STUDIA METODOLOGICZNE
NR 39 • 2019, 123–152
DOI: 10.14746/sm.2019.39.5

Andrzej Jarynowski, Michał B. Paradowski, Andrzej Buda

Modelling communities and populations: An introduction to computational social science

Abstract. In sociology, interest in modelling has not yet become widespread. However, the methodology has been gaining increased attention in parallel with its growing popularity in economics and other social sciences, notably psychology and political science, and the growing volume of social data being measured and collected. In this paper, we present representative computational methodologies from both data-driven (such as "black box") and rule-based (such as "per analogy") approaches. We show how to build simple models, and discuss both the greatest successes and the major limitations of modelling societies. We claim that the end goal of computational tools in sociology is providing meaningful analyses and calculations in order to allow making causal statements in sociological explanation and support decisions of great importance for society.

Keywords: computational social science, mathematical modelling, sociophysics, quantitative sociology, computer simulations, agent-based models, social network analysis, natural language processing, linguistics.

1. One model of society, but many definitions

Social reality has been a fascinating area for mathematical description for a long time. Initially, many mathematical models of social phenomena resulted in simplifications of low application value. However, with their growing validity (in part due to constantly increasing computing power allowing for increased multidimensionality), they have slowly begun to be applied in prediction and forecasting (social engineering), given that the most interesting feature of social science research subjects – people, organisations, or societies – is their complexity, usually based on non-linear interactions. Consequently, the application of modelling in applied social science has expanded in the 21st century, with a gradual shift towards data science with the growing availability of Big Data. Both abroad and in Poland one can now observe the spread of modelling and analytical knowledge and competences among young quantitative sociologists, so a shift of priorities in sociology towards computational social science is to be expected.

1.1. Defining models and methods

As in sociology the term 'model' may refer to a range of concepts, in this paper we focus on the narrower, strict understanding of a 'hard' model as a projection of society and its properties which allows computational operations. In this limited sense, a 'hard' model should be described by a mathematical formula or formalism, while its 'soft' counterpart may be describableed with words or infographics attempting to systematise the data or knowledge [Pabjan, 2004]. Natural science is not monolithic [Jarynowski, 2017], thus a branch of science dealing with the example described above will be called system science or system dynamics by computer scientists, dynamical systems or non-linear dynamics by mathematicians, and complex systems or complexity science by physicists.

1.2. Aims of this paper

In this paper, we discuss the characteristics of different types of models, notably genesis, limitations and advantages, and their main impact in the field of modelling communities. The main dividing line distinguishes rule-based and data-driven approaches. In the historical section, we present the roots of modelling and the pairing of concepts from philosophy and hard science in order to help understand the development of modern sociology and its subfield of computational social science [Kułakowski, 2012]. We then describe in detail subcategories of models as well as a few more lines of division corresponding to the specificity of the research questions asked and the methods applied.

1.3. Culture of modelling

Sociologists have always been trying to measure properties of society (such as social sentiments) and how to predict and understand social changes. Natural scientists and mathematicians would then incorporate the results of these measurements and experiments into models, usually based on an analogy with a known natural phenomenon (*approach via analogy*). On the other hand, computer scientists and statisticians have developed a set of "black box" tools to predict states of society. Those techniques (*approach via black boxes*) are currently used in a "Big Data" world, and even if the underlying models are not understandable to a layperson, they usually perform more reliably than rule-based techniques.

2. Modelling by Analogy (rule-based phenomenology)

The interface of sociology (social science) with mainly physics and biology (natural science) is supported by mathematics and computer science. Following the positivist tradition in social sciences of imitating natural sciences [Pabjan, 2004], which will be discussed more extensively in the section on the history of the fields, is usually based on the belief that there exists an analogy between particles, atoms, or molecules on the one hand and living organisms, humans, and even entire ecosystems and societies on the other. This analogy has been made in order to answer questions about the relationships, evolution and functions of society. These postulates and research directive seem to be different from the mainstream (especially Polish) sociological perspective, relying on the role of the experiment [Ossowski, 1962]. For a very long time natural science methodologies in social sciences were reduced to statistical hypotheses and the running of regressions [Sztompka, 1973]. Nature obviously provides many interesting examples of phenomena that can occur in humans [Lesniewska, Lesniewski, 2016], for instance coordinated movement of animals such as flocks of birds or schools of fish. The animals that move in such patterns usually lack a leader, and their knowledge is local, but in most cases the herd is moving in the right direction. Their models can explain the gregarious properties of human communities if the variable trait represents the position of the individuals. However, a breakthrough in the practical application of new techniques for social sciences has only become possible at the turn of the 21st century, when the volumes and complexity of digitalised information on social activities have forced a paradigm change in social studies. The need for understanding social phenomena within a broader, interdisciplinary perspective emerged as a natural step in detailed studies. Rule-based approaches allow building more comprehensive, multidisciplinary planes of social knowledge.

3. Black Box modelling (data-driven)

A different modelling approach is data exploration, where prior knowledge about interaction between people is unnecessary. This may reveal complex relations between agents in various systems without the need to understand the ongoing processes. Statistical and computational visualisation methods and quantitative techniques are currently being fully exploited in research devoted to social systems. Basic quantitative methods such as statistical regressions and data-mining procedures are extended to a range of various innovations that improve data on the evolution of social systems and, most importantly, their objective prediction [Komosiński, 2011]. This area has changed the quantitative sociology of the 21st century. Analysis and modelling of human behaviour has been widely applied in academia, and even more so in commerce. Content targeting such as that used in algorithms behind Google, Amazon or Facebook adverts is a nearly trillion-dollar business [Tyler, 2018], and the accuracy of these tools relies on solid data-driven social models. Moreover, understanding of quantitative methods is taken for granted, so current and future generations of social scientists must be increasingly more familiar with data-driven modelling.

4. History of modelling societies

The modelling of societies has difficulty being recognised as part of social science. Still, the application of computer simulations and mathematical methods to the modelling of social behaviour should not come as a surprise, after they proved their value in the understanding of physical and biological systems [Komosiński, 2011]. The current era of the modelling of societies began with the works of Schelling [Schelling, 1971] and Galam [Galam et al., 1982]. The philosophical background, however, is much older.

4.1. The era of positivism and the establishment of modelling tools and theory

In 1766 Daniel Bernoulli presented probably the first mathematical model[1] of society on a population scale [Jarynowski, 2010] and stated, *I simply wish that, in a matter which so closely concerns the well-being of mankind, no decision shall be made without all the knowledge which a little analysis and calculation can provide*.[2] In the 21st century his wish came true in many countries.

In 1784 Immanuel Kant spoke of 'universal laws' which, *however obscure their causes, permit us to hope that if we attend to the play of freedom of human will in the large, we may be able to discern a regular movement in it. Moreover, that what seems complex and chaotic in the single individual may be seen from the standpoint of the human race as a whole to be a steady and progressive though slow evolution of its original endowment* [Kant, 1784].[3]

Moreover, Kant claimed that elementary mathematics could be synthetic *a priori* because its statements may provide new knowledge not derived from experience [Kant, 1781].

[1] On a side note, Bernoulli used census data of the 17th-century city of Wrocław [Bernoulli, 1766], the home town of the first author.

