

# Emergence in the U.S. Science, Technology, Engineering, and Mathematics (STEM) workforce: an agent-based model of worker attrition and group size in high-density STEM organizations

Ronald Iammartino<sup>1,2</sup> · John Bischoff<sup>1</sup> · Christopher Willy<sup>1</sup> · Paul Shapiro<sup>1</sup>

Received: 15 December 2015 / Accepted: 4 April 2016 / Published online: 19 April 2016  
© The Author(s) 2016. This article is published with open access at Springerlink.com

**Abstract** The United States government forecasts a shortage of 1,000,000 Science, Technology, Engineering, and Mathematics (STEM) workers over the next 10 years, putting STEM workforce sustainability at risk. The U.S. federal government has launched a range of programs, initiatives, and commissions to address this critical shortage. Past research has shown that organizational group-oriented culture, smaller work team groups, and worker homogeneity relate to lower attrition rates and higher levels of reported worker satisfaction. Applying Complex Adaptive Systems, this study takes a systems approach to examine worker attrition rates and worker group sizes in relation to STEM density across high-density STEM organizations. The base-case organization for this study is the National Aeronautics and Space Administration (NASA) because it maintains the highest density of STEM workers across all U.S. federal organizations and consistently ranks first in teamwork, innovation, and worker satisfaction in the annual Office of Personnel Management (OPM) Federal Employee Viewpoint Survey. The methodology for this study comprises an empirically validated NETLOGO agent-based model of worker attrition using OPM data sets for high-density STEM organizations. Model initialization parameters are represented and validated by historical attrition data for NASA and two control group organizations. The control groups include a lower-limit high-density STEM model and a medium high-density STEM model for validation with historical worker attrition data for the Environmental Protection Agency and Federal

Communications Commission. The findings for this study confirm that STEM worker density is negatively related to attrition rate across all high-density organization types. The model output observations further show the emergence of a negative relationship between organizational STEM density and average worker group size, yet the opposite association is observed for STEM workers. The agent-based modeling approach is an important addition to the current line of academic focused research on STEM workers because it provides a bottom-up insight which helps inform theories and policy effects on ways to mitigate forecasted STEM shortages. Future research could be extended to apply Tipping Point Theory to STEM density and attrition rate variability to better understand STEM threshold ranges across high-density STEM organizations. Future models could also investigate specific STEM and Non-STEM worker characteristics to include age, gender, salary, or education level.

**Keywords** Complex adaptive systems · Emergence · Computational organizational theory · Agent-based modeling

## Introduction

The United States (U.S.) government forecasts a shortage of 1,000,000 Science, Technology, Engineering, and Mathematics (STEM) workers over the next 10 years, putting STEM workforce sustainability at risk [1]. The U.S. federal government has launched a range of programs, initiatives, and commissions to address this critical shortage [1,2]. In 2013, the Office of Personnel Management (OPM) was charged to use the existing body of federal workforce data to understand the STEM workforce life-

✉ Ronald Iammartino  
riammartino@gwu.edu

<sup>1</sup> The George Washington University, Washington, DC 20052, USA

<sup>2</sup> 12499, Sweet Leaf Terrace, Fairfax, VA 22033, USA

cycle, including worker attrition and work environment characteristics [2].

In recent years, a substantial body of systems and agent-based modeling research on worker attrition and group-oriented culture has focused on linkages to organizational performance and employee satisfaction at large [3–8]. These studies have shown that worker homogeneity and smaller work groups generally relate to improved worker performance, satisfaction, and lower attrition rates across all worker-types [6–8]. Worker propensity for attrition has been shown to be most significantly influenced by those most similar in terms of culture, values, and frequency of interaction [6–8].

There are few agent-based studies that have specifically investigated STEM attrition and group-oriented culture within a work organizational environment. STEM studies are largely focused on educational curriculum and academic policy effects. There is further scant evidence of research that has explored STEM organizations across the worker lifecycle. An agent-based modeling approach is an important addition to the current line of research on STEM workers because it provides a bottom-up insight which helps inform theories and policy effects on ways to mitigate forecasted STEM shortages. The existing STEM research on attrition and group-oriented culture is primarily related to macro and meta-level analyses of STEM students [1,3,9]. In academic settings, these studies show that STEM students are more likely to be retained if immersed in top-rated academic programs dedicated singularly to STEM education [1,9]. More frequent interaction with groups of high-quality STEM-oriented students, STEM classes, and STEM teachers have been shown to relate to lower student attrition rates [1,3,9].

The purpose of the present study was to model and investigate the micro-level work-group behaviors related to STEM worker attrition in order to inform theories on how best to address forecasted STEM worker lifecycle shortages at the organizational level. Applying Complex Adaptive Systems (CAS), our study draws from historical data-sets of the NASA base-case organization, and two control-group organizations to validate worker attrition as a basis for analyzing emergent observed grouping behavior in a mixed STEM and Non-STEM organizational model environment. It is proposed that increased STEM density, defined as percentage of STEM workers, relates to lower attrition rates for high-density STEM organizations. The study further posits that STEM density is negatively related to worker group size. Model observations provide insight into emergent grouping properties not easily identifiable through traditional computational analysis [6]. As in past agent-based modeling studies, we conclude with model observations which support future theory generation for mitigating the negative organizational impacts of STEM worker attrition [2].

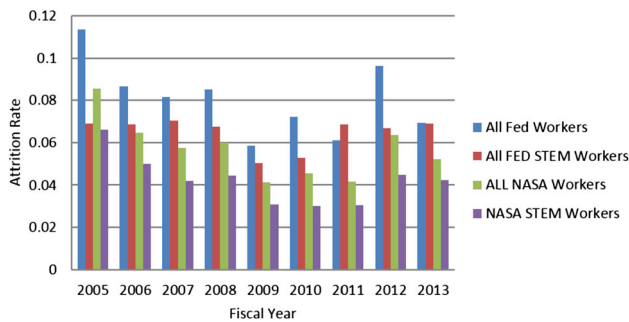
## Conceptual framework

This study defines organizational STEM density based on the percentage of STEM workers relative to all workers within the organizational workforce. Historically, the OPM data show STEM workers comprised no more than 13 % of the total U.S. federal workforce from 2005 to 2014 [1,2]. We define a high-density STEM organization as a large federal organization with greater than 13 % STEM workers within the total organizational workforce. Large federal organizations are identified here based on the OPM classification for large independent federal agencies that have 1000 employees or more.

