of neighbors	p-value	number of neighbors	p-value
2 vs. 4	8.286e-13	2 vs. 4	2.151e-03
2 vs. 6	8.386e-05	2 vs. 6	0.236
2 vs. 8	1.418e-04	2 vs. 8	0.210
4 vs. 6	0.269	4 vs. 6	0.589
4 vs. 8	0.641	4 vs. 8	0.967
6 vs. 8	0.202	6 vs. 8	0.749
2 vs. 468	5.009e-18	2 vs. 468	3.231e-03

Figures 4a and 4b show the clustering results. The left plot is for 2 neighbors, and the right plot is for 4, 6, and 8 neighbors. Different clusters are marked with different colors and numbers, 1 through 4. In Figure 4a, the black cluster is the least active, and the blue cluster is the most active. In Figure 4b, the blue cluster is the least active, and the green cluster is the most active.

5.2 Quantifying Uncertainty For Model Parameters

Table 3 shows one set of $\mathbf{z}$ values, for the group $k = 2$, and we are using these $\mathbf{z}$ values to generate heterogeneous probability vectors. One can draw a random sample from the confidence region of $\hat{\boldsymbol{\beta}}$ or directly draw a random sample from the distribution of $\hat{\boldsymbol{\beta}}$. Table 4 shows generated probability vectors for group $k = 2$ and clusters 3 and 4 and these values are from direct draw, and these clusters can be found in Figure 4a. Players in cluster 4 have lower to-idle ($a_1 \to a_1$) transition probablity than players in cluster 3, so players in cluster 4 are more active than players in cluster 3. This also can be confirmed in Figures 4a.



(a) Cluster plot for 2 neighbors

(b) Cluster plot for 4, 6, and 8 neighbors

Figure 4: (a) Scatter plot of words against number of replies and requests (engagements) for 2 neighbors. (b) Scatter plot of words against number of replies and requests (engagements) for 4, 6, and 8 neighbors.

Table 3: **z** vector values of $Z_B(t)$, $Z_L(t)$, $Z_W(t)$, and $Z_C(t)$.

Initial state	Number of neighbors	buffer	letter	word	constant
a_1 (idle)	2	0	3	1	5

Table 4: Samples of randomly generated probability vectors using the asymptotic normal distribution, where the initial state is a_1 (idle) and the group is $k = 2$. The left part is for cluster 3, and the right part is for cluster 4. One can see these clusters in Figures 4a. The deviance of the multinomial logistic regression models for cluster 3 and cluster 4 are 12294.71 and 8830.27, respectively. Each row is a probability vector for next actions. The $a_i \to a_j$ means the previous action is a_i and the next action is a_j, so the value under each column represents the probability of next action (e.g., idling, replying with letter, request letter, and form word). The mean probability vector is calulated using Equation (1).

Cluster 3					Cluster 4				
	$a_1 \to a_1$	$a_1 \to a_2$	$a_1 \to a_3$	$a_1 \to a_4$		$a_1 \to a_1$	$a_1 \to a_2$	$a_1 \to a_3$	$a_1 \to a_4$
mean	0.9620	0.0059	0.0127	0.0194	mean	0.8711	0.0235	0.0559	0.0496
sample 1	0.9661	0.0028	0.0113	0.0198	sample 1	0.8731	0.0218	0.0525	0.0526
sample 2	0.9604	0.0045	0.0158	0.0193	sample 2	0.8678	0.0256	0.0602	0.0464
sample 3	0.9624	0.0050	0.0124	0.0202	sample 3	0.8725	0.0229	0.0552	0.0494
sample 4	0.9681	0.0060	0.0097	0.0162	sample 4	0.8606	0.0276	0.0621	0.0497

6 SIMULATIONS OF EGO AGENT WITH NEIGHBORS

We use the models developed in Section 4 to build agent models of GrAG player behavior. In this section, we provide results from ABSs, using data in Section 5 for explanations.

Simulation scenarios. Simulations focus on an ego node and its neighborhood. This enables us to illustrate the value of the UQ approach, and that simulation results can change significantly for different groups and clusters, while using a minimal networked game configuration. This system illustrates the effects of the groups and clusters. We include both groups: (i) group 1 ($g = 1$): four nodes with degrees 1, and (ii) group 2 ($g = 2$): one node with degree 4. See Figure 5. We include cluster 1 for node 0 and clusters 3 and 4 for nodes 1 through 4. For each node, we identify the group and cluster pair by $[g, c]$. The ego node is node 0 (in group 2), with four neighbors (nodes 1 through 4, which are leaf nodes, and are in group 1). Each player is assigned four letters that they can use to form words and to share with their neighbors.