[2] « Je souhaite simplement que, dans une affaire qui concerne de si près au bien-être de la race humaine, aucune décision ne sera faite sans toutes les connaissances dont une petite analyse et de calcul peuvent fournir. » [Bernoulli, 1760]

[3] „[…] so tief auch deren Ursachen verborgen sein mögen, läßt dennoch von sich hoffen: daß, wenn sie das Spiel der Freiheit des menschlichen Willens im Großen betrachtet, sie einen regelmäßigen Gang derselben entdecken könne; und daß auf die Art, was an einzelnen Subjecten verwickelt und regellos in die Augen fällt, an der ganzen Gattung doch als eine stetig fortgehende, obgleich langsame Entwickelung der ursprünglichen Anlagen derselben werde erkannnt werden können." [Kant, 1781]

In 1851-54 August Comte in his book "*Système de politique positive*" introduced term 'social physics' as 'mécanique sociale', a mechanical social science based solidly on statistics and reflecting positivist approaches of world scientification. He wrote that whatever concerns the human species considered *en masse* belongs to the domain of physical facts; the greater the number of individuals, the more the individual will be submerged beneath the series of social facts which depend on the general causes according to which society exists and is conserved.

In 1904 Max Weber presented the concept of the "ideal type" which permits drawing up a schema where some features are very strongly emphasised in order to analyse the complex and amorphous social world. This fundamental axiom (methodological reductionism) is the quintessence of positivism and reductionism in sociology and lies at the foundation of similar concepts of rationality in economics. In the same year George Simmel formulated the theory of social interaction, where for example fashion is a form of social relationship. As a consequence, at the beginning of the 20th century the first mathematical models of societies appeared within this concept, such as that of evolutionary language adaptation by Otto von Wiener, or Pareto's distribution of wealth.

In 1918, when Marian Smoluchowski concluded that probability is the central problem of modern science, he introduced the methodology of social science into physics. He said: *The probability calculus, which since the beginnings of its development found the most successful application—besides mathematical approaches—mainly in the less anticipated areas of social and biological processes, has recently earned itself a (new) very important field of application: physics.* [Smoluchowski, 1918; transl. MBP].[4]

If physics had failed to learn from sociology at that time, we do not know for how long the former field would have remained in the utopia of determinism and reductionism.

[4] "Die Wahrscheinlichkeitsrechnung, welche seit Beginn ihrer Entwicklung mit größtem Erfolg hauptsächlich in dem sonst der mathematischen Behandlung wenig zugänglichen Bereich sozialer und biologischer Vorgänge angewendet wurde, hat sich in den letzten Zeiten ein überaus wichtiges Anwendungsgebiet erobert: die Physik." [Smoluchowski, 1918]

4.2. Anti-positivism/Post-modernism era. Great recession of modelling abilities

Up to end of World War I the main easily translatable sociological theories were grounded in structural functionalism, interactionism, utilitarianism, exchange and conflict theories. However, the development of modelling in social sciences was later suppressed by the boom of the humanist and anti-positivist philosophies of interwar science in the 1920s and the 1930s and weakened by the subsequent discovery of quantum mechanics, which partially altered classical logic. Individualistic mainstream sociologists interested in post-modernism, with the attitudes of "anything goes" of Feyerabend and Kuhn, led to a dispute over the methods of social sciences which, following [Ossowski, 1962] and [Sztompka, 1973], resulting in a "big divorce" of Polish sociology and the techniques of natural sciences. Perhaps one consequence of this misunderstanding was a crisis of Polish sociology, which was even unable to predict the determinant role of the solidarity movement "Solidarność" in the political transformation in Poland. However, owing to the development of computer calculation techniques, modelling was proven in application in other, sometimes fairly remote, areas such as biology.

4.3. Neo-positivism. Modelling comes back on stage

In the same post-war time, with the development of statistics and computations (where computers started to solve equations) one can observe the emergence of independent fields of social science: analytical sociology [Merton, 1968] and social cybernetics [Wiener, 1950]. Although both areas use mathematical descriptions of human behaviour, in most cases they lack analogies to the laws of nature, because the models only rely on engineering (cybernetics) or sociological (analytical sociology [Hedström, Bearman, 2009]) descriptions of processes. For example:

- Merton [1968] was able to explain the rich club phenomenon (most resources are owned by a very small number of people);
- Coleman [et al., 1957] was able to explain the spread of adaptation;

- Axelrod and Hamilton [1981] were able to explain the rationality of cooperation and conflict;
- Schelling [1971] was able to explain segregation processes.

In Poland, these approaches never got enough attention from social science representatives [Szaniawski, 1971], even though cybernetics [Mazur, 1966] and sociotechnics [Kossecki, 1996] were extensively studied in the USSR.

In the 1950s, the frontiers of science propelled the modelling of societies thanks to the works of ([Bertalanffy, 1950]; "open systems") and later ([Anderson, 1972]; "more is different"). Bertalanffy criticised attempts to describe living systems on the basis of closed thermodynamic systems and called for a completely new approach, taking into account the systems' openness and exchange of energy and information with the environment. Anderson's seminal [1972] paper argued that *the behavior of large and complex aggregates of elementary particles [...] is not to be understood in terms of a simple extrapolation of the properties of a few particles. Instead, at each level of complexity entirely new properties appear*

Sociologists such as Durkheim argued that changes in the evolution of social systems occur not only incrementally, but also rapidly, in a manner corresponding to non-equilibrium statistical physics. Thus, the fundamental question of social theory is the concept of human action [Granovetter, 1985], which takes into account not only internalised norms and values (oversocialised — collective behaviour), but also the self-interest of the actors (undersocialised — linear changes) in the individual decision-making process (linear changes). The approaches to describing interactions offered in this framework correspond to Newtonian mechanics (undersocialised systems) and statistical physics (oversocialised systems).

4.4. The digital era – Greatest expansion of modelling

A real boom in the field of modelling societies occurred around the turn of the centuries, when a plethora of new models were developed (mainly around Social Network Analysis), aiming to recreate the entire spectrum of social phenomena [Barabasi et al., 2002]. It is worth mentioning the visible representation of Polish scientists in the process, such as Katarzyna Sznajd-Weron (Wrocław), Andrzej Nowak (Warsaw) or Jacek Szmatka (Kraków) [Sznajd-Weron, 2001].

5. Sociology as a hard science

As we have shown above, modelling may be applied to social science because all the predicted results can be verified. However, this condition is not sufficient to consider sociology a 'hard science' in the same way as physics, chemistry or biology, because of lack of falsifiability. The pioneers of logical positivism such as D'Alambert or The Vienna Circle grounded science in observation, while non-science was deemed non-observational and hence meaningless. Falsifiability is the central property of science and every genuinely scientific claim has the potential of being proven false, at least in principle [Popper, 1959]. According to the principles of logical empiricism, physics may fulfil these conditions more than any other science. (Thus, sociology has been partially considered as a science in the narrower domains of sociobiology [Carnap, 1928] and, later on, sociophysics.) Therefore, the concepts of modelling in sociology must be consistent with physicalism defined by the general principles of the positivist program and its conceptual bases: 1) the construction of a universal system which would encompass all of the knowledge furnished by the various sciences, 2) the absolute rejection of metaphysics, in the sense of any propositions not translatable into verifiable scientific sentences [Neurath, 1931].