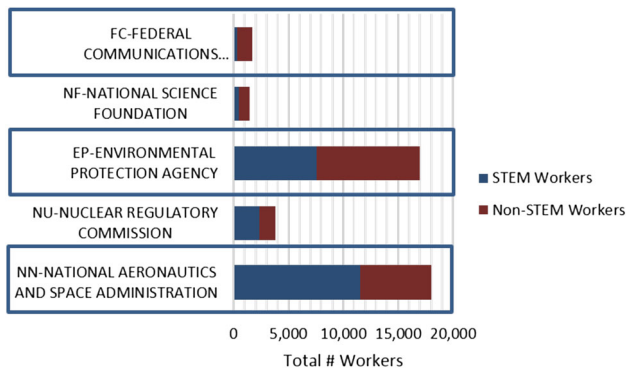
Figure 1 illustrates the historical trends in overall STEM and Non-STEM worker attrition rate differences across the federal workforce and for this study's model base-case high-density STEM organization, NASA. The units for the y-axis, Attrition Rate, is based on the ratio of workers that separate from an organization to workers that remain in an organization during a fiscal year. The OPM definition for separation includes retirement, resignation, termination, transfers, and death. As indicated in these charts, STEM workers historically exhibit lower attrition rates across the entire federal workforce and even less so within the base-case high-density STEM organization, NASA. Figure 2 further shows that Non-STEM workers also exhibit lower attrition rates at NASA relative to the entire federal workforce. This is an important empirical observation, suggesting STEM workers and Non-STEM workers are influenced by their peers within the organizational environment at NASA.

Selection of NASA as the base-case high-density STEM organization for this study was based on the depicted empirical analysis across all large federal organizations for STEM density distribution, worker satisfaction, and group survey data. In addition to being the highest-density large STEM organization in the federal government, the OPM data show that NASA has consistently ranked first amongst all federal organizations in worker-reported survey data on innovation, satisfaction, and teamwork [10]. These factors make NASA an appropriate base-case organization for examining STEM organizations and emergent work-group behavior.

Control group selection was based on variability in terms of STEM density distribution within large independent federal agencies. Figure 2 depicts the total number of STEM and Non-STEM workers for each of the five highest density STEM organizations across all large federal organizations in 2013. The Environmental Protection Agency (EPA) and the Federal Communications Commission (FCC) were selected as model validation control group organizations. The EPA and FCC have historically been ranked third and fifth, respectively, in terms of STEM density across all large federal organizations. Empirical data show approximately a 20 % difference in STEM density across each of the three selected



**Fig. 1** Comparative hist. attrition: all fed workers and NASA



**Fig. 2** Five highest-density stem organizations total stem and non-stem workers (Opm, large federal agencies)

high-density STEM organizations, NASA, the EPA, and the FCC. These three organizations comprise 65 % of all STEM workers in the 21 large federal organizations during the past 10 years.

## Literature review

This study draws upon existing systems-based research in Complex Adaptive Systems, Computational Organizational Theory, Dynamic Social Impact Theory, and Thomas Schelling's Tipping Point Theory. It further draws upon past research on worker performance and worker group sizes.

### Dynamic Social Impact Theory

An extension of Social Capital Theory, Dynamic Social Impact Theory (DSIT) asserts that work groups are formed based on four elements of self-organization: clustering, correlation, consolidation, and continuing diversity [3, 11]. Groups are represented as complex systems that continuously adapt and change based on differing local environmental conditions [3]. Prominent to this theoretical framework is that the strength of group influence is proportional to the proximity, or immediacy, of the main sources of influence [11]. The proximity of the main sources of influence on a given subject or agent is determined by the number of influences

and the power of those influences based on factors such as frequency of interaction and similarity in age, education, and gender [3, 11].

There is a limited body of DSIT-related work on organizational-level agent-based models. Shapiro [6] found that general federal worker attrition is related to the social influences of organizational culture and age-groups dispersion across the career lifecycle. Rouchier et al. [12] found that worker similarities in individual values and social influence proximity mitigate the effects of worker attrition on information promulgation, thus having an impact on knowledge transfer. Along these lines, conventional research has showed that increased technical knowledge worker participation on project teams is associated with better overall organizational performance [3–22]. There is limited evidence of research using DSIT for STEM worker density within an organizational environment.

Applying DSIT, our first hypothesis posits that STEM worker density relates to lower overall organizational attrition. We assert that STEM worker density relates to total organizational-level workforce attrition decision-making, which is in line with DSIT, historical data analysis on individual STEM and Non-STEM attrition, and research on the relationship between technical worker density and team-level performance.

*H1o:* There is a positive or neutral relationship between organizational STEM worker density and worker attrition in high-density STEM organizations.

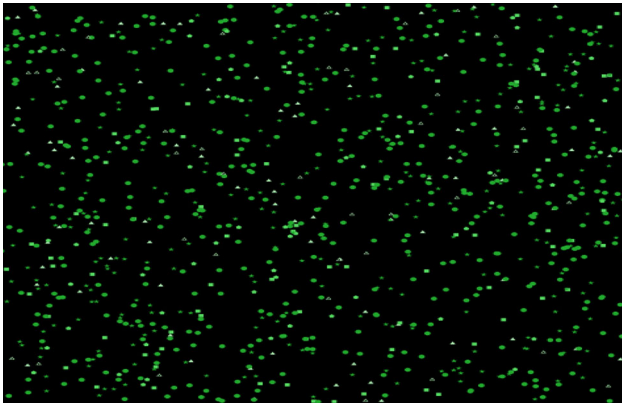
*H1a:* There is a negative relationship between organizational STEM worker density and worker attrition in high-density STEM organizations.

### Emergence

Rooted in systems theory, emergence research spans a wide-spectrum of both natural and social system environments [13]. Emergence refers to simple self-organizing agents that interact at the local micro-level to produce unpredictable patterns of effect or phenomena at the macro-level [13–15]. In general, emergence occurs when the sum of the system's parts or micro agent interactions produce an irreducible whole-system outcome [13–15].

Emergence within organizations is broadly defined by the complex interaction effects of individual human agents [15]. Unlike in natural environments, human agents are able to communicate, learn, and change at varying degrees within an organization. These agents are also able to impact, and be impacted by, both local and environmental dynamic forces that are in an indeterminable state of changing conditions [10, 13–15].

