ResearchGate

See discussions, stats, and author profiles for this publication at: https://www.researchgate.net/publication/276298576

An Overview of Modeling and Simulation using Content Analysis

Article *in* Scientometrics · June 2015

DOI: 10.1007/s11192-015-1578-6

CITATIONS
21

READS
94

4 authors:

Saikou Y. Diallo
Old Dominion University
162 PUBLICATIONS 2,248 CITATIONS
SEE PROFILE

Ross Gore
Old Dominion University
84 PUBLICATIONS 1,184 CITATIONS
SEE PROFILE

José julian Padilla
University of Guadalajara
108 PUBLICATIONS 1,281 CITATIONS
SEE PROFILE

Christopher J. Lynch
Old Dominion University
71 PUBLICATIONS 876 CITATIONS
SEE PROFILE

All content following this page was uploaded by Ross Gore on 08 May 2024.

The user has requested enhancement of the downloaded file.

Scientometrics (2015) 103:977–1002
DOI 10.1007/s11192-015-1578-6

An overview of modeling and simulation using content analysis

Saikou Y. Diallo[1] · Ross J. Gore[1] · Jose J. Padilla[1] · Christopher J. Lynch[1]

Received: 25 September 2014 / Published online: 26 March 2015
© Akadémiai Kiadó, Budapest, Hungary 2015

Abstract Over the past six decades, Modeling and Simulation (M&S) has been used as a method or tool in many disciplines. While there is no doubt that the emergence of modern M&S is highly connected with that of Computing and Systems science, there is no clear evidence of the contribution of M&S to those disciplines. Further, while there is a growing body of knowledge (BoK) in M&S, there is no easy way to identify it due to the multi-disciplinary nature of M&S. In this paper, we examine whether M&S is its own discipline by performing content analysis of a BoK in Computer Science. Content analysis is a research methodology that aims to identify key concepts and relationships in a body of text through computational means. It can be applied to research articles in a BoK to identify the prominent topics and themes. It can also be used to explore the evolution of a BoK over time or to identify the contribution of one BoK to another. The contribution of this paper is twofold; (1) the establishment of M&S as its own discipline and the examination of its relationship with the sister disciplines of Computer Science and Systems Engineering over the last 60 years and (2) the examination of the contribution of M&S to the sciences as represented in the Public Library of Science.

Keywords Modeling and Simulation (M&S) · Content analysis · Simulation research · Association for Computing Machinery (ACM)

✉ Saikou Y. Diallo
sdiallo@odu.edu

Ross J. Gore
rgore@odu.edu

Jose J. Padilla
jpadilla@odu.edu

Christopher J. Lynch
cjlynch@odu.edu

[1] Virginia Modeling Analysis and Simulation Center, Old Dominion University, 1030 University Boulevard, Suffolk, VA 23435, USA

Mathematics Subject Classification 00B99

JEL Classification C8

Introduction

Modeling and Simulation's (M&S) progression from a series of techniques into a discipline of its own continues into the twenty-first century (Hollocks 2006; Sokolowski and Banks 2010a; Wierzbicki 2007). Historically, M&S has been viewed as a tool, method, or application used by scientists and engineers to explore problems in their domain. While it is multidisciplinary in nature, there is some agreement that it is rooted in Computer Science (CS) and Systems Engineering (SE) (Padilla et al. 2011; Tolk 2010, Mustafee et al. 2014). A vibrant international M&S community exists and in fact M&S is now taught as a discipline at the undergraduate and graduate level (Banks and McGinnis 2008). It is also true that M&S remains embedded within its application domain and is sometimes called different but synonymous names such as "simulation modeling" or "simulation and games."

There has been much discussion on what makes M&S its own discipline (Banks 2006; Sarjoughian and Zeigler 2001; Yilmaz et al. 2008). Sokolowski and Banks (2010b) explore the growth of M&S from an education viewpoint and how various aspects of M&S are pushing the discipline forward. However, none of these works conclusively state what makes up the M&S BoK. Several efforts have identified parts of the BoK of M&S by examining the co-citation pattern of journals that specialize in M&S (Mustafee et al. 2014). While these approaches are useful and insightful they need to be complemented by studies that reveal the important concepts that make up the discipline of M&S. Further, we need to explore the commonalities and differences between M&S, CS, and SE in order to establish whether M&S has a distinct and identifiable BoK.

In this paper, we seek to identify if there exists a distinct and identifiable BoK for the M&S domain. To do this we profile M&S as a discipline and explore its contribution to computing using content analysis. Content analysis is both qualitative (Hsieh and Shannon 2005) and quantitative (Katsaliaki et al. 2010; Marsh and White 2006). According to Marsh and White (2006), qualitative content analysis is an inductive approach to identify patterns and emergent concepts within a body of text. Qualitative content analysis has four steps which are (1) formulating research questions, (2) sampling, (3) coding and (4) method of analysis. The idea of formulating research questions in step one is to ask overarching and open questions and use them as a guiding influence to evaluate the body of text. For our purposes, we ask (1) to what extent M&S is different from CS and SE and (2) what is the contribution of M&S to the sciences? The idea of sampling is to identify a body of text that is broad and inclusive enough to capture all relevant patterns and concepts are identified. For sampling, we rely on a corpus of data from (1) the Association for Computing Machinery (ACM) and (2) the Public Library of Science (PLOS). We use the ACM dataset to gain an understanding of the contribution of M&S to computing. We use the PLOS dataset to broaden the profile outside of computing and discover the application of M&S within the sciences. The purpose of coding is to iteratively categorize the body of text using a coding scheme in order to identify overarching concepts and relationships. In this paper, we use an automated mathematical approach that we describe in the "Methodology, tool and data" section. Finally, the role of the method of analysis is to answer the research questions formulated in one. Analysis is part of the coding process but at this stage the focus is on providing a textured answer to the research questions using not only the concepts and relationships but also other information gleaned through the coding process.

The rest of the paper is organized as follows: "Overview of modeling and simulation" section provides an overview of modeling and simulation; "Content analysis" section provides a brief introduction to content analysis; "Methodology, tool and data" section provides the methodology, data, and the tool used in this study; "ACM as a combination of disciplines" section provides the results and analyses; and we conclude in "Conclusions" section.

Overview of modeling and simulation

Modeling and Simulation is a discipline of engineering that uses a purposeful abstraction of a referent to answer a modeling question. The main philosophy behind the discipline is to simplify complex problems into a meaningful representation called a model. The model is often expressed mathematically, logically, computationally, visually, verbally or a combination thereof. A simulation is the execution of a model over time. After a simulation is executed, results are collected and analyzed in order to formulate an answer to the modeling question. A simulation must be verified to ensure that it faithfully represents the model and the model must be validated against the referent to ensure that it is a meaningful representation of the referent and did not leave out any important aspects. As shown in Table 1, the practice of Modeling and Simulation reflects its multidisciplinary nature with journals varying from the medical field to the gaming industry. The importance of M&S is further highlighted by the variety of conferences in multiple domains as shown in Table 2.

Table 1 Sample of M&S journals

Journal	Domain	Journal impact factor (2013)
ACM Transactions on Modeling and Computer Simulation (TOMACS)	The modeling and simulation life cycle across all sciences	0.829
Journal of Simulation (JOS)	Research and practice in discrete-event simulation	0.383
Journal of Artificial Societies and Social Simulation (JASSS)	The exploration and understanding of social processes	1.733
Simulation Modelling Practice and Theory	The theory, methodology, and application of modeling and simulation dealing with systems	1.050
SIMULATION: Transactions of the Society for Modeling and Simulation International	The theory and application of general modeling and simulation issues	0.656
Simulation and Gaming: An Interdisciplinary Journal of Theory, Practice and Research	Simulation and gaming methodologies used in education, training, consultation, and research	0.447
Simulation in Healthcare: Journal of the Society for Simulation in Healthcare	Healthcare simulation methodologies covering basic, clinical, and translational research	1.593
Building Simulation	Environmental and human behavioral sciences	0.631