On the other hand, modelling in hard science may also explain emergent and critical phenomena. However, the existence of a correlation does not always mean causation [Paradowski, 2011]. In this case, we recommend to use the criteria of logical empiricism. For example, whenever a correlation between processes could be explained by interdisciplinary research based on culture and biology, it is more reasonable to consider biological reasons as a cause. "Metaphysical" approaches have been massively criticised and can only be considered as complementary methods, for example in hypothesis generation [Sedlak, 1980] in model explorations.

6. Some techniques of modelling societies

Mathematical methods and computer simulations are becoming increasingly popular and successfully applied in explanations of phenomena observed in real-world social, economic and biological systems. Here we explore two main computational methodologies, agent-based modelling and system dynamics, and one main analytical tool – social network analysis, all of which allow to represent complicated and complex non-linear social systems [Pabjan, 2004], as well as several other techniques. We also mention various other approaches, such as machine learning, decision trees, and neural networks.

7. Data exploration and data analysis

Collecting data on human activity can rely on linear and non-linear techniques of data analysis. Stochastic processes and data-mining allow investigating the properties of the system. One can analyse the data sets (survey, register-based, time series, spatial, panel, longitudinal data, etc.) and (re) construct (simulate processes) with similar characteristics (e.g. distributions), to predict future states [Jarynowski, 2017]. Series may be analysed from a hierarchical [Mantegna, 1998] or fractal [Kwapień, Drożdż, 2012] perspective to explore the features of the processes. A structured database allows running quantitative tree decision algorithms, though in the case of machine learning and neural network approaches this is not even required.

7.1. Regressions and multivariate analyses – artificial intelligence "in people's homes"

Logistic regression is a mathematical model which allows describing the influence of many variables $X = (X_1, \ldots, X_i, \ldots)$ on dependent variable $Y(X)$. A multivariate analysis consists of those statistical techniques that consider two or more related variables as an entity and attempt to produce an overall result taking into account the relationship among the variables. Regression analyses are widely used for risk/likelihood prediction and forecasting future statuses [Duch et al., 2000].

7.2. Structural equation modelling (SEM)

The idea behind SEM is the possibility of finding out causal relationships by systematic analysis [Konarski, 2009]. Given a set of questions (manifest variables) corresponding to theory-related statements (latent variables), one can provide [Jarynowski, 2017]:

- causal modelling, or path analysis, which validates relationships among the variables and tests them with manifest and latent variables;
- confirmatory factor analysis, an extension of factor analysis in which specific hypotheses about the structure of the factor loadings and intercorrelations are tested against the data.

7.3. Natural Language Processing (NLP)

In the field of computational linguistics, (statistical) modelling has usually aimed at the development of probabilistic models able to accurately predict the most likely next word in a provided string [Goodman, 2001]. As sociology has processed human written or spoken signals since the beginning of this field [Thomas, Znaniecki, 1918], NLP techniques have been widely applied in qualitative analyses. NLP is thus central to a range of real-world language modelling, machine-learning/deep learning and AI (artificial intelligence) applications such as automatic speech recognition, speech-to-text software, handwriting recognition (think smartphone or tablet input), information retrieval, text summarization, or machine translation. A lay reader will have encountered algorithms relying on NLP while dealing with Google Suggest or using mobile phone interfaces while typing.

Recent advances in NLP have shifted to the use of recurrent neural networks [Jozefowicz et al., 2016] and networks with long-term memory [Sundermeyer, Ney, Schlüter, 2015]. The main advantage of connectionist models [Bengio et al., 2003] over traditional statistical NLP such as non-parametric *n*-gram models that had been developing since the 1980s [Jelinek, Mercer, 1980], or subsequent shallower feed-forward neural network models, is their improved classification accuracy alongside the increased ability to generalise and scale [Józefowicz et al., 2016]. This line of enquiry is often termed Neural Language Modelling. Usually, words in the training dataset are parametrized as vectors, whereby lexemes sharing similar functional feature values (e.g. expressions both grammatically and semantically close) are analogously proximal in the induced vector space [Yoon et al., 2016].

7.4. Computational language identification (classifiers)

One noteworthy recent development has been computational identification of the native tongues of second-language users [Paradowski, 2017: 69–71]. This area has grown out of the field of stylometry (concerned with authorship attribution based on statistical calculation of textual features [Barr, 2003]) and automated text classification (which frequently applies machine learning algorithms to sort texts by their type or author attributes [Stemle, Onysko, 2015]).

Resting on the assumption that language users from different native tongue backgrounds exhibit distinguishable profiles in their language production, native language detection likewise uses computer classification techniques and machine learning algorithms trained on large databases of learner texts in an attempt to discover the constellations of words, multiword sequences, error types and other textual features most predictive of authors' mother tongue [Paradowski, 2017: 69–71]. The first application of automated text classification by the author's native languages was a study aiming at distinguishing between Chinese and Japanese learners of English [Mayfield, Tomokiyo, Jones, 2001]. The next two decades have seen studies of learners from various language constellations [Paradowski, 2017]. The accuracies of the best classifiers currently range between 80 and 84 per cent [Tetreault, Blanchard, Cahill, 2013]. For instance, the employment of support vector machines managed to differentiate between original and translated Italian texts in 86.7% of the samples in one corpus [Baroni, Bernardini, 2006].

Models can also be used to uncover insights for second-language user pedagogy. For instance [Rezaee, Golparvar, 2016] used random forest modelling to detect the most salient predictors of clause ordering in academic English.

8. Rule-based methodology

Differential equations (used by dynamical systems, system dynamics and other approaches) were the first to be applied to describe and predict phenomena, but recently even more frequent have been agent-based models (ABM) [Jarynowski, 2017]. Sometimes one problem could be solved in multiple ways. Despite the parallel development of numerical methods for differential equations, agent-based models usually give more accurate predictions and hints for decision-makers. On the other hand, differential equations allow us to understand the core process, something that could be missing in an agent-based approach. As a result, both perspectives are common among specialists and depending on the theoretical or applied aspects, their respective prevalence differs.

8.1. System dynamics (SD)

The main idea in SD is to draw up a set of differential equations representing social phenomena. The model (equations or diagrams) can be solved by numerical or approximative procedures, easily available with several types of computer software used for SD, such as Vensim, Dynamo, iThink or Stella. Their graphical notation allows non-mathematicians such as sociologists to build and solve sets of differential equations [Brouwers, 2009]. The dynamical variables are represented as stocks and rates of change as flows.

8.2. Agent-based models (ABM)

ABM is a computational technique used for experiments with artificial systems populated by agents which interact in non-trivial ways. This is probably the most common approach to modelling used by sociologists. The available toolkits include Netlogo, Swarm, RePast or MASON. In NetLogo, an agent (an autonomous, interacting entity), is represented by a turtle, while a patch is the elementary spatial unit in the grid. The goal is to imitate real patterns by running (often computerised) ABMs under different treatments and conditions. This approach is used much more often by economists than by sociologists [Kaminski, 2012].