In 2010, Allen and Davis built an agent-based model to analyze the effects of emergence in STEM students. They



**Fig. 3** Netlogo Agent Interface (represents initial agent randomized dispersion)

found that better teacher quality during earlier STEM student years and more specific STEM topic segregation related to lower overall attrition and better STEM career workforce placements. Additionally, conventional STEM workforce studies show that STEM workers prefer a more decentralized work environment characterized by lesser supervisor interaction, greater autonomy, and more task-independence than Non-STEM workers [12, 22]. To evaluate the model behavior associated with this finding, our second hypothesis is as follows:

*H2o: There is a positive or neutral relationship between organizational STEM worker density and emergent model worker group size average in high-density STEM organizations.*

*H2a: There is a negative relationship between organizational STEM worker density and emergent model worker group size average in high-density STEM organizations.*

### Agent-based modeling with NETLOGO

The model environment is established and executed using NETLOGO agent-based modeling (ABM) software [13]. NETLOGO is an open-source ABM modeling tool appropriate for analyzing emergent phenomena in complex systems [6]. Global environmental conditions for time and distance from other agents are defined using the NETLOGO standard protocols of “ticks” and “patch radius.” “Patch radius” represents the two-dimensional distance within the model space between agents. As shown in Fig. 3, the dispersion of agents within the model space is randomized upon initialization.

### Data

OPM maintains open-source historical employment data for the past 17 years across all large independent federal organizations and agencies. Since 2005, these data-sets

have included STEM worker employment trend data. In the present model, the data-sets for 2005–2014 are evaluated to establish initialization modeling parameters and validation criteria (satisfaction threshold, attrition propensity, age range, and STEM or non-STEM occupation code). As noted in the introduction section, the agent-based model initialization population is selected based on STEM worker density reported in OPM data sets. STEM workers are categorized in accordance with OPM occupational codes for all Science, Technology, Engineering, and Mathematics workers. For verification and validation purposes, attrition data collection was also conducted for two additional high-density STEM organizations, the EPA and FCC.

### Methodology

The Overview, Design Concepts, and Details (ODD) protocol [10] follows the widely accepted structure used to describe the agent-based model (ABM) for STEM worker and Non-STEM worker attrition and group-size in high-density STEM organizations [6, 16].

#### Overview: purpose

The purpose is to propose and illustrate hypotheses on STEM organizational systems and associated emergent phenomena. Both the base-case high-density STEM organization, NASA, and control group organizations, the EPA and FCC, are represented to investigate attrition rate and emergent group-sizes across a distributed scale of organizational STEM density. Using the base-case high-density STEM organizational model, two additional variable high-density STEM organizational models are tested to validate and assess lower and upper-limit STEM density thresholds for attrition and observed group sizes during a 10-year period, 2005–2014.

#### Overview: entities, state variables, and scales

Adapted from Shapiro [6] U.S. Federal Workforce Departure Model, the present ABM includes STEM and Non-STEM workers represented as agents within an agent-space environment of a large U.S. government organization. Agents are defined by breed type, STEM or Non-STEM, and by the satisfaction or STEM satisfaction variable on a scale of 0 to 100. Starting satisfaction levels are assigned based on 2005–2014 NASA historical attrition rates in each of the 11 OPM age group categories. The agents are further categorized by age-group and distance to other STEM or Non-STEM agents. Distance is measured based on the NETLOGO model protocol with a starting patch radius of one agent-worker per ten model patches. Agents are set to randomly move two patches per time “tick” representing a week. The criterion for two-



dimensional distance and agent movement amongst workers is an assumption based on the precedent set by the Shapiro, Mazzuchi, and Sarkani model. It does not directly reflect real-life distance. As workers depart the model, new workers are randomly introduced to the agent environment based on NASA 2005 STEM and Non-STEM age-group worker replacement rates [6].

### Overview: process and scheduling

The model sets empirically defined rule-based thresholds for agent satisfaction at the individual and agent-group level to simulate worker attrition and group size starting behavior. The objective for agents is to seek and find an agent breed group that is at or above their individual satisfaction threshold level.

Individual STEM and Non-STEM agents have defined satisfaction threshold levels based on NASA 2005 attrition data. Agents that seek and find a local group at or above their assigned satisfaction level remain in the model organization; agents that fail to find a local agent group at or above their assigned satisfaction level continue to seek local group acceptance until reaching the global environment satisfaction decay limit of zero. Model run results are aggregated and validated by the historical empirical data-sets.

### Design concepts

The basic concept examined in ABM is the seeking of an acceptable shared work environment for both STEM and Non-STEM workers based on individual worker propensity for attrition and work group selection [6]. These design concepts are general in nature and can be applied across a wide-variety of future ABM settings and environments. Key design concepts for this study include the following:

- *Emergence* The main model results that are expected to vary in unpredictable ways include micro-behaviors associated with group size and attrition rate relative to STEM or Non-STEM workers.
- *Adaptive behavior* Agents adapt to their environment based on distance to other STEM or Non-STEM workers and relative satisfaction levels. Agents migrate across the model space seeking agent groups with satisfaction threshold level averages above their individual agent satisfaction threshold level. The agent migration rate is assumed for agents seeking an acceptable agent-group at two patches per time “tick.”
- *Objective seeking* Workers continue to seek acceptable groups at a prescribed decay rate until reaching the global decay limit of zero at which time they exit the model organization. Seeking agents have a global decay rate of

0.075 per time “tick” and a prescribed age-group decay rate factor. Agents that find an acceptable group prior to reaching the global decay limit of zero have a global satisfaction growth rate of 1.00 per time “tick” and a prescribed age-group growth rate factor. Both decay and growth rate factors for all age-groups are listed in Table 1. It is important to note that decay and growth rates and factors are assumptions based on model calibration.

- *Sensing* Agents sense proximity to other STEM or Non-STEM agents based on agent migration across the model space. Agents sense other agents or groups of agents at an acceptable average satisfaction level using a pre-defined sensing satisfaction radius. This satisfaction radius is assigned based on like-agent organizational density.
- *Collectives* The model assumes that STEM and Non-STEM workers must only find other acceptable STEM or Non-STEM worker groups to remain in the environment organization.

### Details: initialization and input data

Table 1 depicts the initialization parameters for a 1000-agent model, the NASA Workforce in 2005 [6]. All initialization parameters in Table 1 are based on OPM historical data-sets and structured in accordance with the variables established in the Shapiro, Mazzuchi, and Sarkani model [6]. The model organizational workforce is distributed into two job classifications, STEM and Non-STEM. Total starting population distribution is represented by the starting STEM or Non-STEM distribution at NASA in 2005. Each of the two job classifications has defined satisfaction threshold levels in three segments: satisfied, committed, and unsatisfied; these segments are based on the NASA 2005 OPM data set for age-group attrition rates. The starting satisfied segment includes workers in age-group 30–54 years old; starting committed segment includes workers in age-group 25–29 and 50–54 years old; starting unsatisfied segment includes workers in age-group 24 and under, 55–59, and 60 and over. New worker probability is based on the NASA 2005 OPM data set distribution of new workers. These new workers enter the model environment based on the probability of STEM or Non-STEM worker replacement at NASA in 2005 using OPM data-sets. STEM and Non-STEM age-group starting satisfaction and minimum satisfaction thresholds are uniformly assigned to all same job classification agents within the satisfied, committed, or unsatisfied segments; this is an abstraction based on assumptions from the NASA 2005 OPM data set and model calibration for approximating historical attrition rates [6].