Table 2 Sample of M&S conferences

Conference	Domain
AlaSim International Conference and Exposition (AlaSim International)	The science and technology of simulation.
Conference for the Australian Society for Simulation in Healthcare (SimHealth)	Healthcare simulation—patient safety, clinical outcomes, and education pedagogy
The IASTED International Conference on Applied Simulation and Modeling (ASM)	Modeling and simulation, operations research, and artificial intelligence and soft computing
The IASTED International Conference on Modelling, Identification and Control (MIC)	Modeling and simulation, adaptive control, and intelligent systems
Conference on Behavior Representation in Modeling and Simulation (BRIMS)	Modeling and simulation of human factors and human–machine systems
Epistemological Perspectives on Simulation Conference (EPOS)	The epistemological aspects of modeling and simulation within the social sciences, computer science, engineering, and the natural sciences
European Modeling and Simulation Symposium (EMSS)	Modeling and simulation methodologies, techniques, and applications in industry, business, finance, and commerce
International Conference on Bond Graph Modeling and Simulation (ICBGM)	Bond graph modeling techniques for dynamic systems
International Conference on Chaotic Modeling, Simulation and Applications (CHAOS)	The ideas, methods, and techniques in the fields of nonlinear dynamics, chaos, and fractals
International Conference on Computer Modeling and Simulation (ICCMS)	Computer modeling and simulation
International Conference on Harbor, Maritime & Multimodal Logistics Modelling and Simulation (HMS)	The applications of simulation and computer technologies to logistics, supply chain management, multimodal transportation, maritime environments, and industrial logistics
International Conference on Modelling & Applied Simulation (MAS)	Modeling and simulation of logistics, supply chain management, production control, business, and industrial organization
International Defense and Homeland Security Workshop (DHSS)	Modeling and simulation innovations for Defense and Homeland Security applications
International Workshop on Energy, Sustainable Development & Environment (SESDE)	Modeling and simulation applications related to energy, sustainability, and environmental issues
Interservice/Industry Training Simulation and Education Conference (I/ITSEC)	The fields of modeling and simulation, training, education, STEM, analysis, and defense and security
Powerplant Simulation Conference (PowerplantSim)	Focuses on the special needs of nuclear and fossil power plant simulations and simulators
Simulation Interoperability Workshops (SIW)	Modeling and simulation interoperability and reuse
Spring Simulation Multi-Conference (SpringSim)	Theory of modeling and simulation, communications, networking, high performance computing, and medicine
Summer Simulation Multi-Conference (SummerSim)	Focus on modeling and simulation, tools, theory, methodologies, applications, and hybrid systems
Symposium on Simulation for Architecture and Urban Design (SimAUD)	Modeling and simulation techniques for architecture design and construction and urban design and society
The European Simulation and AI in Games Conference (GAMEON)	Ideas on programming, programming techniques, hardware design and applications for simulation and artificial intelligence in gaming

Table 2 continued

Conference	Domain
The Industrial Simulation Conference (ISC)	Industrial simulation research including the use of artificial intelligence, agents, and modeling techniques
The SIMEX Conference (SIMEX)	Simulation tools to bring simulations closer to the non-simulation, engineering community
Winter Simulation Conference (WinterSim)	Advances in system simulation within the industry, service, government, military, and academic sectors

Traditionally, attempts to map M&S as a discipline focus on an application domain such as Gaming (Crookall 2010) or a specific area within a discipline such as functional magnetic resonance imaging (Welvaert and Rosseel 2014). We are now beginning to conduct broader studies such as that conducted in Mustafee et al. (2014) where a co-citation analysis of *Simulation–Transactions of the Society of Computer Simulation* identifies key authors, articles, and cited journals. From a science of science perspective, the multidisciplinary aspect of M&S combined with its heavy dependence on SE and CS make it very difficult to identify its core papers, authors, or topics. In particular, M&S methods and techniques are tightly related to CS and SE. M&S relies on SE to formulate a model given a set of requirements, identify a modeling question, and understand restrictions associated with developing, funding, and managing the life cycle of model development. M&S relies heavily on CS to implement, execute, and improve the performance of simulations. Table 3 shows a sampling of M&S methods and techniques that are currently used in multiple domains and Table 4 shows a list of institutions and research centers active in the field.

While there is no debate on the usefulness of M&S, it is still not clear whether M&S is a discipline that stands on its own independent from CS and SE. In particular, we focus on the following question: "Is M&S a discipline that is distinct and separate from CS and SE?" It is important to note that (1) we focus on computable models, meaning models that can be successfully executed on a digital computer, and therefore exclude other useful areas of M&S such as data modeling or cognitive architectures and (2) we assume that the BoK we have selected is a significant representation of the domain of computing and systems science. We conjecture that if M&S is its own discipline then an examination of the BoK in all three disciplines over time will reveal a separation in focus areas between the disciplines. Thereby, reflecting the fact that each discipline has become specialized and concerned with its own fundamental problems even though they might have started as computing. In the next section we introduce content analysis which we will use as a method to attempt to answer our research question.

Content analysis

Content analysis is a systematic and objective technique for creating a collection of content categories from the words contained within a larger group of text (Berelson 1952; United States General Accounting Office 1996; Kassarjian 1977; Stemler 2001). This technique is capable of identifying the main themes, ideas, and topics within the text to create the resulting content categories (United States General Accounting Office 1996). Themes are

Table 3 Sample of M&S methods

M&S techniques and methods
Mathematical and equation-based modeling
Bond graph modeling
Petri nets
Markov-chain modeling
Multi-paradigm modeling
Statistical modeling
Stochastic modeling
Visual interactive modeling
Bayesian networks
Discrete-time modeling; GERT—graphical evaluation and review technique
Markov chains; Semi-Markov model
Network Modeling and Simulation
Discrete event simulation
Monte Carlo simulation;
Numerical simulation
Finite element modeling
System dynamics
Trace-based simulation
Continuous simulation/flow simulation;
Statistical simulation (including Regression and Poisson Simulation)
Rare events simulation; software-in-the-loop simulation;
Stochastic simulation; virtual reality simulation; web-based
Spreadsheet simulation
Agent-based modeling and simulation
Multi-agent systems
Agent-based geo-simulation
Devs—Cell-Devs
Composable cellular automata formalism; Devs—Devs/soa; Devs—Dsdev; Devs—eUdevs; Devs—Gdevs; Devs—Rtdevs; Devs—Cell space approach (note: this is different from Cell-Devs);
Formal specification and analysis (Maude); heterogeneous
Flow system specification formalism

critical components of content analyses as they highlight the assertions that are made about the topic and "issues, values, beliefs, and attitudes" are commonly presented in this manner (Kassarjian 1977). The systematic nature of content analysis allows for the analysis of a large amount of text at one time while creating content categories that encompass the entire set of text. Inferences and interpretations are then made from an analysis of the content categories to learn from the text (Ahuvia 2001; Weber 1990). The power of content analysis lies in the ability to identify underlying topics, domains, focuses, trends, and patterns within the data set.

This technique is applicable to all forms of communication that can be translated into a textual format. This includes documents, text passages, interviews (Wiitavaara et al. 2009), newspapers (Tse et al. 1989), case studies, written conversations, evaluations (Schredl

Table 4 Sample of M&S research and development centers

R&D centers	Domain
Virginia Modeling, Analysis and Simulation Center, Old Dominion University	Modeling and Simulation
Institute for Simulation and Training, University of Central Florida	Modeling and Simulation
Center for Medical Simulation, www.harvardmesim.org	Medical
Simulation Laboratories, NASA AMES Research Center	Engineering
Center for Modeling, Simulation, and Analysis, University of Alabama in Huntsville	Modeling and Simulation
Northwestern University Transportation Center	Transportation
University Transportation Center, Georgia Institute of Technology	Transportation
Massachusetts Institute of Technology	Transportation
Illinois Center for Transportation	Transportation
Sandia National Laboratories, www.sandia.gov	Engineering
Center for Sensed Critical Infrastructure Research, Carnegie Mellon University, www.ices.cmu.edu/censcir	Engineering
Simulation Research Group, Lawrence Berkeley national Laboratory	Engineering
Forest Biometrics Research Institute	Forestry
Massachusetts Institute of Technology, Engineering Systems Division	Engineering
RAND Corporation	Policy
Idaho National Laboratory	Engineering
Santa Fe Institute	Finance
Iowa State University, Department of Economics	Economics
Yale University, Cowles Foundation for Research in Economics	Economics
Hinkley Center For Solid And Hazardous Waste Management	Waste Management
Cornell Waste Management Institute	Waste Management

et al. 2003), and abstracts (Cretchley et al. 2010) among other textual formats in both structured and unstructured formats (Berelson 1952). The primary focus of this work is on the use of content analysis to look at journals; however, there are also many examples of content analysis being using on other types of text. Tse et al. (1989) use content analysis on newspaper advertisements to examine cultural differences between Hong Kong, the People's Republic of China, and Taiwan. Schredl et al. (2003) use content analysis to examine the similarities and differences of dreams across countries using the country grouping as the separation of bodies of knowledge.

Profiling is a form of content analysis which uses the text within a body of work to make determinations about that body of work (Lonner et al. 2010; Mustafee et al. 2014, Sagar et al. 2013). This type of content analysis is the focus of our work. Profiling can target a single journal; compare different journals; study contributions to fields as a whole; or target application domains (Mustafee et al. 2014). This provides avenues for gaining insight into a journal as a whole, gaining insight into a single topic spread across multiple journals, and comparing journals against each other.

The benefit of profiling comes from the ability to look at the development of a field in terms of data trends, field composition, and overlapping concepts over time. This provides a method for discovery within a BoK. For example, Katsaliaki and Mustafee (2011) conducted a study of simulation usage across journals to identify the main simulation techniques used in healthcare. Studies can focus on the relevance of themes within a body of work (Grimbeek et al. 2005) and contributions to specific domains, fields, or bodies of knowledge. Studies can also focus on research trends including the emergence or reemergence of topics (Cretchley et al. 2010; Gasiorek et al. 2012).