8.3. Dynamical systems and chaos

The evolution of dynamical systems is ruled by an implicit relation of input and output. If the behaviour of a dynamical system is highly sensitive to the initial conditions, it can be described in terms of chaos.

8.4. Control theory and cybernetics

The general goal of control theory and cybernetics is discovering patterns and finding principles that can be applied to prediction or control [Poczybut, 2006]. Apart from feedback loops—the central point of system theory-many other techniques are used, such as neural networks, artificial intelligence, artificial life, swarming and floating algorithms. A very important issue is emergence – a process whereby interactions among basic entities exhibits properties that cannot be derived from a simple sum of the entities. After Anderson [1972], this has often been often referred to as the "more is different" principle.

8.5. Critical phenomenon and self-organisation.

The critical phenomena theory [Jarynowski, 2007] has been applied to solving problems in behavioural, social and political sciences, although its roots can be traced to the mathematics and physics of the early 20th century. Catastrophes, self-organising processes and chaos have been broadly overlapping in social and natural science within the complex system paradigm. Many socio-economic systems, where the equivalent of the Reynold number is high enough, can be described with the turbulent flow theory of fluids and gases.

8.6. Evolutionary models

Evolution may explain the formation of different population structures, speciation through the rivalry of two forces of nature – natural selection and genetic drift [Leśniewska, Leśniewski, 2016]. This approach can be derived from the theory of thermodynamics. In reference to sociology, different analogies are used depending on the model, but genetic drift may correspond to the transmission of cultural information, and natural selection can be observed in societies favouring the fittest individuals. There is still a lot of variability due to various factors, like fashion, wars or external. Natural selection results in the differentiation of adjacent societies, because each social group adapts to its local conditions. The cultural transmission of civilisation tends to blur the differences between populations [Dybiec, et al., 2012]. In addition to vertical transmission from parents to descendants, there is horizontal transmission (the media, schools, etc.), whose mission is to maintain the social structure [Bourdieu, 1977]. According to the knowledge of biological systems, the average value of the characteristics of a metapopulation depends on external conditions. Sometimes small changes in external conditions lead to drastic changes in the metapopulation. The analogy to genetic drift in a human population can be inserting through social mechanisms in models. Some of the most popular examples of theories based on evolutionary models in the 20th century were memetics, followed by sociobiology, and subsequently human behavioural ecology. Mathematical models and computer simulations had already been applied in memetics, however 21st-century research prioritizes other approaches.

8.7. Algorithms and heuristics

Models can use deterministic algorithms to find stable states and assume rational behaviour of the agents. This is a well-known problem in mathematics called the stable marriage problem, which also frequently surfaces in other fields, primarily in economics and sociology[5] (Memorial Nobel Prize in 2012).

[5] Actually one of only two memorial Nobel prizes linked to sociology (the other being [Brian, 1994] on risk perception).

In the problem, each agent tries to maximise its own satisfaction (find the best partner) without respecting the rest. However, the algorithms stop at a point where the agents are more or less satisfied with their partners and cannot change them any more; these are called stable states or Nash equilibria. Usually there are several possible stable states with one set of initial conditions.

8.8. Game theory (e.g. the prisoner's dilemma)

Game theory problems are well-studied in economics and can lead to an understanding of individual human decisions [Axelrod, Hamilton, 1981]. The question of whether to cooperate or defect can be answered according to mathematical rules. There are many types of games from single to repetitive (which can illustrate an adaptive strategy based on history). An interesting case of the problem initially considered is a minority game, when the payoff decreases dramatically when too many players choose the same strategy.

8.9. Language simulations

A separate chapter in trying to capture language phenomena in a formal manner while incorporating a social component is simulations devoted to language dynamics, which over the past two decades have been the topic of both dedicated workshops and articles posted on arXiv and published across scientific journals [Paradowski, Jonak, 2012b]. The presentations and papers have mainly been penned by physicists, and cover such phenomena as language acquisition [e.g. Nicolaidis, Kosmidis, Argyrakis, 2009], language evolution and change [de Oliveira, 2013, and papers from the thematic issue of the *Mind & Society Symposium* on "A multi-methodological approach to language evolution"], language spread [Atkinson, 2011], linguistic typology and diversity [Baronchelli et al., 2010, or papers from the theme issue of *Philosophical Transactions of The Royal Society B* 'Cultural and linguistic diversity: Evolutionary approaches'], the emergence of creoles [Tria et al., 2015], diglossia/bilingualism [Castelló et al., 2007], language competition [Abrams, Strogatz, 2003], naming consensus [Lipowski, Lipowska, 2009], and semiotic dynamics. While most of the early works focused primarily on formal representations and regular-lattice *in silico* experiments that were not infrequently grossly inadequate to the scenery of the 21st century [Paradowski, 2012], over time the field has tried to move from coarser-grained game-theoretic [Nolfi, Miroli, 2010] and agent-based models [Nowak, Komarova, Niyogi, 2001] towards increasingly accurate and sophisticated work based on the results of rigorous data-driven research and empirical studies that recreate the necessary conditions and parameters as faithfully as possible.

9. Networks models

9.1. Complex networks

The most important aspect of complex networks is their topology: who is connected to whom. Each item is a node, connected to others by links (edges). The degree is the number of links attached to a given node. The shortest path length is the minimum number of connections to go through to get from one node to another. The clustering coefficient is a measure of whether the neighbours of a node are connected to each other (at the level of the network it tells us how tightly clustered the individuals are in general), while centrality tells us which nodes (or links) are the most important (e.g. act as 'brokers' between the most individuals or have the highest degree). The 'small-world' concept comes from the fact that most of us are linked by small chains of acquaintances. Community detection algorithms in turn identify the intermediate scale structure by breaking the network up into separate social groups. The preferential attachment property means that a node is linked with a higher probability to a node that already has a large number of links.

Network theory is useful when it comes to the study of nature from a systems perspective, and there are already several examples where it has helped understand the behaviour of complex systems. Genetic regulatory networks, Internet transfer protocols, financial market dynamics and social interactions [Fronczak, Fronczak, 2008], such as those involved in the social diffusion of linguistic innovation [Paradowski, Jonak, 2012a] or second language acquisition [Paradowski et al., 2020; Paradowski et al., subm.]. The most exciting property of these systems is the existence of emergent phenomena which cannot be simply derived or predicted solely from the knowledge of the system's structure and the interactions between their individual elements. A modelling methodology proves helpful in analysing many issues of complex systems properties including collective effects and their coexistence with noise, long range interactions, the interplay between determinism and flexibility in evolution, scale invariance, criticality, multifractality and hierarchical structure. Thereby, complex networks are mostly an artificial concept developed by physicists and mathematicians and (at least in theory) they obey universal rules. Complex network analysis helps better understand social behaviour and determine the degree to which individual agents build a functioning and working system.