Global variables include worker satisfaction radius, satisfaction growth rate, and satisfaction decay rate. The constant satisfaction radius is set to three patches for both STEM and

**Table 1** Initialization parameters base-case model, NASA

	Undecided	Undecided	Committed	Satisfied	Satisfied	Satisfied	Satisfied	Satisfied	Satisfied	Committed	Undecided	Undecided
Initialization parameters	20 and under	20–24	25–29	30–34	35–39	40–44	45–49	50–54	55–59	60–64	65 and older	
Population (NonSTEM)	1	17	17	21	40	62	71	82	60	25	8	
Population (STEM)	0	10	26	29	74	139	125	79	55	36	23	
Starting satisfaction (NonSTEM)	75 %	75 %	83 %	95 %	95 %	95 %	95 %	95 %	83 %	75 %	75 %	
Starting satisfaction (STEM)	74 %	74 %	90 %	96 %	96 %	96 %	96 %	96 %	90 %	74 %	74 %	
Min satisfaction threshold (NonSTEM)	87 %	87 %	78 %	71 %	71 %	71 %	73 %	74 %	78 %	84 %	84 %	
Min satisfaction threshold (STEM)	90 %	90 %	90 %	70 %	70 %	71 %	71 %	70 %	84 %	87 %	87 %	
Growth(+)/decay factor(-) (NonSTEM)	2.05/1.2	2.05/1.3	84.7/13	94.8/10	95/10	95/7.5	95/7	94.8/6	3.2/100	0.40/100	0.40/100	
Growth(+)/decay factor(-) (STEM)	2/16	2/16	83.7/14	91.9/10	94.8/10	94.2/10	92/7.5	92/7.5	2/100	0.35/100	0.35/100	
New worker probability (NonSTEM)	0.05	0.50	0.65	0.72	0.80	0.86	0.93	0.97	1.00	1.00	–	
New worker probability (STEM)	0.00	0.17	0.34	0.54	0.65	0.80	0.91	0.95	0.98	0.98	1.00	

Non-STEM agents. For the base-case model and the medium high-density STEM model, the satisfaction radius assumes both STEM and Non-STEM agents have approximately the same distance, or satisfaction radius, within the organization to interact with same-breed agents. Agents can group with other agents within three patches of their current location.

Based on DSIT, the satisfaction radius for the lower-limit high-density STEM model is modified to reflect the greater distance of other STEM agents in the model space; STEM agents have a one-patch satisfaction radius and Non-STEM agents retain a three-patch satisfaction radius. The modified satisfaction radius supports the theory that the influence of like-agent STEM workers will be lesser in the more sparsely populated lower-limit STEM organizational environment [3, 11].

Satisfaction growth and decay rates are globally set based on job classification, STEM or Non-STEM [6]. Satisfaction growth and decay factors are globally set based on STEM or non-STEM age-group.

**Base high-density model (NASA: 59 % STEM)**

Scaled to the OPM NASA 59 % STEM-density data set for parameter-setting year 2005, the 1000-agent model is set to run 100 times to assess attrition and group-size self-organization. STEM and Non-STEM agent attrition data are analyzed and compared to OPM data on NASA average attrition between 2005 and 2014.

The primary aim for model validation is to ensure output results appropriately align with OPM historical attrition data findings. A two-sample *t* test and confidence interval testing is conducted for comparative analysis of base-case model attrition with OPM NASA data-set averages, 2005–2014. The *t* test compares the mean of 100 runs of the base-case model attrition rate with the mean attrition rate in the OPM NASA data-set, 2005–2014.

For H1o, the model and empirical data are fitted for linear regression analysis to test the across high-density STEM organization relationship between STEM density and attrition rate. For H2o, model data are fitted for linear regression analysis to test the across high-density STEM organization relationship between model STEM density and observed model average work group size.

**Medium high-density model (EPA: 43 % STEM)**

Scaled to OPM Environmental Protection Agency (EPA) 43 % STEM density for parameter-setting year 2005, the 1000-agent model is set to run 100 times to assess attrition and group-size self-organization based on individual agent interaction. Initialization parameters for satisfied, committed, and unsatisfied workers are adjusted at the macro-level to align with the EPA historical STEM and Non-STEM

distribution; specific initialization parameters for age-group and starting satisfaction rates remain aligned with the initial NASA data set. STEM and Non-STEM agent attrition and group size averages are analyzed.

For model validation, a two-sample *t* test and confidence interval testing are conducted for comparative analysis of medium high-density model attrition to OPM EPA data-set averages, 2005–2014. The *t* test compares the mean of 100 runs of the medium high-density model attrition rate with the mean attrition rate in the OPM EPA data-set, 2005–2014.

**Lower-limit high-density model (FCC: 19 % STEM)**

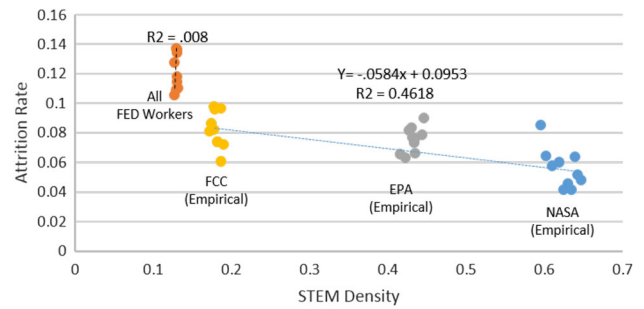
Scaled to OPM Federal Communications Commission (FCC) 19 % STEM density for parameter-setting year 2005, the 1000-agent model is set to run 100 times to assess attrition and group-size self-organization based on individual agent interaction. Initialization parameters for satisfied, committed, and unsatisfied workers are adjusted at the macro-level to align with the FCC historical STEM and Non-STEM distribution; specific initialization parameters for age-group and starting satisfaction rates remain aligned with the initial NASA data set.

For model validation, a two-sample *t* test and confidence interval testing are conducted for comparative analysis of lower-limit model attrition to OPM FCC data-set averages, 2005–2014. The *t*-test compares the mean of 100 runs of the lower-limit model attrition rate with the mean attrition rate in the OPM FCC data-set, 2005–2014.