Another use of profiling with content analysis is the ability to compare bodies of knowledge. This requires first creating an initial profile of each BoK so that they can then be compared against one another. Comparing bodies of knowledge helps to examine differing perceptions of concepts between the bodies of knowledge (Crofts and Bisman 2010), examine differences in topics of interest between the bodies of knowledge, and identify gaps in the literature within the bodies of knowledge (Noltemeyer et al. 2013). Comparing bodies of knowledge can also identify the supporting disciplines and theories that support the underlying research for each BoK (Chang et al. 2010). Additionally, profiling can also be used to identify the common characteristics from each BoK in terms of cultural tendencies and appeals that can then be used to determine how to best interact with the bodies of knowledge (Jeon et al. 1999). For example, a comparison of magazine advertisements in Korea against United States advertisements shows that companies seeking to advertise in Korea should utilize emotional appeals as this is the main form of advertisement within Korean magazines (Jeon et al. 1999).

Due to the size and heterogeneity of our datasets, manual coding of the data would require a human coder years of experience in various aspects of computing and science as well as an inordinate amount of time to conduct. Therefore, we selected to use an automated approach. However, there are several limitations associated with conducting automated content analysis. An initial limitation is that the material must be obtained in a textual format. For interviews and conversations this requires that the content be converted into text before conducting an analysis. From an analysis viewpoint, the themes contained within a text are difficult to analyze since a single sentence can contain multiple themes and it is important to properly categorize these themes (Kassarjian 1977). Determining the type of content analysis (i.e. profiling vs. co-citation analysis) that is most useful for a study can be difficult to assess and can vary on a case-by-case basis (Ahuvia 2001). The most common limitation for conducting a content analysis falls on the ability of the coder to design the analysis and determine how the themes will be generated from the text (Ahuvia 2001; Kassarjian 1977; Weber 1990). Additionally, the importance of a concept can be underestimated if synonyms are used throughout the text (Stemler 2001). The methodology and the tool used to conduct the content analysis presented in the following section were selected to account for these limitations.

Methodology, tool, and data

The ACM digital library archives several thousands of articles about computing and information systems in periodicals, magazines, and journals. Furthermore, ACM has a special interest group on simulation and modeling (SIGSIM) and is home of two of the premier M&S conferences, The Winter Simulation Conference (WSC) as well as the Parallel and Distributed Simulation Conference (PADS). ACM is therefore an appropriate place to explore the relationship between M&S, CS, and SE. While SE and CS topics appear as top level category in the ACM Computing Classification System[1] (CCS); it is not the case for M&S. Rather, M&S is considered a sub-category of computing and the term "simulation" appears in sixteen different sub-categories (mainly levels three and four).

[1] http://www.acm.org/about/class/ccs98-html.

This indicates that even though M&S is an integral part of computing, its contribution is diverse and wide ranging which makes it difficult to assess its aggregate contribution. It is important to note that usually authors are asked to provide classification tags for their own work using the CCS; however this *tag-only* approach is subjective, especially since authors can provide multiple and possibly conflicting tags for their publication. In addition, some conference proceedings do not use the CCS which disqualifies an important source of information. In order to evaluate the overall place of M&S in computing, we need to look beyond the classification and into the content of the publications. The idea is to rely on the preponderance of evidence embedded within abstracts to identify key concepts, relationships, and themes associated with M&S regardless of the publication. In order to do so, we use a three part approach.

In part one, we wish to know whether there is a conceptually distinct BoK in ACM that captures the discipline of M&S as separate from that of CS and SE. We classify the ACM publications into the disciplines of CS, M&S, SE, Electrical Engineering, and Computer Engineering based on proceedings and periodicals as shown in Table 5.

We then remove the Electrical Engineering and Computer Engineering publications since we are only concerned with testing the relationship between CS, SE, and M&S. The remaining publications form the ACM corpus. The discipline of a particular publication in

Table 5 Classification of ACM publications

Discipline	Publication
Modeling and Simulation	DS-RTDistributed Simulation and Real-Time Applications, SIGMETRICS, MSWiMModeling, Principles of Advanced and Distributed Simulation, Winter Simulation Conference, Transactions on Modeling and Computer Simulation, ACM/SPEC International Conference on Performance Engineering, Performance Evaluation Methodologies and Tools, SACMAT Symposium on Access Control Models and Technologies
Systems Engineering	Journal on Emerging Technologies in Computing Systems, Transactions on Autonomous and Adaptive Systems, Transactions on Computer Systems, Transactions on Information Systems, Transactions on Intelligent Systems and Technology, Transactions on Interactive Intelligent Systems, Transactions on Management Information Systems, Transactions on Reconfigurable Technology and Systems, System Level Interconnect Prediction, Systems Programming and Applications, Ubiquitous Computing, Architecture for Networking and Communications Systems, Architecture for Networking and Communications Systems, Architecture for Networking and Communications Systems, Conference on Human Factors in Computing Systems, Document Engineering, Engineering Interactive Computing Systems, Geographic Information Systems, ACM SIGCOMM
Electrical Engineering	Design Automation Conference, Design Automation and Test in Europe, International Symposium on Low Power Electronics and Design, Transactions on Design Automation of Electronic Systems, International Symposium on Field Programmable Gate Arrays, Transactions on Computer Systems, Engineering Interactive Computing Systems, Embedded Systems Week, Embedded Network Sensor Systems, Tangible and Embedded Interaction, Computing Frontiers Conference
Computer Engineering	Great Lakes Symposium on VLSI, International Symposium on Physical Design, ACM Symposium on Parallel Algorithms and Architectures, Transactions on Architecture and Code Optimization, International Symposium on Computer Architecture
Computer Science	Everything else[a]

[a] For a complete list of publications go to http://dl.acm.org/

the ACM dataset is determined by the title of the conference or periodical in which it appears. For example, a publication in the Winter Simulation Conference pertains to the discipline of modeling and simulation. Similarly, a publication in the Transactions on Programming Languages pertains to the discipline of CS. In most cases the mapping between the periodical and discipline is straightforward. However, in some cases it is possible to disagree on the classification and in ambiguous cases we choose to classify periodicals and their publications as CS. We made this choice because CS is seen as the overarching discipline of the ACM. We use automated content analysis to confirm our manual classification and to discover commonalities and differences between the disciplines that we have identified. The *ACM as a Combination of Disciplines* section expands upon part one.

In part two, we wish to the answer the question of what is the contribution of M&S to computing? In order to do so, we examine the ACM corpus as a whole over the last six decades to explore the role of M&S in computing over time. We use automated content analysis to compare the evolution of ACM overall over the last 60 years to the evolution of the M&S, CS and SE disciplines over the same time span. *The ACM Map—An overview of the six decades* section expands upon part two.

In part three, we use the M&S corpus to train a machine learner to recognize M&S concepts and ask it to discover and explore M&S across the sciences as captured in PLOS. The *What is the contribution of M&S to the sciences?* section expands upon part three. In the next section we discuss the datasets in more detail.

Data

ACM provided our team with a download site where we retrieved all ACM publications from 1960 to 2011. From this data set we extracted the abstracts only, resulting in a total of 213,725 abstracts. This dataset is broken into decades starting from the sixties and ending in the two thousand and tens. It is important to note that some decades are incomplete due to the fact that there are years when a publication did not appear. The 2010s only include the years 2010 and 2011, so we only have a partial outlook of that decade. We use this set to create a profile of ACM and identify the role of M&S in each decade.

As discussed earlier, we also divide the ACM dataset into the disciplines of CS, M&S, SE, Electrical Engineering, and Computer Engineering. The initial 213,725 abstracts are mapped to the disciplines as follows: CS (160,293), M&S (12,823) and SE (19,235). We further breakdown each discipline into decades. In total, the ACM dataset is broken into the subsets shown in Table 6.

Table 6 Description of ACM subsets

Subset	Content
ACM all	Abstracts of publications in ACM CS, M&S and SE
ACM CS	Abstracts of publications in ACM CS
ACM M&S	Abstracts of publications in ACM M&S
ACM SE	Abstracts of publications SE
ACM by decade	Abstracts of publications in ACM CS, M&S and SE by decade starting from the sixties ending in two thousand tens
ACM (CS, SE, M&S) by decade	Abstracts of publications in ACM CS, M&S and SE respectively by decade starting from the sixties ending in two thousand tens

Finally, we form the PLOS dataset by collecting all of the abstracts of the six domain specific Public Library of Science periodicals from 2003 to 2011. The domain specific PLOS periodicals are: PLOS Biology, PLOS Medicine, PLOS Computational Biology, PLOS Genetics, PLOS Pathogens, and PLOS Neglected Tropical Diseases. The six disciplines we define from these periodicals match their respective titles: biology, medicine, computational biology, genetics, pathogens, and disease. The disciple of a particular publication is determined by the journal in which it is published. Ultimately, we categorized 13,095 abstracts which are mapped to the disciplines as follows: biology (1334), medicine (832), computational biology (2976), genetics (3041), pathogens (3348), and disease (1564). Next we present the tool used for creating a profile in more detail.

Analytical tool

The analytical tool used in this research is Leximancer. Leximancer is a content analysis tool that uses the frequency of words and their relative co-occurrence to infer the concepts most associated with a text or a corpus of text. Concepts are then grouped into themes based on a function of their frequency of occurrence and strength (Leximancer 2011). Smith and Humphreys (2006) analyzed Leximancer in terms of the algorithms used (face validity), repeatability (stability), reproducibility, and correlative and functional validity. They found that the tool is useful in extracting valuable information from text and make recommendations on how to setup Leximancer parameters to obtain reproducible, repeatable, and valid results.