9.2. Social Network Analysis

One can currently observe an exponentially growing interest in—and importance of—multi-layered and multifaceted interactions. A method which has been more and more widely applied in approaching complex networks from a societal perspective is social network analysis (SNA), made all the more powerful with the growing arsenal of versatile tools [Borgatti et al., 2009, Christakis, Fowler, 2007]. Conceptually, the social network was introduced in the 19th century by Durkheim (1893), who compared the structure and functioning of societies to biological systems consisting of interconnected components. He concluded that social phenomena are not an effect of the actions of particular individuals, but of the interactions between them. In a similar vein within sociological concepts of structural functionalism and interactionism, Malinowski (1924-44) skilfully combined anthropological study with knowledge from the borders of psychology, mathematics and economics in an attempt to get a better grasp of how societies work. In the 1990s, Jacek Szmatka and his team at Jagiellonian University in Kraków actively participated in the development of Social Network Analysis in one of the first labs of this kind in Europe. Currently, SNA serves as a methodology and set of tools enabling a multifaceted in-depth exploration of interacting systems. The topological properties of networks have been proven to determine dynamic processes above the network level, such as cascades of information adoption or default contagion in culture networks [Dybiec et al., 2012]. Dynamical network models are crucial for dealing with adaptive systems, such as those investigating the relationship between interactions and change of behaviour. In the complexity science paradigm, models have been proposed assuming interactions at the network level [Holme et al., 2012], but an integrative framework is often missing that would combine both theoretical and empirical approaches.

9.3. Homophily, social contagion and external field

There are correlations between the properties of ego and its alters in the network of social ties. The famous Framingham Heart Study Network [Christakis, Fowler, 2007] showed that the phenomenon of obesity is linked to social networks of relations between peers. This means that people who have many friends with similar characteristics (e.g. overweight friends) were more likely to share this feature or faced an increased likelihood of getting it in the future. Surprisingly, the greatest effect is seen among close friends and not among people sharing the same household or sex. This observational study is a very good example helping understand the core processes of network formation (topology) and the processes taking place in the network (spread of norm). Thus, homophily can drive topological network dynamics, but social contagion and external field influence the process on the top of the network.

10. How to model societies?

There are a huge variety of possible approaches to modelling [Pabjan, 2004]. The main classification based on fundamental methodological and technical differences was discussed in the previous sections. Focusing on research questions may lead to yet other divisions [Gilbert, Troitzsch, 2005]. Modelling a target system requires the researcher to choose the model's ingredients, including its form, structure, content, and properties [Hardt, 2016]. If we are interested in correspondence to reality, models can be precise, but very complicated. If we agree on lower precision, models can be simple and easy to analyse. A model can be deterministic or stochastic; interactions can be implied by forces, energy, or rules; variables can be discrete or continuous. Models may be further divided into two major, substantially different types, macroscopic and microscopic [Jarynowski, Nyczka, Buda, 2014].

10.1. Macroscopic models

Macroscopic models attempt to answer the questions "how?" and "how much?" They do not care what happens at the micro level of individual units of analysis, only how the respective average values behave. Here, we are dealing mainly with all kinds of structural equations. This description is similar to the macroscopic description of complex systems, such as in the case of thermodynamics, which includes temperature, pressure, volume, etc. This approach can answer many quantitative questions; it can also generate more or less accurate predictions. One example is population growth. Until the mid-20th century, population growth on Earth was observed to be exponential. In the Malthus model, the world's population keeps increasing in size exponentially (J-curve), while in the Verhulst model, it slows down in a logistic (S-)curve. This also points to the very important issue of the reliability of models. In 1972, economists gathered around the Club of Rome predicted that human population would by now (2018) have exceeded 10 billion in the slowest growth scenario, which was not the case (overestimation of more than 30%).

10.2. Microscopic models

The main disadvantage of macroscopic models is the lack of answers to the question about the causes of the occurrence of the phenomena ("why?"). Enter microscopic models. In the case of social and economic sciences, microscopic models are further divided into:

- microsimulations, where the objects change their state due to deterministic or stochastic rules [Jarynowski, 2010];
- Agent Based Models, wherein the system is a collection of "agents" interacting according to some dependent model rules [Jarynowski, Nyczka, 2013].

An agent, as the basic element of the system, has some characteristics (described numerically) and usually interacts with other agents or external factors. The features of a single agent, as well as the rules, are affected depending on the specific model. One can say that agents are a generalisation of the concept of particles, many-body systems, etc. known from natural sciences.

10.3. Goals of modelling

The application of adequate theoretical methods and the empirical approaches of rigorous sciences such as mathematics and physics to economic and social issues has many faces. From a historical perspective, any science starts by collecting and systematising empirical observations, then moves on to the search for regularities and patterns, which finally results in theoretical formulation which captures the observed behaviours and mechanisms. Although the current scientific community is trying to make progress on all three stages, there are still methodological and conceptual issues that we believe should be addressed in that context. Here, we present the most important open or re-opened methodological issues [Berman, Jarynowski et al., 2016]:

- How to use statistical mechanics to approach social issues? This should be addressed at all levels of social systems, which sometimes lack micro-, meso- or macro-foundations.
- How can conclusions from mathematical and physical models be translated into practical policies that would affect society? Theoretical predictions and conclusions drawn from mathematical models should eventually be put into simple, applicable ideas.
- How should empirical data be collected and shared? Data sources are the key to good science and should be trustworthy and broadly shared.
- Universality versus specificity. The tendency in the hard sciences is towards the generalisation of phenomena and properties, while in so-

cial sciences there is a tendency for particularity and specificity. These natures should synergise in order to make progress and on the one hand come up with new ideas, on the other make them applicable.

11. Conclusions

Researchers devoted to the modelling paradigm postulate that in order to capture the complex and dynamic nature of social phenomena, there is a necessity to unify non-structural factors into universal laws. However, simplistic models are not satisfactory in predicting or even reproducing known sociological observation [Biccheri, 2006; Paradowski, 2012]. For example, many natural scientists claim the universality of the power law degree distribution of social networks. Based on accurate empirical studies [Jarynowski, Buda, Nyczka, 2014], this is considered not true any more. On the other hand, modellers increase our general understanding of the mechanisms of social dynamics and the impact of social interaction. The authors' own exploratory research aims at investigating issues which have so far remained virtually unexplored, by applying computational approaches and a methodological complex systems apparatus in combinations never before successfully carried out *in natura*.

11.1. Main applications of modelling

There are many applications of modelling societies and computational social science. Modelling introduces new concepts to sociology to understand changes, new questions that can be asked, and offers new explanations for phenomena. Rule-based models such as system dynamics are useful in explaining processes and identifying the main factors. Data-driven models such as machine learning can be applied in classification tasks and predict future states. Social Networks Analysis has been found to play an increasing role in connecting human behaviour with the attainment of individuals and has already been part of standard organisation management [Żak, Zbieg, Możdżyński, 2014]. On the other hand, the application of stochastic models based on physicalism may confirm in a scientific manner hypotheses proposed by sociology or aesthetics [Buda, 2012].