Simulation process. A **simuation** is composed of a collection of simulations instances or iterations. An **iteration** is sequence of simulation steps from time $t = 0$ to t_{max}, such that actions of players (see Table 1) are computed at each t. The state of the system at $t = 0$ constitute the initial conditions for the simulation instance. The process of forming words and sharing letters in an iteration is that shown in Figure 1. In this work, each simulation is comprised of 50 iterations, and results are presented as time point-wise averages over all 50 instances.

Simulation parameters. Simulation parameters are: the configuration of game players as in Figure 5; four letters initially assigned to each player; game duration $t_{max} = 300$ seconds, consistent with the GrAG experiments; and model properties ($\boldsymbol{B}$ matrices) assigned to each player via a pair $[g, c]$ based on node degree for group g, and cluster c.

Simulation results. Figure 6 shows results for node 0, assigned the behavior model according to $[g, c] = [2, 1]$ and nodes 1 through 4, assigned the model behavior for $[g, c] = [1, 3]$. Figure 6a depicts the time histories for forming words for the five nodes; the leaf nodes form fewer words than the hub node 0. All leaf nodes form roughly the same number of words. Figure 6b provides time histories for node 0: numbers of replies received, replies sent, requests received, requests sent, and words formed. Node 0 receives letter requests

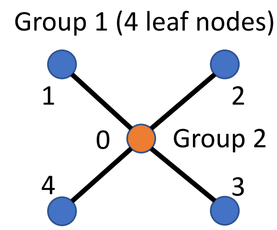
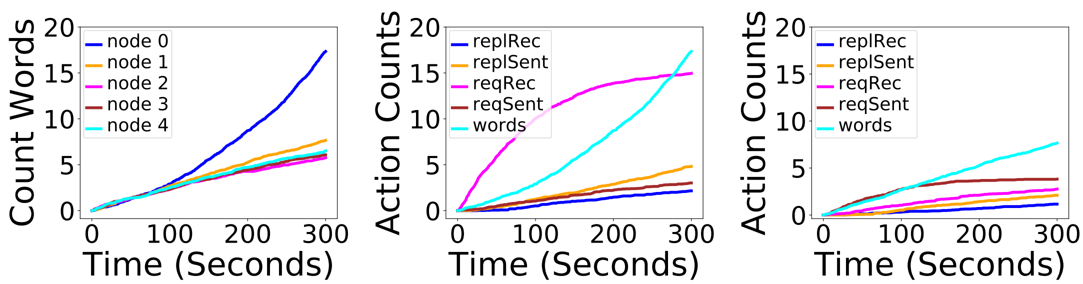


Figure 5: Star network used for simulations. Nodes are game players and edges represent communication channels that can be used to share letters. The center node (hub) has degree $k = 4$ and hence is in group $g = 2$ and the four leaf nodes each have degree $k = 1$ and so are in group $g = 1$.

from all nodes, but only replies to five of the 16 requests made by the leaf nodes. Figure 6c contains the same data for leaf node 1. All nodes form significant numbers of words, consistent with the data in cluster 1 for group 2 and cluster 3 for group 1; see Figure 4.



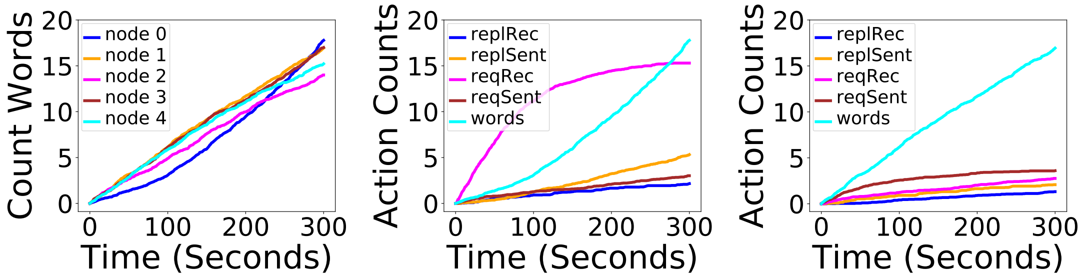
(a) all nodes (b) node 0 (c) node 1

Figure 6: Simulation results for game configuration of a star, with hub node 0 connected to four leaf nodes (nodes 1 through 4). Each player has four initial letters. (a) Number of words formed by each player. (b) Counts of each action for node 0. (c) Counts of each action for node 1. Node 0 has properties $[g,c] = [2,1]$ and all leaf nodes have properties $[g,c] = [1,3]$.