While our studies' application to M&S is unique, the use of Leximancer to perform content analysis is not new. In a study similar to the one conducted in this paper, Liesch et al. (2011) used Leximancer to profile the evolution of international business as a field by exploring the leading journal of that domain from 1970 to 2007. Crofts and Bisman (2010) use Leximancer to profile the term "accountability" across 114 journal articles published in leading journals from 2000 to 2007. This is similar to the approach we employed in parts two and three of our study.

Leximancer was selected to conduct content analysis for this work based on its abilities: (1) to handle synonyms, (2) to provide access to a built-in machine learner, and (3) to quickly analyze large amounts of text at one time. The issue of potentially needing to identify multiple themes per sentence is handled directly by Leximancer through a thesaurus. This thesaurus creates a ranked list of terms associated with each concept (Leximancer 2011). This allows the researcher to objectively see which terms are most heavily associated with a specific concept and to identify which terms connected to that concept are more heavily associated with other concepts (Smith and Humphreys 2006). Using Leximancer removes the need for the researcher to code a learner from scratch and provides a configurable learner that can be set to best meet the needs of a study. See the *Variables Setup* section for further details on this aspect. Finally, Leximancer accounts for synonyms during the learning process and removes the threat of underestimating the importance of certain concepts due to the use of synonyms throughout the body of text.

Leximancer provides a number of benefits to the content analysis process; however, there are still limitations. While it provides the main themes and concepts that are contained within a body of work, it remains the responsibility of the researcher to interpret the meaning of the results. By default, Leximancer looks for the co-occurrence of one word to another word, such as "modeling" and "simulation." If it is the intent of the researcher to look for "modeling and simulation" as a term, then that term must be specifically added to Leximancer's term list by linking the two words together during the setup process

(Leximancer 2011; Poser et al. 2012). However, insight into which words should be linked together can be gained by first running Leximancer on the default settings and identifying which words are most closely associated with each other. It remains the responsibility of the researcher to identify when this is necessary or beneficial.

Variables setup

To allow for reproducibility of results, we present the setup of variables for each profile within Leximancer in Table 7. The main difference between the setups for part one and parts two and three is the use of a supervised learner. In part two, the learner uses the folder names as additional information to group concepts and themes, since the folder names reflect which decade and discipline the papers within it belong. In part three, we extract concepts from one corpus to be used as seeds to train the learner which we then use to explore another corpus. We highlight in gray the differences between the three setups.

Table 7 Leximancer setup

Stage	Category	Variable	Setup 1	Setup 2	Setup 3
Generate concept seed	Text processing options	Sentence per Block	2	2	2
		Prose test threshold	1 (Default)	1 (Default)	1 (Default)
		Identify name-like concepts	Yes	Yes	Yes
		Break at paragraph	Yes	Yes	Yes
		Auto-paragraphing	Yes	Yes	Yes
		Merge word variants	No	No	No
		Apply folder tags	Yes	Yes	Yes
		Apply file tags	Yes	Yes	Yes
		Apply dialog tags	No	No	No
	Concept seeds identification	Automatically identify concepts	Yes	Yes	No
		Total number of concepts	Automatic	Automatic	Automatic
		Percentage of name-like concepts	Automatic	Automatic	Automatic
		Concept specificity	No	No	No
		Boilerplate cutoff	Stronger	Stronger	Stronger
Generate thesaurus	Concept seeds	Auto concepts	Yes	Yes	Yes
	Thesaurus settings	Learn concept thesaurus using source documents	Yes	Yes	Yes
		Learn once	No	No	No
		Concept generality	12 (default)	12 (default)	12 (default)
		Learn from tags	No	Yes	Yes
		Learning type	Normal	Supervised	Supervised
		Sampling	Automatic	Automatic	Automatic
		Phrase separation	3	3	3
		Sentiment lens	No	No	No
		Number to discover	Off	Off	Off
		Themed discovery	Concepts in Any	Concepts in Any	Concepts in Any
		Only discover name-like concepts	No	No	No

Leximancer (2011) provides a definition of each of these variables and explains in detail how they can be used depending on the purpose of the study. In the remainder of the paper, we will refer to the setup for the first experiment as Setup 1; the setup for the second experiment as Setup 2; and the setup for the third experiment as Setup 3.

Next, we present the results of our analysis.

ACM as a combination of disciplines

In order to conduct this study, we break down the ACM corpus into a CS, SE, and M&S subsets as described in the *Data* section. In Leximancer, we use the supervised learner to associate concepts within each subset as being correlated and by doing so help put concepts that are shared between the three disciplines in context.

The quadrant chart in Fig. 1 shows that most concepts are shared between the three corpora. The axes are "Relative Frequency" which is a measure of the conditional probability of the *Concept*, given the *Category*. For instance, given that we are looking in the CS corpus, how likely is it that the concept "simulation" is mentioned. "Strength" on the other hand is a measure of the likelihood that a concept comes from a given corpus. For example, when looking at occurrences of the concept "simulation" in the CS corpus how likely is it to come from that corpus? As such, concepts in Quadrant 1 occur seldom and are not unique to the Category. Concepts in Quadrant 4 on the other hand occur seldom and are strongly associated with the Category.

We observe that:

- CS concepts occur seldom in the overall corpus but are unique to the CS domain. We also observe that simulation is as a unique but rarely occurring concept associated with the CS corpus.
- M&S concepts occur seldom and are not unique with the exception of the concepts of simulation and development. The concept of simulation occurs more often in the M&S corpus than in the CS corpus and the SE corpus.
- SE concepts are neither unique nor frequent with respect to the overall corpus. This is to be expected since the size of the SE corpus is ten times smaller than that of the CS corpus. However, it tells us that SE in ACM deals with research domains associated with the development and use of computing, including environment development, project engineering, concept development, interfaces, and virtual environments. The concept "simulation" occurs seldom and is not unique to SE.

In summary, the concepts of simulation and development are unique to CS, rarely occur in SE, and are frequent in M&S.

Based exclusively on this analysis, we cannot conclusively say that M&S and SE have distinct contributions to computing and further exploration is needed. As a result, we investigate each corpus to find the main concepts of CS, M&S, and SE in ACM over the last 60 years.

Table 8 shows a side by side comparison of the most relevant concepts for each discipline including the likelihood of a concept co-occurring with the top concept in a corpus (highlighted). As expected, we see that CS is concerned with CS (the concept of science has a strong likelihood of co-occurrence with computer), M&S with simulation design, and SE with systems design. However, we also observe that while most concepts are shared between the three, the context in which they are used is different. For instance, while

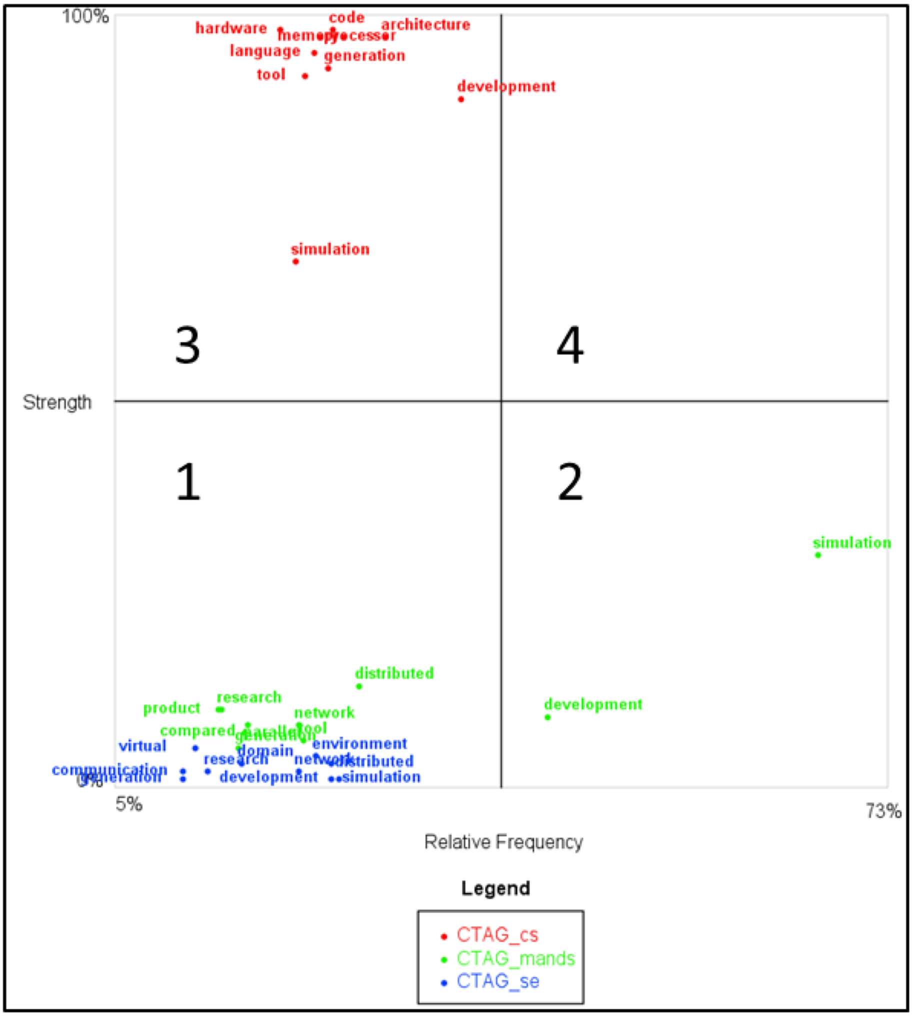

Fig. 1 Comparison of the CS, M&S and SE Corpora

"design" is important to all three domains, the CS domain talks about it most often in the context of computer design, while M&S and SE talk about simulation and system design, respectively. Similarly, the concept of "documentation" is important in CS and SE; however, in CS it most often refers to computer documentation (code documentation) while in SE it refers to system documentation. Furthermore, there are concepts that appear relatively more frequently within one domain than in others. We have coded the concepts with italics representing concepts that are shared by at least two corpora and bold for those that are shared by all. The remaining concepts are uniquely associated with a discipline according to the data.