**Results**

**Model verification**

For model verification, we use statistical methods to first analyze real-world phenomena as a basis for validation with simulated model results. This approach aligns with foundational principles outlined in Carley’s [17] seminal work on using simulation for generating systems-based hypotheses. Figure 4 shows the linear regression output for the OPM historical data sets for both the entire federal workforce and across each of the three high-density STEM organizations. The units for the *y* axis, Attrition Rate, is based on the ratio of workers that separate from an organization to workers that remain in an organization during a fiscal year. The units for the *x* axis, STEM density, is based on the ratio of STEM workers to all workers within an organization. The differences in the coefficient of determination,  $R^2$ , confirms that high-density organizations are distinguishable from the U.S. federal workforce at large in terms of worker attrition rate. Combined with Figs. 1 and 2, this supports verification of the model’s conceptual framework as a basis for hypothesis test-



**Fig. 4** Historical stem density and organizational attrition

**Table 2** Model validation method #1: descriptive statistics (hist and model attrition)

<i>t</i> test: two-sample independent means (unequal variance)		
	NASA empirical	NASA model
Mean	0.05594	0.05755
Variance	8.22E–05	1.73E–05
Observations	10	100
$P(T \leq t)$ two-tail	0.3476**	
<i>t</i> critical two-tail	2.306	
	FCC empirical	Lower limit model
Mean	0.08213	0.08388
Variance	0.000142	1.42E–05
Observations	10	100
$P(T \leq t)$ two-tail	0.6654**	
<i>t</i> critical two-tail	2.2621	
	EPA empirical	Median high model
Mean	0.076545	0.07944
Variance	7.36E–05	1.28E–05
Observations	10	100
$P(T \leq t)$ two-tail	0.1855**	
<i>t</i> critical two-tail	2.2621	

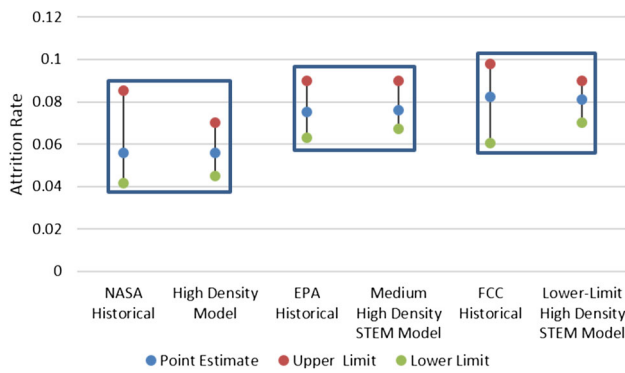
\*\* Fail to reject null hypotheses (no significant differences between model and empirical data)

ing. There is no evidence of prior research that has examined the micro-behaviors specifically exhibited by high-density STEM organizations workers.

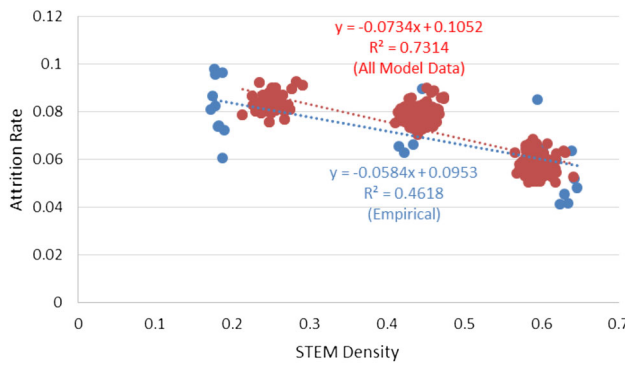
**Model validation**

The results of running each of the three model variations 100 times produced average attrition rates that were validated to have no significant statistical differences with the historical averages for NASA, the EPA, and FCC between 2005–2014.

Table 2 shows the two-sample independent means *t*-test descriptive statistics and *p* values for each of these data for model validation. The *p* values for each set of data are greater



**Fig. 5** Model validation method #2: confidence interval testing



**Fig. 6** Linear regression: all three high density organizations (empirical and model)

than 0.05; this indicates the null hypothesis cannot be rejected at the 95 % confidence level. There is no significant difference between the average attrition rate in the OPM historical data set and the model data output for any of the three model types.

Figure 5 illustrates the confidence interval method results for comparative model validation analysis. These results show that each of the model data output attrition rates have averages and ranges that fall within the confidence interval range for the historical data sets. This second method further validates the acceptability of these model results for simulating the OPM data.

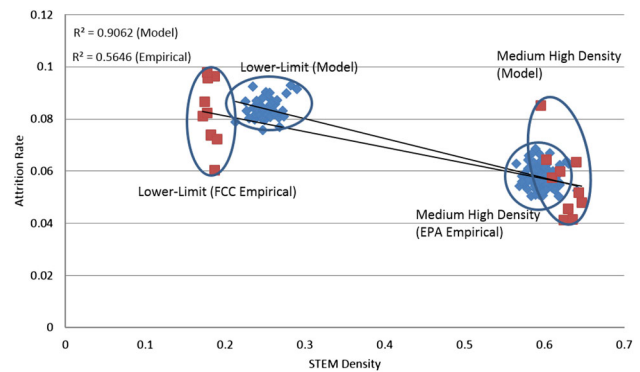
**III. Hypothesis testing: (H1o)** There is positive or neutral relationship between organizational STEM worker density and worker attrition in high-density STEM organizations.

Figure 6 depicts the linear regression results for hypothesis testing of H1o using all three sets of empirical and model organizational data. In Table 3, the *p* value for each of the regression coefficients is less than 0.05; this indicates the null hypothesis can be rejected at the 95 % confidence level. For both the empirical and model data, there is a statistically significant negative relationship between the average attrition rate and STEM density. This is consistent with the model verification linear regression analysis for empirical attrition data in Fig. 4. Figures 7, 8 and 9 illustrate differences in correlation across organizational and model variation types. These

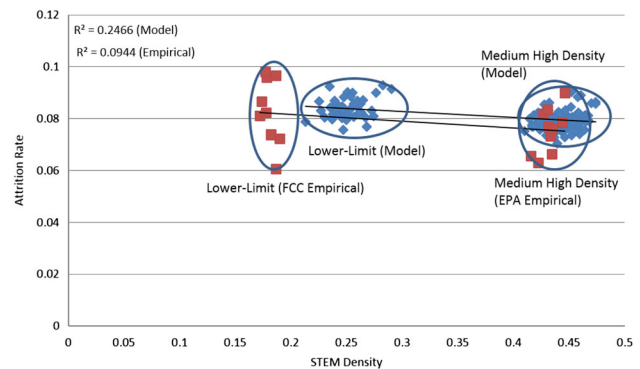
**Table 3** Hypothesis testing: H1o: regression stat. (all three ORG. and models)