11.2. Evolution of modelling in sociology

Modelling is an interdisciplinary venture. The development of sophisticated tools can be observed in every scientific discipline. A growing body of 21st-century research has sought to integrate methods and techniques from different fields of science in order to further understanding of societies and the governing principles therein. Sociology has gained more quantitative, mainly statistical tools in order to process the increasing amounts of data. In consequence of the ability to comprehend the calculus, modelling has begun to play a leading role in investigating social structures through the prism of mathematical formalism. Its growth has been facilitated by the interactions among traditionally defined scientific disciplines. Natural scientists discovered (and social scientists recently agreed) that alongside many differences in research objects and methodologies, there are also similar principles that underlie seemingly unrelated phenomena [Guastello, 2017]. Recent progress in this field can be also attributed to developments in communication technologies. Powerful computer facilities and ubiquitous GPS devices have made possible the collection, processing and storage of data on individual behaviour [Belik, 2008]. Traffic jams and evacuation process models have not only been successfully applied, but have become standard procedure for urbanism and building planning [Helbing, 2016]. In conclusion, modelling already played an important, but still not a dominant role in sociology, and even lived to see dedicated periodicals (e.g. *Journal of Mathematical Sociology* or *Journal of Artificial Societies and Social Simulation*, devoted to sociological enquiries carried out with the application of statistical methods and numerical evaluations). However, it seems that the 'European fortress' of humanistic sociology is less susceptible to mathematical physicalism than the 'Anglospheric' scientific style of thought. The fault also lay with those representatives of European natural sciences (especially in post-communist countries) who preferred to stay in their 'sphere of comfort' rather than try and solve real social problems. During deep Polish social-realism Hugo Stainhaus, one the best Polish applied mathematicians, said to his students at Wrocław University [in a recollection by Roman Duda, p.c.]: *The USA is not richer than Poland, because Poland can afford raising and educating good theorists without having any profits from their work. The USA cannot afford that.*[6] Even today, most natural scientists working in the field of sociophysics have not shown much interest in solving social issues (beyond mentioning them in grant applications) and are only focussing on the 'physical' properties of their models. Due to their disinclination to take off the blinders, collaboration with sociologists may still not be appropriate, at least in the Polish context.

11.3. The future of modelling in sociology

In the authors' opinion, modern computational sociology is mainly about basic research on understanding social systems. The role of sociologists consists in education, influencing public debate and actively enforcing social change. Sociology, and computational social sciences in particular, is a public service for developing better societies [Sitek, 2007]. For example, any possible social intervention can be verified *in silico* before application [Bernoulli, 1766]. Sociotechniques and knowledge form other kinds of modelling outcomes have been already applied and is being increasingly applied by governments and companies. Even some effects of social paradoxes[7] of modelled intervention have been noticed. However, some paradoxes could be mitigated in certain scenarios. The most frequent argument used by some opponents – that the social world is too complicated to be described and explained by models – is untenable. While in the past the modelling of multi-agent systems was limited by the lack of reliable data and computer power, one can easily combine many perspectives in quantitative sociology. Thus, some modelling techniques and complementary tools such as regressions, structural equation modelling, social network analysis, and agent-based modelling, have become necessary part of curricula of sociology programmes at world's leading universities (e.g. Oxford, Columbia, UNSW). There is no other choice for sociology than to adopt the

[6] „Stany Zjednoczone nie są bogatsze od Polski, bo Polskę stać na wykształcenie oraz utrzymanie teoretyków, z których gospodarka nie ma korzyści. USA na to nie stać."

[7] For example the self-fulfilling prophecy or the self-defeating prophecy.

computational social science paradigm to a larger extent, with the possibilities coming from computer power. Deep Blue (the computer chess-playing system) won its first game using heuristic algorithms against a human world champion in 1996, but in 2017 a standard personal computer needs only 72 hours of unsupervised learning to beat any human player. The problem of quantitative and computational deficit [Pabjan, 2004] will lead to the 'End of Sociology' [McKie & Ryan, 2015] as it was traditionally defined. Therefore, computational sociology theories based on modelling or physicalism may be synthetic *a priori* with additional properties such as falsifiability. Such a narrowing down gives sociology the opportunity to get closer to hard science. 21st-century societies are changing at an increasing speed. Much of that change is being driven by developments in Information and Communications Technology (ICT) [Helbing, 2016]; thus, the same ICT techniques as modelling must be applied to investigate societies. The suitability and feasibility of the conceptually novel approaches and applicability of the methodology to still new domains, at least in Poland in sociology [Winkowska-Nowak et al., 2004], [Winkowska-Nowak et al., 200], [Nowak et al. 2009] is highly promising and some results are very encouraging.

Acknowledgments

The authors acknowledge support from COST Action CA 15109 "European Cooperation for Statistics of Network Data Science (COSTNET)". MBP and AJ's research is sponsored by SONATA-BIS grant № 2016/22/E/ HS2/00034 from the National Science Centre of Poland. MBP is also supported by COST Action 15130 "Study Abroad Research in European Perspective (SAREP)". The authors are grateful to the anonymous reviewers for their helpful commentary.

Literature

Abrams, D.M. & S.H. Strogatz (2003). Modelling the dynamics of language death. Nature, 424, 900. doi: 10.1038/424900a

Atkinson, Q.D. (2011). Phonemic diversity supports a serial founder effect model of language expansion from Africa. Science, 332(6027), 346-349. doi: 10.1126/science.1199295

Axelrod, R, W.D. Hamilton (1981). The evolution of cooperation. Science, 211(4489), 4489, 1390-1396. doi: 10.1126/science.7466396

Baker, A. (2008). Computational approaches to the study of language change. Language and Linguistics Compass, 2(2), 289–307. doi: 10.1111/j.1749-818X.2008.00054.x

Albert, R., A.L. Barabási (2002). Statistical mechanics of complex networks. Rev. Mod. Phys. 74: 47–97. doi: 10.1103/RevModPhys.74.47

Baronchelli, A., T. Gong, A. Puglisi, V. Loreto (2010). Modeling the emergence of universality in color naming patterns. PNAS, 107(6), 2403-2407. doi: 10.1073/pnas.0908533107

Baroni, M. & S. Bernardini (2006). A new approach to the study of translationese: Machine-learning the difference between original and translated text. Literary and Linguistic Computing 21(3), 259–74. doi: 10.1093/llc/fqi039

Barr, G.K. (2003). Two styles in the New Testament epistles. Literary and Linguistic Computing 18(3), 235–48. doi: 10.1093/llc/18.3.235

Belik, V. (2008). Transport, disorder and reaction in spreading phenomena. Niedersächsische Staats-und Universitätsbibliothek Göttingen.

Berman, Y., Jarynowski, A., Borysov, S., Balatsky, A., et al. (2016). Manifest on Predictive Data-Driven Research of Social and Economic Systems: From Data Science to Statistical Physics, Nordita.

Bernoulli D. (1760). Essai d'une nouvelle analyse de la mortalité causée par la petite vérole et des avantages de l'inoculation pour le prévenir. Histoire et Mémoires de l'Académie des Sciences, Paris. [English translation by Sally Blower]

Bengio, Y., Ducharme, R., Vincent, P., Jauvin, C. (2003) A neural probabilistic language model. Journal of Machine Learning Research 3, 1137-1155. http://www.jmlr.org/papers/volume3/bengio03a/bengio03a.pdf

Bicchieri, C. (2006). The Grammar of Society. Cambridge UP.