We observe that M&S has the most unique concepts containing the least likelihood of occurrence; whereas, CS and SE share most of their concepts but with a high likelihood of occurrence. The concept of "computer security" emerges as being unique to CS while that of "system standardization" appears only in SE.

Table 8 Cross comparison of concepts across corpora

Ranked concepts for CS only		Ranked concepts for M&S only		Ranked concepts for SE only	
Word-like	Likelihood (%)	Word-like	Likelihood (%)	Word-like	Likelihood (%)
Computer	100	Simulation	100	Systems	100
Science	94	**Design**	95	*Design*	92
Design	91	**Performance**	93	Engineering	83
Performance	83	*Algorithms*	88	*Performance*	79
Theory	80	**Theory**	78	*Management*	71
Algorithms	79	*Verification*	61	*Algorithms*	70
Management	79	*Human*	57	**Theory**	70
Languages	76	Network	36	*Human*	67
Human	75	Model	35	*Languages*	65
Factors	74	Modeling	35	*Measurement*	62
Experimentation	71	Development	27	*Factors*	61
Measurement	67	Process	24	*Reliability*	60
Verification	58	Study	24	*Documentation*	58
Reliability	56	*Computer*	23	*Experimentation*	57
Documentation	54	Program	22	*Verification*	55
Security	52	Method	21	Standardization	48

The importance of computer theory (80 %), simulation theory (78 %), and systems theory (70 %) is also noteworthy. While there is a well-established computer theory in CS and systems theory in SE, the notion of simulation theory is just as strong in M&S. This can be seen as evidence of the emergence of an M&S discipline in ACM over the last 60 years. Further evidence can be found in the fact that simulation model and simulation modeling are unique to M&S despite low likelihood of co-occurrence.

While these findings do not unequivocally show that M&S is its own discipline concerned with the design, performance, algorithms, and theory of simulation, we can safely say that according to (1) the overall ACM corpus, (2) the unique and overlapping identified concepts from the CS, M&S, and SE corpora and (3) the likelihood of co-occurrence of concepts with the simulation concept in the M&S corpus, that M&S is conceptually distinct from CS and SE. Furthermore, we can say that all three disciplines are conceptually distinct with strong overlaps. Based on this analysis we can say that CS focuses on the CS aspect of computing, M&S focuses on the simulation aspect of computing, and SE focuses on the engineering aspect of computing. While rhetorical, the question now is how M&S reached this unique corpus from CS. To answer this question, we provide an overview of ACM over the six decades and identify the role of M&S in that time span.

The ACM map: an overview of the six decades

In the first level of analysis, the focus is on profiling ACM as a whole. We use the ACM corpus as a whole without conference distinction or journal provenance. Our goal is to gain an understanding of ACM as an organization through its publications. ACM as an organization is concerned with the study and advancement of computing as a scientific discipline. This is reflected in the concept map of the overall ACM corpus and the themes

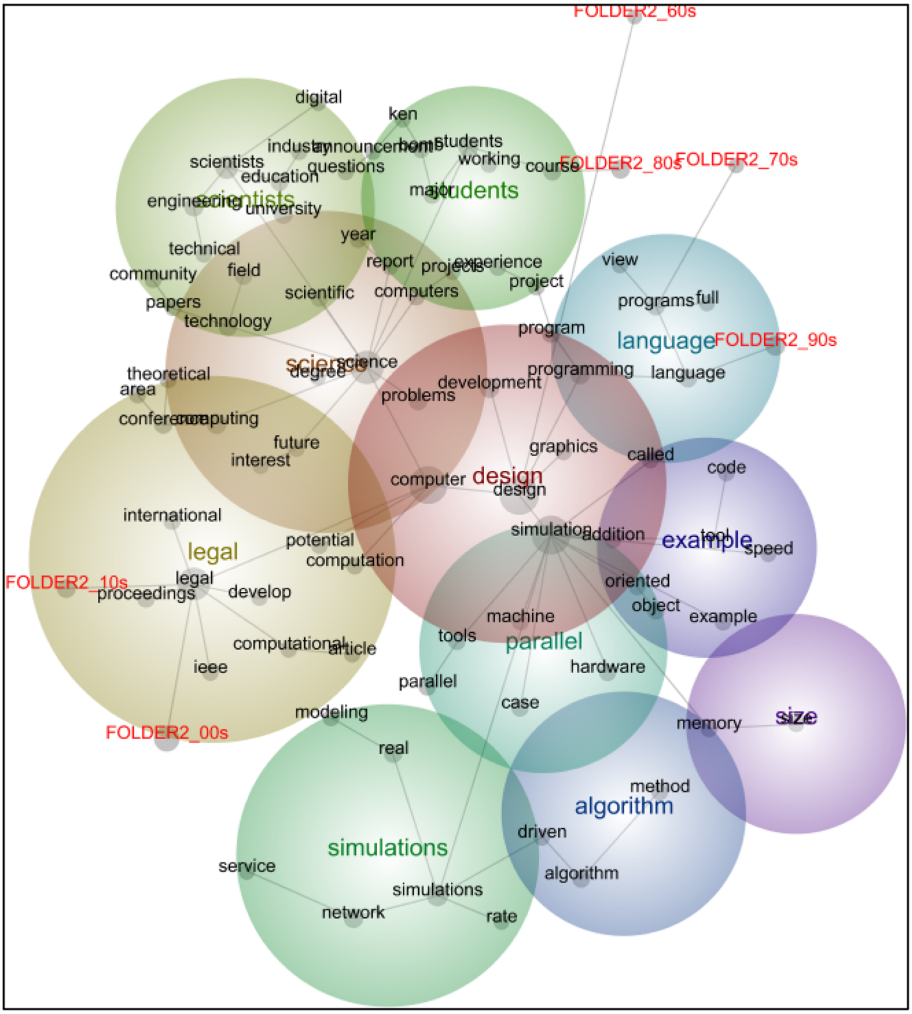

Fig. 2 Themes in ACM—sixty years view. (Color figure online)

identified in Fig. 2. Themes are color coded with darker colors indicating higher prominence.

We focus on the two most prominent themes from the ACM corpus as determined by Leximancer:

- *Design* (design, simulations, computer, general, case, structure, and international) The design theme focuses on the design aspects of computing, including code, data structures, and computer graphics. The emergence of the design theme is not surprising since (1) in the CCS, the term “design” appears in forty-seven different sub levels as opposed to sixteen times for “simulation” for instance and (2) the profiles of M&S, CS, and SE showed that the concept of design is likely to appear over ninety percent of the time with the main concept of that discipline. It is also interesting to observe the strong link between computer, hardware, design, and simulations. This suggests that simulations were envisioned as a method and tool for designing better software and hardware, as well as for designing and improving computer programs and programming languages. This is also reflected within the CCS in B.1.2 (Control Structure Performance Analysis and Design Aids) and B.4.4 (Performance Analysis and Design Aid), for instance.

- *Science* (science, university, program, scientist, students, engineering, field, projects, report, community, knowledge, Industry, technical, and social) The science theme intersects with the students, scientists, legal, and design themes and focuses on the theoretical and practical aspects of computing as a discipline, industry, and application. This theme is associated with the birth of modern CS and its establishment as a discipline that deals with computational problems at the theoretical and practical level. We also note the strong link between CS and computer scientists, which emphasizes the focus on the educational and scientific aspects of computing. This is also apparent within the CCS in the J (Computer Applications) and K (Computing Milieux) categories which deal with computers and society (K.4) or computer and information science education (K.3.2).

It is interesting to observe the emergence of a separate "simulations" theme that is associated with modeling, networks, algorithms, and methods. We will analyze this theme in more detail when we compare the M&S corpus to the ACM corpus.

In order to further explore the ACM corpus, we study the evolution of leading concepts over the decades starting with the sixties. We use a quadrant report contained in Fig. 3 to summarize our findings. As a reminder, there are four pertinent areas to the Quadrant. Concepts in Quadrant 1 occur seldom and are not unique to the Category. Concepts in Quadrant 4 on the other hand occur seldom and are strongly associated with the Category. For instance, we observe that the concepts computer, science, and simulation are unique in the 2000's while the concepts computing and scientists are not very frequent but are unique to the 2000s. The concept of simulation appears in all four quadrants. We observe that the concept occurs often in all decades but is unique in the two thousands. Since, we did not merge word variants simulation and simulations appear as different concepts with the concept of simulations appearing in the nineties with low frequency then increasing in frequency in the two thousands.