	Empirical	Model
Regression statistics		
Multiple R	0.67957	0.85522
R square	0.46182	0.73140
Adjusted R square	0.44260	0.73050
Standard error	0.01189	0.00630
Observations	30.00000	300.00000
Intercept	0.09527	0.10521
X variable	-0.05844	-0.07335
<i>t</i> Stat	-4.901	-28.486
<i>P</i> value	0.00003627*	4.66E-87*

\* Reject null hypothesis (H0) (there is a statistically significant negative relationship between STEM density and attrition)



**Fig. 7** Linear regression: lower-limit high density to NASA base case (empirical and model)

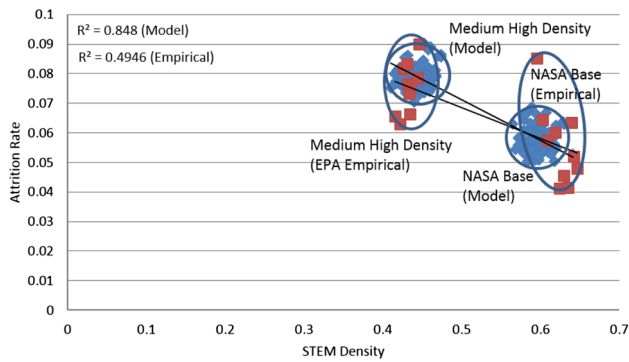


**Fig. 8** Linear regression: lower-limit high density to medium high density (empirical and model)

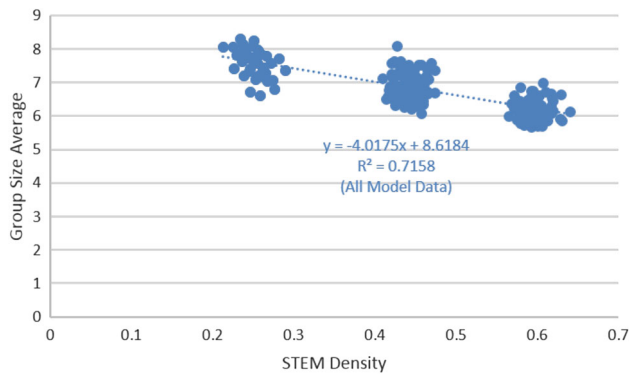
findings support the conclusion to reject the null hypothesis, H1o.

**IV. Hypothesis testing (H2o):** There is a positive or neutral relationship between organizational STEM worker density and observed model worker group size average in high-density STEM organizations.





**Fig. 9** Linear regression: medium high density to NASA base case (empirical and model)



**Fig. 10** Linear regression: group size average (model observation data)

Figure 10 depicts the linear regression results for hypothesis testing of H2o using all three sets of model organizational data. It is important to note that these data are singularly derivative of emergent model output observations; there is no empirical data set to validate group size results. In Table 4, the *p* value for the regression coefficients is less than 0.05; this indicates that the null hypothesis can be rejected at the 95 % confidence level. The linear regression results confirm a negative relationship between model organizational STEM worker density and observed model worker group size. These results support the conclusion to reject the null hypothesis, H2o.

## Discussion

### Model run observations: attrition

The figure below represents the model output for a single model run of each of the three model variations, NASA (Base Model), EPA (Medium Density Model), FCC (Low-Density Model). It is important to note that a single model run is not wholly representative of all model runs in the data set due to the stochasticity of the agent-based approach [6]. This output is intended to illustrate comparative trends across each of the model variations (Figs. 11, 12, 13).

**Table 4** Hypothesis testing: H2o: regression stat. (all model data)

	Model
Regression statistics	
Multiple R	0.84608
R square	0.71584
Adjusted R square	0.71489
Standard error	0.35865
Observations	300
Intercept	8.61844
X variable	−4.01749
<i>t</i> Stat	−27.399
<i>p</i> value	2.07E−83*

\* Reject null hypothesis (H2o) (there is a statistically significant negative relationship between STEM density and observed group size average)

As illustrated, the model output for a single model run of each of the three model variations shows trends across each of the three tested environments. For Non-STEM workers, attrition remains stable latest in the representative 10-year model lifecycle for the low-density model, which lags the early lifecycle peak in STEM attrition for low-density model STEM workers. Together with the medium- and high-density models, the behavior of Non-STEMs in the low-density model suggests that environmental conditions and worker STEM or Non-STEM percentages have the most significant impact on Non-STEM attrition early in model lifecycle. Though the low-density model shows peak attrition for Non-STEM workers latest in the lifecycle, all three models show stable attrition past year 5 of the model.

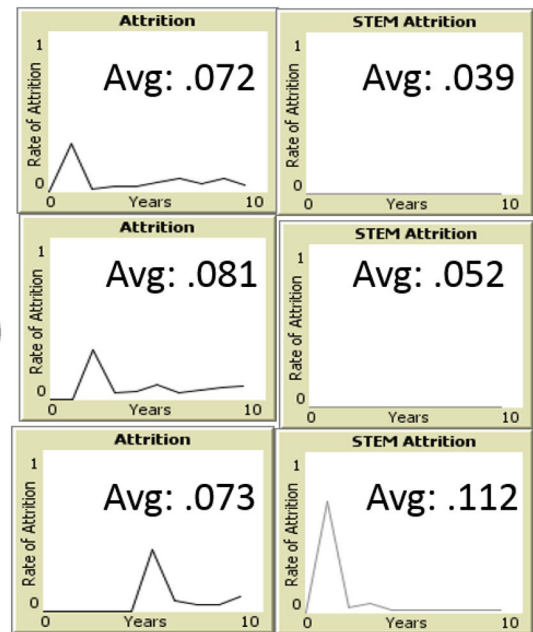
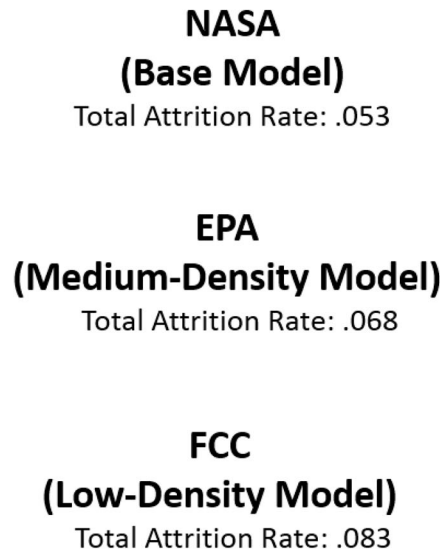
Following initial surges in attrition, Non-STEM attrition for both the high-density and medium-density models remain generally stable through the entire lifecycle. Non-STEM workers appear to experience an initial period of onboarding and socialization followed by stable, low attrition, suggesting peer influence effects of the high organizational commitment STEM worker cohorts.

