Theory of Complex System, Smoluchowski Institute, Jagiellonian University in Cracow

PhD thesis

Exploring the dynamics and the complexity of human behavior using nonlinear physics methods

Mgr Andrzej Jarynowski

Supervisor: dr hab. Andrzej Grabowski, prof. CIOP-PIB

Wrocław, Kishinev, Kraków 2017

Streszczenie

Praca doktorska "Badanie dynamicznych własności ludzkich zachowań metodami fizyki nieliniowej" składa się z szeregu analiz oraz symulacji komputerowych różnych układów społecznych (głównie w paradygmacie badań eksploracyjnych) podzielonych na 2 kategorie: I- modele epidemiologiczne i układy dynamiczne, II- analizę sieciową i modelowanie danych. Wstęp teoretyczny do załączonych prac własnych stanowi również wprowadzenie do warsztatu badawczego socjofizyki oraz obliczeniowych nauch społecznych.

Do głównych społecznych wyników pracy należą: 1) wskazanie braków polskiego programu profilaktyki raka szyjki macicy na postawie symulacji różnych strategii; 2) propozycja kalkulatorów ryzyka dla chorób zakaźnych; 3) analiza zaraźliwości popularności w Europie, ze wskazaniem na państwa nadawców i odbiorców kultury masowej.

Z punktu widzenia fizyki układów złożonych najważniejszymi rezultatami są: 1) ocena wpływu rzeczywistych tymczasowych aspektów sieci społecznych w różnych modelach propagacji; 2) określenie miar szacowanego ryzyka zakażenia w wieloczynnikowej sieci tymczasowej.

Większość wchodzących w skład doktoratu prac łączy techniki fizyki nieliniowej z pogłębioną analizą opartą na danych jakościowych i ilościowych, gdzie ta ostatnia czynność jest za często pomijana przez fizyków zajmujących się układami społecznymi. Przestawiona praca doktorska stanowi spoiwo między metodologią wywodzącą się z fizyki układów złożonych, a zjawiskami społecznymi

Summary

This thesis contains simulations and analyses of social systems (mostly in exploratory perspective) in two categories: I epidemiological modelling and system dynamics, II networks and data modelling. The theoretical introduction to my own research papers aslo provides the basis of sociophysics and computational social science.

The social impact of this thesis includes: 1) answers to some questions on the best strategy for cervical cancer prevention; 2) infectious diseseases risk calculators ; 3) the intra-European popularity spread mechanisms and classification of countries' roles.

In the field of physics of complex system, main results are: 1) estimating effects of real temporal networks aspects on various propagation models; 2) an infection risk assessment in a real network-type environment.

In conclusion, in most of the presented case studies, the computational aspects of nonlinear physics have been linked with rigorous data-driven and deepend problem analysis, which is too often neglected by many physicists working on social systems. This thesis fills this gap by applying a carefully selected state-of-the-art methodology from complexity science to open social problems.

Table of context

Preface

> "Matematyki stosowanej jako gotowej doktryny nie ma (…) właściwą taktyką było tu przerzucanie mostu matematycznego (…) na obóz socjologów, biologów i lekarzy"
>
> **Hugo Stainhaus** in his talk in 1955

Thesis is describing some of the most applicable tools of complex system in science, by case study of the author's own research. Thesis problem description and solution is organized by main methodological concepts (**I Epidemiological modeling and system dynamics, II Networks and data modeling**). The proposed order comes from few observations. It's not chronically, because networks and data modeling are concepts from end of XX C., while rest comes from beginning of XX c. First of all is application potential, because epidemic models heave already help in coupling with infectious diseases. Networks and data modeling are currently in use by most of public of commercial intelligence. While system dynamics allow only to describe the reality and do not rely on real data as previous concepts. On the other presented order show increasing freedom in modeling. System dynamics gives the most possibility of model inventors, which was already called as „modeling zoo", while data processing is binding in networks and data modeling, where predictability or recall scores are very important. Moreover, presentation or results of epidemic models is even more structuralized, because there are addressed to evident-based medicine. The potential and predictive power of the conceptually novel, cutting-edge interdisciplinary research paradigm will afford answers to both existing but still unresolved, and original burning questions about the role of social (vs. e.g. individual) factors in the investigated processes. In section I 'Epidemiological modeling and system dynamics', I provided solution – based methodology. I conducted also research of few problems that have not been clearly defined yet and papers from section II 'Networks and data modeling' are mostly explanatory.