Figure 7 shows the same data as in the previous figure, with one change: the cluster of the leaf nodes changes from 3 to 4. The dominant effect is the increases in numbers of words formed by the leaf nodes. This is consistent with the data in Table 4, showing increases in the probabilities of transition $a_1 \rightarrow a_4$. These results illustrate the changes in game player performance that can be produced in ABMs for different assignments of model properties. In particular, since monetary payouts to game players are proportional to the number of words formed by the team, and all players evenly split total earnings (Cedeno et al. 2020), a comparison of Figures 6a and 7a indicates that the players in Figure 7 would receive more than $2\times$ the earnings as those in Figure 6.

7 CONCLUSION

Our work presents and evaluates a method to quantify uncertainty and build ABMs of human behavior. This work uses as starting points (i) data from group anagram games, and (ii) a mean logistic regression model. Motivation, novelty, and contributions of our uncertainty quantification approach and ABM are provided in Sections 1.1, 1.2, and 1.3. The method works best when a data set can be partitioned along natural parameter dimensions as is the case in this work. Note that the multinomial logistic regression used in this work assumes a linear parametric form, which may not be satisfied in some sophisticated social experiments. However, we can modify the parametric statistical model (i.e., multinomial logistic regression) by some nonparametric statistical model such as generalized Gaussian process. Future work



(a) all nodes (b) node 0 (c) node 1

Figure 7: Simulation results for game configuration of a star, with hub node 0 connected to four leaf nodes (nodes 1 through 4). Each player has four initial letters. (a) Number of words formed by each player. (b) Counts of each action for node 0. (c) Counts of each action for node 1. Node 0 has properties $[g,c] = [2,1]$ and all leaf nodes have properties $[g,c] = [1,4]$.

includes exploring Bayesian approaches for modeling and uncertainty quantification (van de Schoot et al. 2021).

ACKNOWLEDGMENT

We thank the anonymous reviewers for their helpful feedback. We thank our colleagues at NSSAC and Research Computing at The University of Virginia for providing computational resources and technical support. This work has been partially supported by NSF CRISP 2.0 (CMMI Grant 1916670).

REFERENCES

Alam, M., X. Deng et al. 2015. "Sensitivity Analysis of an ENteric Immunity SImulator (ENISI)-Based Model of Immune Responses to Helicobacter Pylori Infection". *PLoS One* 10(9):e0136139.

Baker, A. 2016. "Simplicity". In *The Stanford Encyclopedia of Philosophy* (Winter 2016 ed.)., edited by E. N. Zalta. Metaphysics Research Lab, Stanford University.

Cedeno, V., Z. Hu et al. 2020. "Networked Experiments and Modeling for Producing Collective Identity in a Group of Human Subjects Using an Iterative Abduction Framework". *Soc. Netw. Anal. Min. (SNAM)*:43.

Cedeno-Mieles, V., Y. Ren, S. Ekanayake, B. J. Goode, C. J. Kuhlman, D. Machi, M. V. Marathe, H. S. Mortveit, Z. Hu, X. Deng, N. Ramakrishnan, P. Saraf, N. Self, N. Contractor, J. M. Epstein, and M. W. Macy. 2018. "Pipelines and Their Compositions for Modeling and Analysis of Controlled Online Networked Social Science Experiments". In *Proceedings of the 2018 Winter Simulation Conference*, edited by M. Rabe, A. A. Juan, N. Mustafee, A. Skoogh, S. Jain, and B. Johansson, 774–785. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.: IEEE.

CedenoMieles, V., Z. Hu et al. 2019. "Mechanistic and Data-Driven Agent-Based Models to Explain Human Behavior in Online Networked Group Anagram Games". In *ASONAM*, 357–364.