### Model run observations: group size

The figure below shows the group size average output for a model run of each of the three model variations. It is important to note that no single model run is wholly representative of all model runs due to the stochasticity of agent-based modeling.

As illustrated, there are three significant emergent properties in observing how model groups are formed through the lifecycle. First, the model output shows a much more stable STEM worker grouping behavior for all STEM density organizations throughout the 10-year lifecycle. This could be attributed to the more clearly defined roles and tasks assigned to STEM workers relative to a more diverse Non-STEM

**Fig. 11** Comparative attrition rates: (model observation data)



**Fig. 12** Comparative group size averages: (model observation data)

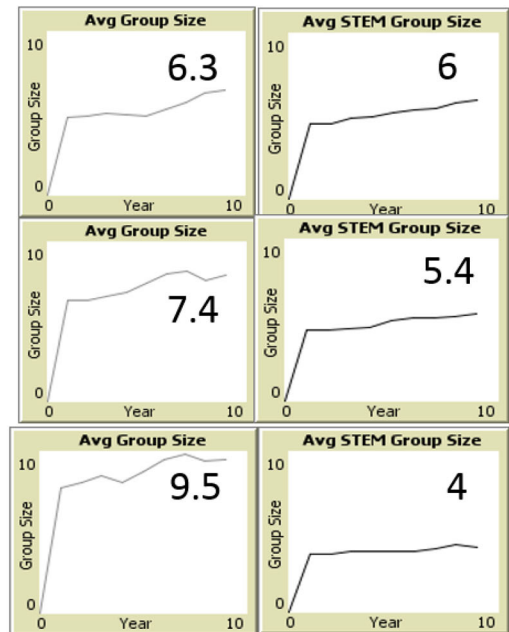
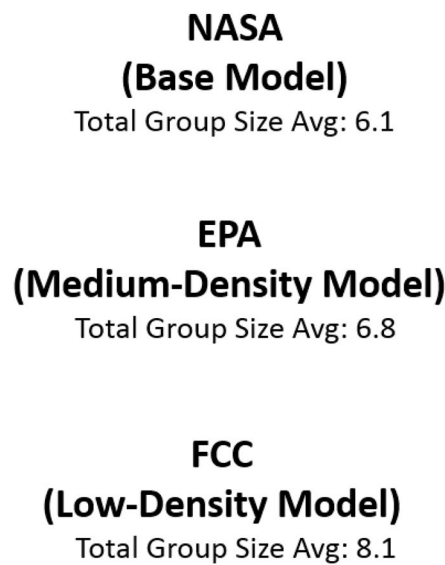


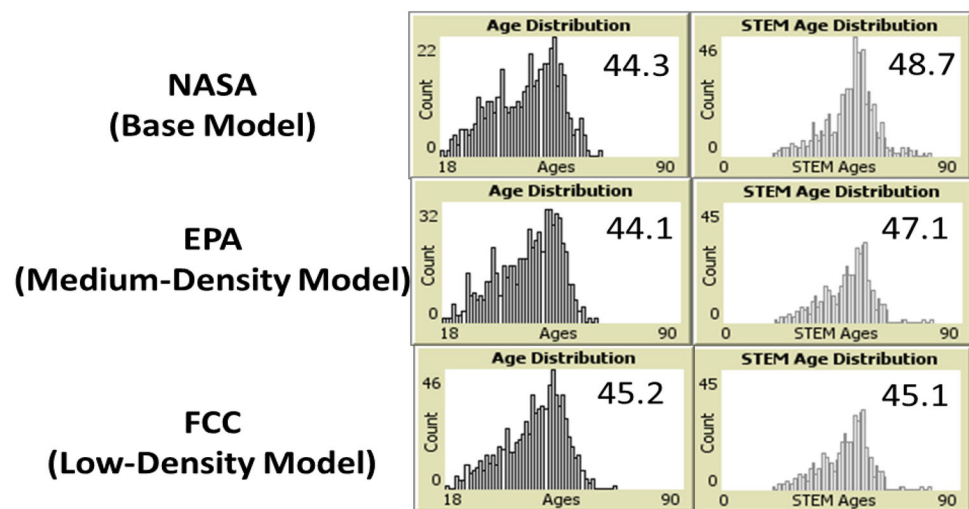
FIGURE 12

workforce. It might also be attributable to the on-average smaller STEM work groups relative to the on-average larger Non-STEM work groups.

Second, STEM worker group size averages increases as organizational STEM density increases, while the opposite effect is observed for both Non-STEM workers and total organizational group size averages. This indicates that STEM worker grouping behavior is opposite the relationship posited in hypothesis 2. This could be attributed to the peer-influence of STEM workers on Non-STEM workers as they interact more or less frequently based on

overall STEM density within each of the model organizations. It further suggests that high-density organizations are better suited to STEM worker preferred worker style, characterized by greater autonomy and stability. Also, a noticeably larger group size increase for all three category types (STEM, Non-STEM, and overall) is observed in the difference between the medium- and low-density STEM organizational models; this suggests the existence of a tipping point where change in group size occurs more rapidly. Further research would be needed to confirm and validate this observation.

**Fig. 13** Comparative average age: (model observation data)



### Model run observations: age distribution

The figure below shows the age distribution output for a model run of each of the three model variations.

Observing the model age distribution effects, the high-density and medium-density models show higher average STEM worker ages within the organization, suggesting the presence of a more experienced STEM workforce relative to the low-density model. These findings suggest a relationship between STEM density and average STEM age; thus future hypotheses could be applied to confirm if greater percentages of STEM workers are associated with longer STEM worker tenure and associated organizational benefits such as group performance or workforce knowledge transfer. This model observation is significant because it has direct implications on a major policy concern outlined in the literature review that showed STEM workers depart the workforce earlier than Non-STEM workers, thus compounding a technical worker shortage attributed to baby-boom generation retirements.