Borgatti S.P., Foster P.C. (2003) The Network Paradigm in Organizational Research: A Review and Typology, Journal of Management 29(6) 991–1013

Bourdieu, P., & Passeron, J.C. (1977). Reproduction in education, society and culture. Sage studies in social and educational change.

Brian, A. (1994). Inductive Reasoning and Bounded Rationality. American Economic Review (Papers and Proceedings) 84 .

Brouwers, L. (2009). MicroSim: Modeling the Swedish Population.

Buda A.(2012) Does Pop Music Exist? Hierarchical Structure in Phonographic Market, Physica A: Statistical Mechanics and its Applications 391 (21), 5153-5159

Carnap, R. (1928) Der logische Aufbau der Welt, Leipzig: Felix Meiner Verlag

Castelló, X., Loureiro-Porto, L., Eguíluz, V.M., San Miguel, M. (2007). The fate of bilingualism in a model of language competition. In: Takahashi, S., Sallach, D., Rouchier, J. (eds.) Advancing Social Simulation: The First World Congress. Berlin: Springer, pp. 83-94.

Christakis, NA., Fowler, JH. (2007). The spread of obesity in a large social network over 32 years. New England journal of medicine, 357(4), 370-379.

Coleman, JS, Katz E. and Menzel, H. (1957). "The Diffusion of a New Drug Among Physicians". Sociometry 20, 253-270.

Comte, A. (1851-1854). Système de politique positive ou Traité de sociologie instituant la religion de l'Humanité, édition originale en quatre tomes. Paris, Carilian-Gœury et V. Dalmont.

de Oliveira, P.M.C. (2013). Dynamic Ising model: reconstruction of evolutionary trees. Journal of Physics A, 46, 365102. doi: 10.1088/1751-8113/46/36/365102

Duch, W., Korbicz, J., Rutkowski, L., & Tadeusiewicz, R. (2000). Biocybernetyka i inżynieria biomedyczna 2000. Sieci neuronowe, 6.

Dybiec, B., Mitarai, N., & Sneppen, K. (2012). Information spreading and development of cultural centers. Physical Review E, 85(5), 056116.

Fronczak, A., Fronczak P. (2009). Świat sieci złożonych. Od fizyki do Internetu. PWN

Galam, S, Gefen, Y. Shapir J., (1982) Sociophysics: A Mean Behavior Model for the Process of Strike, Journal of Math. Sociology 9, 1-13

Gilbert, N., & Troitzsch, K. (2005). Simulation for the social scientist. McGraw-Hill International

Goodman, J.T. (2001) A bit of progress in language modeling. Computer Speech & Language, 15(4), 403-434. doi: 10.1006/csla.2001.0174

Guastello, S.J. (2017). Nonlinear dynamical systems for theory and research in ergonomics. Ergonomics, 60(2), 167-193.

Hardt, L. (2016) On Similarities in Modelling and Metaphorizing Economic Phenomena. Studia Metodologiczne 36, 147-174

Hedström, P, P. Bearman (2009). The Oxford Handbook of Analytical Sociology. Oxford: Oxford University Press.

Helbing, D. (2016). The Automation of society is next. Zurich, ETH.

Holme, P. (October 2012). Temporal networks. Physics Reports , 519 (3).

Jarynowski, A. (2007). Zjawiska krytyczne w przyrodzie. [online]. Wrocław, Uniwersytet Wrocławski. Dostęp: http://th.if.uj.edu.pl/~gulakov/kracylin.pdf [06.01.2018].

Jarynowski, A. (2010). Human-Human interaction: epidemiology. In Life time of correlation. Wrocław: WN.

Jarynowski, A, Nyczka, P. (2013). Dynamic network approach to marriage/divorces problem, ENIC, DOI 10.1109/ENIC.2014.24

Jarynowski, A, Nyczka, P., Buda, A. (2014). Obliczeniowe Nauki Spoleczne w Praktyce, WN: Wroclaw

Jarynowski, A. (2017), Exploring the dynamics and the complexity of human behavior using nonlinear physics methods, PhD thesis (submitted and received positive review), Jagiellonian University in Cracow

Jelinek, F. & Mercer, R.L. (1980). Interpolated estimation of Markov source parameters from sparse data. In Gelsema E.S. & Kanal L.N. (Eds.) Proceedings of the Workshop on Pattern Recognition in Practice. Amsterdam: North-Holland, pp. 381-397.

Jozefowicz, R., O. Vinyals, M. Schuster, N. Shazeer, Y. Wu (2016). Exploring the limits of language modeling. arXiv:1602.02410v2 [cs.CL]

Kamiński, B. (2012). Podejście wieloagentowe do modelowania rynków: metody i zastosowania. Szkoła Główna Handlowa. Oficyna Wydawnicza.

Kant, I. (1781). Kritik der reinen Vernunft

Kant, I. (1784). Idee zu einer allgemeinen Geschichte in weltbürgerlicher Absicht. Berlinische Monatsschrift, 385-411.

Komosiński, M. (2011). Życie w komputerze: symulacja czy rzeczywistość?. Nauka 2, PAN

Konarski, R. (2009). Modele równań strukturalnych: teoria i praktyka. Wydawnictwo Naukowe PWN.

Kossecki J., (1996) Cybernetyczna analiza systemów i procesów społecznych. Wyd. WZiA WSP, Kielce.

Kułakowski, K. (2012). Od fizyki do socjologii i z powrotem. W: A. Chmielewski, M. Dudzikowa i A. Grobler (red.), Interdyscyplinarnie o interdyscyplinarnosci, Kraków: Impuls, s: 99-111

Kwapień, J., & Drożdż, S. (2012). Physical approach to complex systems. Physics Reports, 515 (3), pp. 115-226.

Leśniewska M, Leśniewski P (2016) Analogy-Making in Biology. An Essay on the Comparative Spirit. Studia Metodologiczne 37, 155-173

Lipowski, A. & D. Lipowska (2009). Language structure in the n-object naming game. Physical Review E, 80, 056107. doi: 10.1103/PhysRevE.80.056107

Mantegna, R. (1999). Hierarchical Structure In Financial Markets. Eur. Phys. J. B .

Mayfield Tomokiyo, L. & R. Jones (2001). You're not from 'round here, are you? Naïve Bayes detection of non-native utterance text. In NAACL '01 Proceedings of the Second Meeting of the North American Chapter of the Association for Computational Linguistics on Language Technologies (pp. 1–8). Stroudsburg, PA: Association for Computational Linguistics. doi: 10.3115/1073336.1073367

McKie, L., & Ryan, L. (Eds.). (2015). An end to the crisis of empirical sociology?: trends and challenges in social research. Routledge.

Mazur M., (1966). Cybernetyczna teoria układów samodzielnych. PWN, Warszawa

Merton, R. (1968). The Matthew effect in science. Science, Jan 5; 159 (3810), s: 56-63.

Neurath, O. (1931). Physikalismus, Scientia : rivista internazionale di sintesi scientifica, 50, 1931, pp. 297–303

Nicolaidis, A., K. Kosmidis, P. Argyrakis (2009). A random matrix approach to language acquisition. Journal of Statistical Mechanics, P12008. doi: 10.1088/1742-5468/2009/12/P12008

Nolfi, S., Miroli, M. (eds.) (2010). Evolution of Communication and Language in Embodied Agents. Berlin: Springer.