Please have in mind that proposed divide is only one of possible and many concepts overlap between categories. If I focus on the object of analysis, the 4 categories would appear (Epidemiological modeling {Ist section}, Computational Social Science {Ist and IId Section}, Digital humanities, System Biology), while only first two are in the scope of this thesis and last two have complementary aspects only. Objects of my whole research areas:
A) Epidemiology modeling {Ist section}: temporal networks and spread of epidemics (sexually transmitted infections – STI, hospital infections), agent-based models of epidemic spread in tempo-spatial portrait, diffusion of ideas (study of Stop-ACTA movement in Poland).
B) Computational Social Science {Ist and IId Section}: social networks in culture (music – networks between artists, literature – networks between characters); social change's models of society (marriage/divorce models), adaptive game theory), cultural diffusion, political system reorganization, finding corruptions patterns (in football).
C) Digital humanities: computational application in archeology, law and literature.
D) System biology: computational biology; ecosystem modeling.

MOTTO of Thesis

„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"

Daniel Bernoulli, presented his model of Smallpox in Wrocław, Paris 1766

The author of the thesis:
- is currently the principal investigator in the SIRS-Z ("System Informatyczny Redukcji Szpitalnych Zakażeń) project (IS-2/195/NCBR/2015) studying the spread of nosocomial infections through computer simulations on two layers of contact networks: organizational and empirical in Polish hospitals. This project combines thesis parts I – Epidemiological modeling with II – Social Network Analysis;
- is co-author of application and current investigator in PEERLANG ("How Peer Interaction Mediates Second and Third Language Acquisition") project (2016/22/E/HS2/00034) studying how interactions among students influence progress in second/foreign language acquisition.

Acknowledgment

This work could not come into light without support from many teachers and researchers in field of science (physics, mathematics, computer science and engineering) sociology and epidemiology. First of all, I would like to thank Fredrik Liljeros (Dep. Of Sociology, Stockholm University) for inspiration to solving real problems. He, together with Barbara Pabjan (Institute of Sociology, Wrocław University) was responsible for my sociological training. I'm grateful for meaningful discussions within Forum of Socio/Econo-Physics, especially to Krzysztof Kułakowski (University of Science and Technology, Cracow), Jarosław Kwapień and Andrzej Buda (Polish Academy of Science, Cracow) and Agnieszka Czaplicka (University of Baleares). I was continuously inspired by new ideas because to close cooperation with project "Centre of Applied Mathematics" from town I'm coming from – Gdańsk. In Wrocław – town I'm leaving in I was in constant contact with "Science meets Social Science", "SNA@PWR" (e.g. Anita Zbieg, Piotr Szymański), NLP@B4 (e.g. Maciej Piasecki, Ewa Rudnicka), and UNESCO Chair (e.g. Piotr Nyczka). I was appreciating sharing very good atmosphere by working with Stephanie Boland at YCCSA in York, Hernan Moldani and Xin Lu in Stockholm, Ghenadie Guebec and Liana Cernov in Kishinev, Nataliia Rudyka in Kiev or Damian Marchewka, Jeremi Ochab in Cracow. I want to thanks my official supervisors and co-supervisors mainly for tolerance and allowance for wide freedom of research: Andrzej Grabowski (CIOP - National Research Institute in Warsaw), Fredrik Lijeros (Stockholm University), Leszek Kubiak (Military Institute of Hygiene and Epidemiology in Warsaw), Zdzisław Burda (University of Science and Technology), Janusz Miśkiewicz (Wrocław University), and Wojciech Okrasiński (Wrocław University of Technology), Luis Rocha (Karolinska Institutet), Lisa Brouwers (Swedish Institute of Infectious Diseases), Florentin Paladi (Moldova State University), Richard Morris (John Innes Centre, Norwich), Dan Franks (University of York), Jadwiga Wójkowska – Mach (Polish Assotiation of Hospital Infection), Anita Orzeł-Nowak (Polish Sanitary Inspection). I would like to thanks all my mates with whom I could realize in hobbies like traveling almost all states and region in Europe, fantasy societies and archaeology (e.g. Piotr Szwangruber, Marta Kliś). Of course I'm grateful for support from my family my wife – Anna, daughter Weronika and mother Irena.