In addition, average age observations show lesser sensitivity to STEM density for Non-STEM workers (as expected, the effect of STEM density on Non-STEM age is in the opposite direction). In contrast to STEM workers, the more limited effects of STEM worker density on Non-STEM worker age distribution could be an indicator that Non-STEM workers are influenced more by diverse work environment factors than peer type, such as organizational pay, mission, or management structure.

### Conclusion

This study contributes to a body of STEM shortage research primarily focused on education policies and academic-setting curriculum changes. Focusing on the existing workforce, our model provides computational evidence for hypotheses that

STEM worker micro-behaviors can influence organizational attrition and grouping behavior. The findings of this study confirm that there is a negative relationship between STEM density and attrition rate in high-density STEM organizations.

The model output also conforms to theories on group-oriented culture; the observed model output exhibits a negative relationship between STEM density and total worker group sizes in high-density STEM organizations. STEM workers are observed to establish and maintain smaller work group sizes than Non-STEM workers in all three models. This is consistent with past research that indicates smaller group sizes are associated with reported STEM worker satisfaction, innovation, and teamwork.

The differences in observed directional effects of STEM density on STEM and Non-STEM workers is a main contribution of this model. Notably, there is an observed positive relationship between STEM density and STEM worker group size. This contrasts with the broader effects identified in the mixed STEM and Non-STEM group size deviations across models. Future hypotheses should explore the positive relationship between STEM group size and STEM density as it relates to the emergence of a critical threshold group size range for attrition and age. For age distribution, model observations show that STEM worker average age increases with STEM density. High- and Medium-Density model results for attrition and group sizes were overall most similar and distinguishable from the Low-density model, suggesting the basis for a critical threshold range in future STEM density hypotheses.

These model observations have implications for theory generation for mitigating STEM shortage risks. Specific theories generated from this study for further investigation include: (1) STEM workers demonstrate a more stable, lower attrition rate in high-density STEM environments above 40% STEM; (2) STEM workers prefer larger work group

sizes within high-density STEM organizations; (3) in contrast to the total workforce, there is a positive relationship between STEM worker attrition and group-size; (4) STEM worker average age is higher in high-density organizations; (5) STEM worker attrition, group size, and average age are most stable and similar in high-density STEM organizations comprised of greater than 40 % STEM workers.

This study was limited to three large U.S. federal agencies and the data-sets available during a 10-year timeframe only. Specific economic conditions during 2005–2014, to include a major economic downturn, might limit the applicability of these findings to past and future time periods. The study measurement for attrition was also limited in scope to those that departed the federal workforce, thus not accounting for workers that switched between agencies or STEM occupation types. The three organizations analyzed in this study differ on many levels beyond STEM density. Model parameters for this study also did not include important variables that contribute to worker attrition, such as organization locations, worker salaries, and the centrality of STEM workers to overall organizational mission.

Finally, the ABM methodology applied in this study could be used in a wide range of work environments and job classification types across disciplines and industries. Future research on the STEM workforce could be extended to examine the optimal organizational design, systems, and models for sustaining critical attrition rates and group sizes within high-density STEM organizations. Further analysis might be conducted to apply Tipping Point Theory to STEM density and attrition rate variability to better understand STEM threshold ranges across high-density STEM organizations. Access to supervisory-level data could be investigated as a means to validate observed model work group size findings. Future models might also investigate STEM organizational and worker characteristics, to include gender, education level, and income.

**Open Access** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

## References

- Holdren JP, Cora M, Suresh, S et al (2013) Federal STEM education, 5-year strategic plan: a report from the Committee on STEM Education. In: National Science Technology Council. Office of the President of the United States, pp 1–30
- OPM Hosts STEM Data Jam (2013) <http://www.opm.gov/blogs/Director/2014/1/28/OPM-Co-Hosts-Data-Jam/>. Accessed 15th Oct 2014
- Allen TT, Davis N (2010) A simple agent-based social impact theory model of student STEM selection. In: Proceedings of the Winter 2010 IEEE simulation conference, pp 278–289
- Vakili G, Tabatabaee F, Khorsandi S (2013) Emergence of cooperation in peer-to-peer systems: a complex adaptive systems approach. *J Syst Eng* 16(1):213–223
- Mohr DC, Young GJ, Burgess JF (2012) Employee turnover and organizational performance: the moderating effect of group-oriented organizational culture. *Human Res Manag J* 22(2):217–233
- Shapiro P, Mazzuchi TA, Sarkani S (2013) Emergent behavior of the US government workforce: an agent-based model of worker loyalty. *Rev Manag Innov Creat* 5(17):100–111
- Wheelen SA (2009) Group size, group development, and group productivity. *Small Group Res* 40(2):247–262
- Glock CH, Jaber MY (2014) A group learning curve model with and without turnover. *J Model Manag* 9(2):179–199
- Chen X, Soldner M (2013) STEM attrition: college students' paths into and out of STEM fields. Statistical analysis report. U.S. Department of Education, USA
- Best Places to Work in the Federal Government. Partnership for public service. <http://www.bestplacestowork.org>
- Latane B (1996) Dynamic social impact. In: Modeling and simulation in the social sciences from the philosophy of science point of view, pp 287–310
- Rouchier J, Tubaro P, Emery C (2012) Opinion transmission in organizations: an agent-based modeling approach. *Comput Math Organ Theory* 20(2):252–272
- Haghnevis M, Askin RG (2012) Modeling framework for engineered complex adaptive systems. *IEEE Syst J* 6(3):520–530
- Anderson P (1999) Complexity theory and organization science. *Organ Sci* 10(3):216–232
- Golds C, Kay R (2010) Emergence in organizations: the reflexive turn. *Complex Organ* 12(3):47–63
- Alaa G (2009) Derivation of factors facilitating organizational emergence based on complex adaptive systems and social autopoiesis theories. *Emerg Complex Organ* 11(1):19–34
- Carley KM (1999) On generating hypotheses using computer simulations. *Syst Eng* 2(2):69–77
- Grimm V, Berger U, DeAngelis DL et al (2010) The ODD protocol: a review and first update. *Ecol Model* 221:2760–2768
- Wilensky U (1999) NETLOGO. <http://ccl.northwestern.edu/netlogo>. Accessed 15th Oct 2014
- Hofaidhllaoui M, Chhinzner N (2014) The relationship between satisfaction and turnover intentions for knowledge workers. *ASEM Eng. Manag. J.* 26(2):3–9
- Waters NM, Beruvides MG (2012) An empirical study of large-sized companies with knowledge work teams and their impacts on project team performance. *Eng Manag J* 24(2):54–70
- Kennedy DAPE (2009) Best before forty: the shelf life of an engineer. *Eng Manag J* 21(1):19–26