Nowak, A, W. Borkowski, K. Winkowska-Nowak (Eds.) (2009). Układy złożone w naukach społecznych. Wybrane zagadnienia. Warszawa: Scholar.

Nowak, M.A., Komarova, N.L., Niyogi, P. (2001). Evolution of universal grammar. Science 291, 114-118.

Ossowski, S. (2005). Wzory nauk przyrodniczych wobec osobliwości zjawisk społecznych. W: Marek Kucia (red.), Socjologia. Lektury, Kraków: Znak, s: 28-34.

Pabjan, B. (2004). The use of models in sociology. Physica A: Statistical Mechanics and its Applications, Volume 336, Issues 1–2, s: 146–152.

Paradowski, M.B. (2011). Dekalog analityka danych i infografika – *quid, cur, quomodo*. In: Kluza, M. (Ed.) *Wizualizacja wiedzy: Od Biblia Pauperum do hipertekstu*. Lublin: Wiedza i Edukacja, 338–346.

Paradowski, M.B. (2012). The role of data-driven sociolinguistic research in developing human-machine and machine-machine interaction. In: Obrębska, A. (Ed.) *Practical Applications of Linguistic Research*. Łódź: Primum Verbum, 22–8. https://depot.ceon.pl/bitstream/handle/123456789/3034/palr.pdf

Paradowski, M.B. (2017). Computational native language identification (classifiers). In same, *M/Other Tongues in Language Acquisition, Instruction, and Use* (pp. 69–71). Institute of Applied Linguistics, University of Warsaw.

Paradowski, M.B., A. Jarynowski, K. Czopek & M. Jelińska (2020, subm.). Peer interactions and second language learning: The contributions of Social Network Analysis in Immersion/Study Abroad and Stay-at-Home environments. In: Mitchell, R. & H. Tyne (Eds.) *Language and Mobility: Study Abroad in the Contemporary European Context*. Oxon: Routledge.

Paradowski, M.B. & Ł. Jonak (2012a) Diffusion of linguistic innovation as social coordination. *Psychology of Language and Communication* 16(2) [special issue *Language as a Tool for Interaction*], 53–64. DOI:10.2478/v10057-012-0010-z

Paradowski, M.B. & Ł. Jonak (2012b). Understanding the social cascading of geekspeak and the upshots for social cognitive systems. In: Galton, A. & Z. Wood (Eds.) *Understanding and Modelling Collective Phenomena*. University of Birmingham, 27–32.

Paradowski, M.B., J. Ochab, A. Cierpich & C.-C. Chen (subm.) Accounting for variable L2 success in Study Abroad.

Popper, K. (1959). The logic of scientific discovery, Hutchison & Co.

Rezaee, A.A. & S.E. Golparvar (2016). The sequencing of adverbial clauses of time in academic English: Random forest modelling of conditional inference trees. Journal of Language Modelling, 4(2), 225–244. http://jlm.ipipan.waw.pl/index.php/JLM/article/view/131/143

Schelling, T. (1971). Models of Segregation. The American Economic Review, Vol. 59, No. 2

Sedlak, W. (1980). Homo electronicus. Państwowy Instytut Wydawniczy.

Sitek, W (2007) Paradoksy prognoz socjologicznych. W: Socjologia jako służba społeczna, Kraków: WUJ.

Smoluchowski, M.V. (1918). Über den Begriff des Zufalls und den Ursprung der Wahrscheinlichkeitsgesetze in der Physik. Naturwissenschaften, 6(17), 253-263.

Stemle, E. & A. Onysko (2015). Automated L1 identification in English learner essays and its implications for language transfer. In H. Peukert (Ed.) Transfer Effects in Multilingual Language Development (pp. 297–321). Amsterdam/Philadelphia: John Benjamins.

Sundermeyer, M., H. Ney, R. Schlüter (2015). From feedforward to recurrent LSTM neural networks for language modeling. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 23(3), 517-529. doi: 10.1109/TASLP.2015.2400218

Szaniawski, K. (1971). Metody matematyczne w socjologii, Ossolineum.

Szmatka, J, J. Skvoretz, J. Berger (Eds.) (1997). Status, Network, and Structure: Theory Development in Group Processes, Nowy Jork: Stanford University Press.

Sztompka, P. (1973). O osobliwościach nauk społecznych raz jeszcze. Studia filozoficzne, nr 8 (105).
Sznajd-Weron, K. (2001), grudzień. Opowieści o fizyce egzotycznej. Wiedza i Życie.
Tetreault, J., D. Blanchard & A. Cahill (2013). A report on the first Native Language Identification shared task. In J. Tetreault, J. Burstein & C. Leacock (Eds.) Proceedings of the Eighth Workshop on Innovative Use of NLP for Building Educational Applications (pp. 48–57). Atlanta, GA: Association for Computational Linguistics.
Thomas, W.I., & Znaniecki, F. (1918). Chlop polski w Europie i Ameryce [The Polish Peasant in Europe and America]. Warszawa: LWS.
Tria, F., V.D.P. Servedio, S.S. Mufwene, V. Loreto (2015). Modeling the emergence of contact languages. PLoS ONE, 10(4), e0120771. doi: 10.1371/journal.pone.0120771
Tyler, J. (2018) The 10 most valuable retail brands in the world. http://www.businessinsider.com/most-valuable-brands-in-the-world-for-2018-brand-finance-2018-2 .
Watts, D., & Strogatz, S. (1998). Collective dynamics of small worlds networks. Nature, 393 (440).
Wiener, N. (1950). The Human Use of Human Beings, London: The Riverside Press.
Winkowska-Nowak, K., Batorski, D., & Peitgen, H.O. (2003). Wprowadzenie do dynamiki społecznej. Wydawnictwo Szkoły Wyższej Psychologii Społecznej" Academica".
Winkowska-Nowak, K., Nowak, A., & Rychwalska, A. (Eds.). (2007). Modelowanie matematyczne i symulacje komputerowe w naukach spolecznych. Academica.
Yoon, K., Y. Jernite, D. Sontag, A.M. Rush (2016) Character-aware neural language models. Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence (AAAI-16), 2741-2749. https://www.aaai.org/ocs/index.php/AAAI/AAAI16/paper/viewFile/12489/12017
Żak, B, Zbieg, A, Możdżyński D. (2014). Mapaorganizacji.pl – partycypacyjna platforma badań sieci organizacyjnych, W: Nauki o Zarządzaniu, 1 (18), Wydawnictwo Uniwersytetu Ekonomicznego, Wrocław, s. 100-110.

Andrzej Jarynowski
Interdisciplinary Research Institute
ul. Oriona 15/9, 67-200 Głogów
e-mail: ajarynowski[-AT-]gmail.com

Michał B. Paradowski
Institute of Applied Linguistics
University of Warsaw
ul. Dobra 55, 00-312 Warsaw
e-mail: m.b.paradowski[-AT-]uw.edu.pl

Andrzej Buda
Interdisciplinary Research Institute
ul. Oriona 15/9, 67-200 Głogów
e-mail: adrzejbudda[-AT-]gmail.com