The profile of ACM as a whole broken down into decades as shown in Fig. 3 confirms the importance of the concepts of *simulation* and *design* among others. As a reminder, this dataset contains abstracts from all three disciplines and are only separated into the decades in which they appeared. An alternative view of the data shown in Fig. 3 is displayed in Table 9. In this table, we focus on the top ten concepts and their relative frequency for every decade in ACM overall.

Since we are interested in identifying the role of M&S in computing, we isolate the M&S dataset (ACM M&S) and conduct a decade by decade content analysis (Table 10).

Table 10 shows that the concept of "simulation" is relevant in every decade and is an essential concept of ACM. In fact, the concept of simulation is as relevant (percent relevance) as that of computer in every decade despite the fact that there are at least ten times more CS abstracts than there are M&S and SE abstracts in this dataset.

A further look at Fig. 3 and Table 10 shows that in the:

- *Sixties* The focus of ACM is on simulation design and computer design. In this decade, simulations are used to improve the design of computers and computer languages and the simulation community is particularly focused on the performance of simulations, jobs, and tasks in terms of time (Araten et al. 1992). Data also confirms a focus on

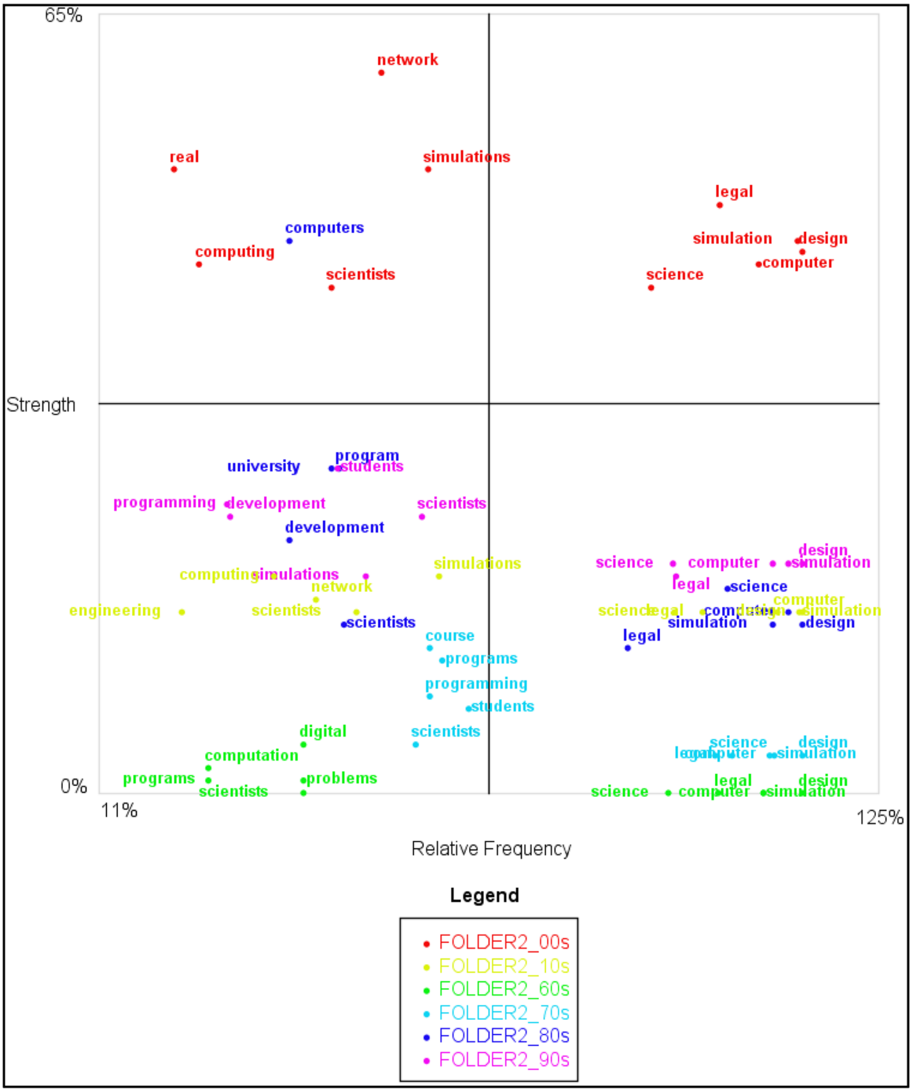

Fig. 3 ACM overall by decade

science and scientists meaning the focus is also on how computing can support science and not just the science of computing even though we see the emergence of an academic field of computing science. This finding is in line with Denning (2005).

- *Seventies* The focus of the ACM community as a whole is very similar to that of the sixties with an added emphasis on the legal aspects of computing (K5 in the CCS). However, the simulation community is now focused on simulation theory as much as performance and the concept of science is now mentioned as frequently as that of simulation. This can be an indication of the growth of computing as a science and the focus of M&S on the simulation aspects of computing science. Denning (2005) asserts that the computing industry is recruiting heavily in the systems ranks during this period but we see the evidence of this mix much later in ACM publications.

Table 9 Decade by decade top ten concepts in ACM

1960s		1980s		2000s	
Concept	Relevance (%)	Concept	Relevance (%)	Concept	Relevance (%)
Design	100	Design	100	Design	100
Computer	89	Computer	96	Simulation	99
Simulation	89	Simulation	92	Computer	88
Legal	78	Science	80	Legal	78
Science	67	Legal	59	Science	63
Digital	22	Program	25	Simulations	32
Problems	22	Scientists	25	Network	28
Scientists	22	University	24	Scientists	24
Computers	17	Computers	21	Computing	16
Engineering	17	Development	21	Real	15
1970s		**1990s**		**2010s**	
Concept	Relevance (%)	Concept	Relevance (%)	Concept	Relevance (%)
Design	100	Design	100	Design	100
Simulation	92	Simulation	96	Simulation	99
Computer	91	Computer	91	Computer	93
Legal	81	Science	68	Legal	74
Science	77	Legal	68	Science	68
Students	36	Scientists	32	Simulations	33
Programs	34	Simulations	27	Scientists	26
Course	32	Students	25	Network	23
Programming	32	Programming	18	Computing	20
Scientists	31	Development	18	Engineering	15

- *Eighties* The focus of the ACM community is the same. However, the M&S community is now focused on the design, use, and management of simulations. This might be an indication of the influence of systems thinking as noted by Denning (2005). It also marks a departure from previous decades where the focus of simulation has been on the computer. Instead, we see the emergence of simulation as a means to conduct studies that involve data and algorithms. This is closer to the use of simulation as an investigative method in the sciences. We will further explore this point when we look at the contribution of M&S to the sciences. The concept of simulation is now consistently more frequent than that of "computer" in the M&S community which indicates a focus on simulation and how simulation can support science in that community.
- *Nineties* ACM as a whole is still focused on the design of computers and simulations. The notion of computing science has now stabilized which might be a reflection of the acceptance of CS as a discipline. The M&S community has once again shifted their focus to the theory, languages, and design of simulations including simulation factors. The focus is back on simulation as part of computer theory and computer languages but we see the re-emergence of a separate simulation theory from the seventies. We also observe a concern over security and most importantly the appearance of system theory as a part of M&S which solidifies the notion of simulation as not just a computing

Table 10 Decade by decade top ten concepts in ACM M&S

1960s		1980s		2000s	
Concept	Relevance (%)	Concept	Relevance (%)	Concept	Relevance (%)
Performance	100	Human	100	Experimentation	100
Simulation	100	Algorithms	100	Design	100
Time	100	Design	100	*Simulation*	99
Control	29	Management	99	Computer	87
Job	29	*Simulation*	81	Legal	76
Times	29	Science	70	Aspects	72
Tasks	29	Paper	62	Science	63
Code	29	Time	54	Network	48
Data	29	Development	45	Time	31
Educational	29	Data	35	Development	28
1970s		1990s		2010s	
Concept	Relevance (%)	Concept	Relevance (%)	Concept	Relevance (%)
Performance	100	Theory	100	Management	100
Theory	100	Languages	100	Measurement	100
Simulation	78	Design	100	Design	100
Science	77	Factors	100	Performance	100
Sharing	25	Computer	97	Algorithms	100
Functions	25	*Simulation*	89	*Simulation*	99
Teaching	19	Security	86	Science	84
Experience	18	Science	72	Systems	74
Includes	16	System	70	Based	52
Behavior	15	Paper	53	Data	38

artifact but as a core part of the overall systems thinking. This marks the intersection between systems science and M&S or perhaps the beginning of the concept of a simulation system as it is currently understood.

- *Two thousands* ACM is still focused on the design of computers and simulations with an emphasis on the legal aspects of the design. We see the appearance of experimentation (100 % relevance) which is interesting especially in light of the remark from Fishwick (2014) that M&S is the key to observing computing as an empirical science. We see the emergence of network design (network science) which might be a reflection of the growth of the internet, social networks, and networked (distributed) simulations. We also see the re-emergence of computing as a science as noted in Denning (2005). The M&S community has now fully embraced the notion of simulation as a method to study and experiment with the real world (relevance of experimentation and science).
- *Early tens* So far in this decade, the focus of ACM as a whole is the same as it was in the sixties with perhaps an emphasis on the networking aspects of computing most probably due to the rise of social networks. M&S on the other hand is now fully concerned with the design, management, performance and algorithms of simulations. Most importantly, M&S is now more than computing and is a part of systems science and systems thinking as well as a method to conduct research.