Fadikar, A., D. Higdon, J. Chen, B. Lewis, S. Venkatramanan, and M. Marathe. 2018. "Calibrating a Stochastic, Agent-Based Model Using Quantile-Based Emulation". *SIAM/ASA J. Uncertainty Quantification* 6(4):16851706.

Hartigan, J. A., and M. A. Wong. 1979. "Algorithm AS 136: A K-Means Clustering Algorithm". *Journal of the Royal Statistical Society. Series C (Applied Statistics)* 28(1):100–108.

House-Peters, L. A., and H. Chang. 2011. "Urban Water Demand Modeling: Review of Concepts, Methods, and Organizing Principles". *Water Resources Research* 47(5):1–15.

Hu, Z., X. Deng, B. J. Goode, N. Ramakrishnan, P. Saraf, N. Self, A. Adiga, G. Korkmaz, C. J. Kuhlman, D. Machi, M. V. Marathe, S. S. Ravi, Y. Ren, V. Cedeno-Mieles, and S. Ekanayake. 2019. "On the Modeling and Agent-Based Simulation of a Cooperative Group Anagram Game". In *Proceedings of the 2019 Winter Simulation Conference*, edited by N. Mustafee, K.-H. Bae, S. Lazarova-Molnar, M. Rabe, C. Szabo, P. Haas, and Y.-J. Son, 169–180. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Lee, S.-J., and E. A. Wentz. 2008. "Applying Bayesian Maximum Entropy to Extrapolating LocalScale Water Consumption in Maricopa County, Arizona". *Water Resources Research* 44(1):1–13.

Li, H., X. Deng, C. Dolloff, and E. Smith. 2016. "Bivariate Functional Data Clustering: Grouping Streams Based on a Varying Coefficient Model of the Stream Water and Air Temperature Relationship". *Environmetrics* 27(1):15–26.

Mason, W., and S. Suri. 2018. "Conducting Behavioral Research on Amazons Mechanical Turk". *Behavior Research Methods*:1–23.

Mason, W., and D. J. Watts. 2012. "Collaborative Learning in Networks". *PNAS* 109(3):765–769.

Papadelis, S., and A. Flamos. 2019. *An Application of Calibration and Uncertainty Quantification Techniques for Agent-Based Models*, 79–95. Cham: Springer International Publishing.

Pearl, J., and D. Mackenzie. 2018. *The Book of Why: The New Science of Cause and Effect*. Basic Books.

Polletta, F., and J. M. Jasper. 2001. "Collective Identity and Social Movements". *Annual Review of Sociology* 27:283–305.

Ren, Y., V. Cedeno-Mieles et al. 2018. "Generative Modeling of Human Behavior and Social Interactions Using Abductive Analysis". In *ASONAM*, 413–420.

Ryzhov, I. O., Q. Zhang, and Y. Chen. 2020. "Advanced Statistical Methods: Inference, Variable Selection, and Experimental Design". In *Proceedings of the 2020 Winter Simulation Conference*, edited by K.-H. Bae, B. Feng, S. Kim, S. Lazarova-Molnar, Z. Zheng, T. Roeder, and R. Thiesing, 1–15. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Sevcikova, H., A. E. Raftery, and P. A. Waddell. 2007. "Assessing Uncertainty in Urban Simulations Using Bayesian Melding". *Transportation Research Part B: Methodological* 41(6):652–669.

Stavrakas, V., S. Papadelis, and A. Flamos. 2019. "An Agent-Based Model to Simulate Technology Adoption Quantifying Behavioural Uncertainty of Consumers". *Applied Energy* 255:113795.

van de Schoot, R., S. Depaoli, R. King et al. 2021. "Bayesian Statistics and Modelling". *Nature Reviews Methods Primers* 1:1–26.

Zhang, H., Y. Vorobeychik, J. Letchford, and K. Lakkaraju. 2016. "Data-Driven Agent-Based Modeling, with Application to Rooftop Solar Adoption". *Autonomous Agents and Multi-Agent Systems* 30(6):1023–1049.

AUTHOR BIOGRAPHIES

ZHIHAO HU is a PhD candidate in the Department of Statistics at Virginia Tech. His email address is huzhihao@vt.edu.

XINWEI DENG is an Associate Professor in the Department of Statistics at Virginia Tech. His email address is xdeng@vt.edu.

CHRIS J. KUHLMAN is a Research Associate Professor in the Biocomplexity Institute & Initiative at the University of Virginia. His email address is hugo3751@gmail.com.