The data shows that M&S in ACM means computing simulation, which is an approach for improving the design and performance of computers. However, M&S as a field of study developed its own theories, evolved to include the notion of a simulation system and became a method to study and experiment with the real world, which is probably a reflection of the use of computer simulation in the scientific domain. Today M&S as a field includes both the computing and systems aspects of simulation. M&S is the link between experimentation and computation or stated otherwise, M&S is the means for conducting experimentation.

Surprisingly, the notion of simulation modeling or modeling associated with simulation is not prevalent in ACM as a whole or in its M&S corpus in any decade. One explanation is that the focus is on simulation as a computing artifact or as system rather than as a method in any particular decade.

While we have focused on computing so far, there are arguments that M&S is more than simulation and by focusing on ACM we have only looked at the computational aspect of what is essentially a multidisciplinary field. Therefore, we further explore M&S by looking into the PLOS dataset in order to discover its profile.

What is the contribution of M&S to the sciences?

In order to explore the PLOS dataset, we extract concepts from the M&S dataset and use them as seeds to train a learner to recognize M&S concepts in other corpora. This corresponds to setup three in Leximancer. It is important to note that without a training set, we could not recognize M&S concepts in the PLOS dataset. Table 11 shows the top ten concepts for each discipline in PLOS.

The emergent concept is that of a simulation study which can take three forms:

- *Simulation as data generators* a simulation is used to collect data as part of a study. For instance, in the Biology, Genetics, and Pathogens subsets, we observe the prevalence of concepts like game and tools (Table 8), which indicate that simulations are used to setup an interactive environment to collect data on a subject. Gaming can also be used for training and improving the performance of students, nurses, and medical practitioners (Medicine). This is a prevalent use of M&S and there are publications solely dedicated to simulation and gaming (Simulation & Gaming for instance[2]);
- *Simulation as a method* a simulation of a system is built and that simulation is used to study the system or evaluate a theory. The study is usually of a complex system (complex networks for instance) which requires simulation as the appropriate method (Table 8).
- *Simulation as a computing tool* a computing environment consisting of hardware and software is developed for the purpose of conducting simulation studies. The focus is on improving the performance of the simulation or the algorithms used within the simulations (Comp-Bio, Table 8).

We note the prevalence of the concept object which reflects the prevalence of the compound concept "object or objective of the study." With these initial findings, we can tailor corpora to further investigate the use of simulation in PLOS. For instance, we can use a dataset from Simulation & Gaming to further look into the use of gaming in sciences (Table 12).

[2] http://sag.sagepub.com/.

Table 11 Contribution of M&S to the Sciences

Comp-Bio		Biology		Medicine	
Concept	Relative freq. (%)	Concept	Relative freq. (%)	Concept	Relative freq. (%)
Study	14	Study	15	Study	20
Method	13	Development	10	Development	8
Structure	9	Human	7	Human	6
Development	9	*Method*	7	*Method*	5
Network	9	Network	6	Control	5
Complex	7	Complex	5	Evaluation	4
Simulation	7	Process	5	Design	3
Performance	7	Interactive	5	Process	3
Process	7	Structure	5	Performance	3
Interactive	7	Control	4	Complex	2
Diseases		**Genetics**		**Pathogens**	
Concept	Relative freq. (%)	Concept	Relative freq. (%)	Concept	Relative freq. (%)
Study	18	Study	20	Study	13
Control	9	Human	7	Human	10
Human	9	Development	7	Development	6
Development	8	Structure	6	Control	5
Case	5	Complex	5	Interactive	5
Evaluation	4	Control	5	Structure	3
Network	3	Interactive	5	Complex	3
Interactive	3	*Method*	4	Process	3
Method	3	Performance	3	*Method*	2
Performance	3	Process	3	Environment	2

Discussion

First it is important to highlight the limitations of our study. M&S involves real objects in a live environment (live), real objects in a virtual environment (virtual), or virtual objects in a virtual environment (constructive). This study is focused solely on the constructive aspects of M&S which are the most easily found but are not the entirety of the domain. As a result, when we say *simulation*, we are exclusively talking about computer simulation. We also focus entirely in computing simulation within ACM which encompasses a large BoK but not the entirety of computing knowledge or M&S knowledge. Finally, we rely heavily of machine learning to identify themes and while the findings are reasonable, it is possible that a human coder might arrive at different categorizations. Therefore, it is important to repeat the process that we applied to the PLOS dataset and augment the learning dataset with new data from as varied a set of domains as possible. With this study, we have identified an initial BoK for M&S within ACM and PLOS but we need to expand it further to other domains and publications.

Based on the analysis of ACM and PLOS, we are able to confirm that M&S is closely related but distinct from CS and SE. Starting from the BoK in computing, we have shown

Table 12 Conceptual analysis of "Simulation Study" in PLOS

Biology: study		Comp-Bio: study		Diseases: study	
Concept	Relative freq. (%)	Concept	Relative freq. (%)	Concept	Relative freq. (%)
Game	33	Case	23	oriented	100
Technology	30	Physical	21	object	48
Object	29	*Game*	21	engineering	45
Science	24	Size	20	design	39
Tools	22	Object	19	theory	38
Performance	22	Technology	18	performance	34
Human	20	Scale	18	evaluation	31
Case	19	Behavior	18	*tools*	31
Environment	19	Interactive	18	parallel	30
Real	19	*Tools*	17	environment	25
Genetics: study		Medicine: study		Pathogens: study	
Concept	Relative freq. (%)	Concept	Relative freq. (%)	Concept	Relative freq. (%)
Object	57	Object	44	Object	65
Game	40	Protocol	35	*Game*	29
Tools	35	Trace	33	Evaluation	28
Design	34	Design	31	Technology	26
Case	31	Memory	30	Computer	25
Performance	31	Evaluation	30	*Tools*	24
Technology	30	Theory	29	Theory	24
Evaluation	29	Performance	29	Network	23
Memory	29	Real	27	Interactive	22
Complex	29	Case	27	Algorithms	21

the existence of a discipline of M&S with its theories, methods, and tools. We discuss the implications of the impact of these findings in three areas:

- *Academic impact* From an academic standpoint, the fact that M&S is a discipline implies that it deserves its own room within the broader engineering house. Publications such as ACM should give M&S a root level code with its own sub-levels as opposed to it being under CS. However, for M&S to be more than a combination of SE and CS, the theory of modeling and simulation needs to be featured prominently. In addition, fundamental problems of M&S need to be identified and taught as separate from those of CS and SE. Ideally, doctoral students in Modeling and Simulation will endeavor to address those problems and help move the discipline forward. Finally, the findings in this paper indicate that there is a BoK of M&S but that it is embedded in every discipline. We suggest that it is the role of the M&S community to undertake the challenge of identifying and cataloging this BoK from all application domains.
- *Professional/policy impact* From a policy standpoint, it is important to note that several M&S professional organizations exist and that a code of ethics has been established. It

is therefore essential that serious practitioners in the field be knowledgeable about the BoK of M&S in addition to that of the domain in which they apply it; especially, when M&S is used for decision support or for safety critical systems. From a policy standpoint, it means that M&S practitioners should be at least licensed or certified in some cases and the M&S industry should be recognized as its own entity rather than a subset of the application domain in which it is used.

- *Funding and research* If M&S is a discipline, basic research in M&S should be conducted and funded independently from potential application domains. The idea is that findings from basic research in M&S can be useful to all domains. Examples of basic research include validation, composability, interoperability, and formal verification which are all themes that do not appear at the top in our findings but that are very important to M&S. The fact that validation in particular does not feature prominently points to a potentially dangerous situation in which M&S is used in almost every domain without validation.

Conclusions

In this paper we use content analysis to develop a profile of M&S by studying its relationships with computing and the sciences. We find that M&S is closely related to computing and system science and confirm that while there are strong overlaps between domains, each domain is conceptually unique. We also show that simulation is the link between experimentation and computation and that the vision of simulation as a method for replicating complex, natural, or artificial processes has existed since the beginning. In the natural sciences represented by PLOS, M&S is used as a means to collect data, train skills, or to replicate complex physical systems. Finally, M&S is used to develop and optimize complex simulation environments (i.e. memory, speed, algorithms, etc.).

An important concept that we expected to be prominent but is not is "validation." This is troubling because it might be an indication that this is a neglected area of M&S. This would bring into question how much faith we can have in simulation results, especially given the increased use of simulation in decision support. Within computing it is also possible that verification is convoluted with validation; however, this is doubtful since (1) the M&S dataset consists of conference proceedings which are purported to report on practical matters and (2) the concept of study confirms the prevalence of the idea of conducting a simulation study which means that somehow the study must be validated. This finding means that more work needs to be done in developing validation methods, techniques, and practices. Other understudied areas are interoperability, composability, and visualization. This is also an indication of the fragmentation of the M&S body knowledge since we know that there are specialized groups working on those areas of M&S that do not necessarily communicate or share results with each other. For instance, interesting findings or methods for validation might not be known in the Agent-based Simulation community and new computational challenges that arise from the human behavior modeling community might be ignored in the simulation computing domain.

We need to combine the BoK of the different subdomains of M&S with the computing aspects of it in order to further study the commonalties and differences and make scientists aware of each other's domains. In addition, we will focus on training machine learning algorithms to identify subsets that we wish to further explore with content analysis. For instance, what is the role of agent based modeling in social sciences or what is the role of simulation and gaming in the medical sciences?

Acknowledgments We would like to acknowledge and thank the ACM Digital Library for granting us permission to use their archives. We would like to acknowledge friends and colleagues who have helped by providing their perspective on their own discipline and how they use M&S.

References

Ahuvia, A. (2001). Traditional, interpretive, and reception based content analyses: Improving the ability of content analysis to address issues of pragmatic and theoretical concern. *Social Indicators Research, 54*(2), 139–172.

Araten, M., Hixson, H. M., Hoggatt, A. C., Kiviat, P. J., Morris, M. F., Ockene, A., Reitman, J., Sussman J. M., & Wilson, J. R. (1992). The winter simulation conference: Perspectives of the founding fathers. In *Proceedings of the 24th conference on Winter simulation* (pp. 37–62). New York, NY: Association for Computing Machinery.

Banks, C. (2006). Academic night: Spring SIW 2006. In *Proceedings of the 2006 Spring Simulation Interoperability Workshop (SIW)* (pp. 1–3). Huntsville, AL: Society for Modeling and Simulation.

Banks, C. M., & McGinnis, M. L. (2008). Compelling challenges and recommended solutions: Developing a continuity of M&S education from public school to graduate studies. In *Proceedings of the 2008 Spring simulation multiconference* (pp. 773–780). San Diego, CA: Society for Computer Simulation International.

Berelson, B. (1952). *Content analysis in communication research*. Glencoe, IL: The Free Press.

Chang, Y. H., Chang, C. Y., & Tseng, Y. H. (2010). Trends of science education research: An automatic content analysis. *Journal of Science Education and Technology, 19*(4), 315–331.

Cretchley, J., Rooney, D., & Gallois, C. (2010). Mapping a 40-year history with Leximancer: Themes and concepts in the Journal of Cross-Cultural Psychology. *Journal of Cross-Cultural Psychology, 41*(3), 318–328.

Crofts, K., & Bisman, J. (2010). Interrogating accountability: An illustration of the use of Leximancer software for qualitative data analysis. *Qualitative Research in Accounting & Management, 7*(2), 180–207.

Crookall, D. (2010). Serious games, debriefing, and simulation/gaming as a discipline. *Simulation & Gaming, 41*(6), 898–920.

Denning, P. J. (2005). Is computer science science? *Communications of the ACM, 48*(4), 27–31.

Fishwick, P. (2014) Computing as model-based empirical science. In *Proceedings of the 2nd ACM SIGSIM/PADS conference on principles of advanced discrete simulation* (pp. 205–212). New York, NY: ACM.

Gasiorek, J., Giles, H., Holtgraves, T., & Robbins, S. (2012). Celebrating thirty years of the JLSP analyses and prospects. *Journal of Language and Social Psychology, 31*(4), 361–375.

Grimbeek, P., Bryer, F., Davies, M., & Bartlett, B. (2005). Themes and patterns in 3 years of abstracts from the international conference on cognition, language, and special education research: Identified by Leximancer analysis. *Stimulating the ''action'' as participants in participatory research Brisbane.* Australia: Griffith University, School of Cognition, Language, and Special Education, pp. 101–113.

Hollocks, B. W. (2006). Forty years of discrete-event simulation—a personal reflection. *Journal of the Operational Research Society, 57*(12), 1383–1399.

Hsieh, H. F., & Shannon, S. E. (2005). Three approaches to qualitative content analysis. *Qualitative Health Research, 15*(9), 1277–1288.

Jeon, W., Franke, G. R., Huhmann, B. A., & Phelps, J. (1999). Appeals in Korean magazine advertising: a content analysis and cross-cultural comparison. *Asia Pacific Journal of Management, 16*(2), 249–258.

Kassarjian, H. H. (1977). Content analysis in consumer research. *Journal of Consumer Research, 4*(1), 8–18.

Katsaliaki, K., & Mustafee, N. (2011). Applications of simulation within the healthcare context. *Journal of the Operational Research Society, 62*(8), 1431–1451.

Katsaliaki, K., Mustafee, N., Dwivedi, Y. K., Williams, T., & Wilson, J. M. (2010). A profile of OR Research and Practice published in the Journal of the Operational Research Society. *Journal of the Operational Research Society, 61*(1), 82–94.

Leximancer (2011). Leximancer Manual: Version 4. Resource document. https://www.leximancer.com/site-media/lm/science/Leximancer_Manual_Version_4_0.pdf. Accessed 28 Aug 2014.

Liesch, P. W., Håkanson, L., McGaughey, S. L., Middleton, S., & Cretchley, J. (2011). The evolution of the international business field: A scientometric investigation of articles published in its premier journal. *Scientometrics, 88*(1), 17–42.

Lonner, W. J., Smith, P. B., van de Vijver, F. J., & Murdock, E. (2010). Entering our fifth decade: An analysis of the influence of the Journal of Cross-Cultural Psychology during its first forty years of publication. *Journal of Cross-Cultural Psychology, 41*(3), 301–317.

Marsh, E. E., & White, M. D. (2006). Content analysis: A flexible methodology. *Library trends, 55*(1), 22–45.

Mustafee, N., Katsaliaki, K., & Fishwick, P. (2014). Exploring the modelling and simulation knowledge base through journal co-citation analysis. *Scientometrics, 98*(3), 2145–2159.

Noltemeyer, A. L., Proctor, S. L., & Dempsey, A. (2013). Race and ethnicity in school psychology publications: A content analysis and comparison to publications in related disciplines. *Contemporary School Psychology, 17*(1), 129–142.

Padilla, J. J., Diallo, S. Y., & Tolk, A. (2011). Do we need M&S Science? *SCS M&S Magazine, 8*, 161–166.

Poser, C., Guenther, E., & Orlitzky, M. (2012). Shades of green: using computer-aided qualitative data analysis to explore different aspects of corporate environmental performance. *Journal of Management Control, 22*(4), 413–450.

Sagar, A., Kademani, B. S., & Bhanumurthy, K. (2013). Research trends in agricultural science: A global perspective. *Journal of Scientometric Research, 2*(3), 185.

Sarjoughian, H. S., & Zeigler, B. P. (2001). Towards making modeling & simulation into a discipline. *Simulation Series, 33*(2), 130–135.

Schredl, M., Ciric, P., Bishop, A., Gölitz, E., & Buschtöns, D. (2003). Content analysis of German students' dreams: Comparison to American findings. *Dreaming, 13*(4), 237–243.

Smith, A. E., & Humphreys, M. S. (2006). Evaluation of unsupervised semantic mapping of natural language with Leximancer concept mapping. *Behavior Research Methods, 38*(2), 262–279.

Sokolowski, J. A., & Banks, C. M. (2010a). *Modeling and simulation fundamentals: theoretical underpinnings and practical domains*. HoBoKen, NJ: Wiley.

Sokolowski, J. A., & Banks, C. M. (2010b). The Geometric Growth of M&S Education: Pushing Forward, Pushing Outward. *SCS M&S Magazine*, *1*(4), 1–5.

Stemler, S. (2001). An overview of content analysis. *Practical assessment, research & evaluation, 7*(17), 137–146.

Tolk, A. (2010). Engineering management challenges for applying simulation as a green technology. In *Proceedings of the 31st Annual National Conference of the American Society for Engineering Management (ASEM)* (pp. 137–147). Fayetteville, AR: American Society for Engineering Management.

Tse, D. K., Belk, R. W., & Zhou, N. (1989). Becoming a consumer society: A longitudinal and cross-cultural content analysis of print ads from Hong Kong, the People's Republic of China, and Taiwan. *Journal of consumer research, 15*(4), 457–472.

United States General Accounting Office (1996). *Content Analysis: A Methodology for Structuring and Analyzing Written Material*. GAO/PEMD-10.3. 1. Washington, DC.

Weber, R. P. (Ed.). (1990). *Basic content analysis (No. 49)*. Beverly Hills: Sage.

Welvaert, M., & Rosseel, Y. (2014). A review of FMRI simulation studies. *Plos One, 9*(7), e101953.

Wierzbicki, A. P. (2007). Modelling as a way of organising knowledge. *European Journal of Operational Research, 176*(1), 610–635.

Wiitavaara, B., Björklund, M., Brulin, C., & Djupsjöbacka, M. (2009). How well do questionnaires on symptoms in neck-shoulder disorders capture the experiences of those who suffer from neck-shoulder disorders? A content analysis of questionnaires and interviews. *BMC Musculoskeletal Disorders, 10*(1), 30.

Yilmaz, L., Davis, P., Fishwick, P., Hu, X., Miller, J., Hybinette, M., Oren, T., Reynolds, P., Sarjoughian, H. & Tolk, A. (2008). What makes good research in modeling and simulation: Sustaining the growth and vitality of the M&S discipline. In *Proceedings of the 40th Conference on Winter Simulation* (pp. 677–688). Miami, FL: Winter Simulation Conference.

View publication stats