Thesis Papers

„It is a common trend in the contemporary science (...), to cross the hitherto fixed, boundaries between disciplines."

Foundation of the UNESCO Chair of interdisciplinary study at Institute of Theoretical Physics, Wrocław 2006

Author of this thesis is first/corresponding author of all paper in main and supplement collection of papers, so author had significant role in data collection, building study, data analysis and writing paper (total input into the paper >=50%). All selected main and supplement collection has been published in reviewed journal in English. Article II.7.Fiction has been including into thesis in original language - Polish.

Main Thesis Content:

I Epidemiological modeling

1) A Jarynowski, A Serafimovic, Studying possible outcomes in model of sexual transmitted virus (HPV) causing cervical cancer for Poland, *Advances in Intelligent Systems and Computing*, Volume 229, pp 129-141, 2014 (I.1.HPV-Pl)
2) A Jarynowski, F Liljeros, Contact networks and the spread of MRSA in Stockholm hospitals, ENIC - conference proceedings, IEEE: DOI 10.1109/ENIC.2015.30. 2015 (I.2.MRSA)
3) A Grabowski, A Jarynowski, Rumor propagation in temporal contact network from Polish polls, ENIC - conference proceedings, IEEE DOI 10.1109/ENIC.2016.19, 2016 (I.3.Polymod)

II Networks and data modeling

4) A Buda, A Jarynowski, Network structure of phonographic market with characteristic similarities between artists, *Acta Physica Polonica A*, vol. 123, no. 3, 2013 (II.4.Music-NET)
5) A Jarynowski, A Buda, Dynamics of popstar record sales on phonographic market--stochastic model, *Acta Physica Polonica B* (PS) 2 (7). 2014 (II.5.Music-StochProc)
6) A Jarynowski, A Buda. Diffusion paths between product life-cycles in the European phonographic market. Control & Cybernetics 45.2. 2016 (II.6.Music-diffusion)
7) A Jarynowski, S Boland, Rola analizy sieci społecznych w odkrywaniu narracyjnej struktury fikcji literackiej, *Biuletyn Instytutu Systemów Informatycznych* 12, 35-42, 2013 (II.7.Fiction)

Papers Not included

I. Books

[1] A Jarynowski, A Buda, P. Nyczka, Obliczeniowe nauki społeczne w praktyce, WN: Wrocław, ISBN 978-83-63089-92-4, 2014

[2] A Buda, A Jarynowski, Life-time of correlations and its applications, WN: Wrocław, ISBN 978-83-915272-9-0, 2010 **(~50 citations)**

II. Original scientific papers (with ministerial points):

[1] A Jarynowski, P Gawroński, K Kułakowski, How the competitive altruism leads to bistable homogeneous states of cooperation or defection, *Lecture Notes in Computer Science*, vol. 7204, pp 543-550, 2012

[2] A Jarynowski, Modelowanie epidemiologiczne na sieciach społecznych na przykładzie zakażeń szpitalnych i chorób przenoszonych drogą płciową, *Studia i Materiały Informatyki Stosowanej* 5 (10), 13-21, 2013

[3] G Morieri, E Martinez, A Jarynowski, H Driguez, R Morris, G Oldroyd, J Downie, Host-specific Nod-factors associated with Medicago truncatula nodule infection differentially induce calcium influx and calcium spiking in root hairs, *New Phytologist*, Nov; 200(3):656-662, 2013 (Impact Factor ~ 7)

[4] A Jarynowski, Modelowanie epidemiologiczne przy wykorzystaniu analizy tymczasowych sieci społecznych, in: Postępy inżynierii biomedycznej, eds. L. Leniowska, Z. Nawrot, InProNa: Rzeszów, 2013

[5] A Jarynowski, J Pawłowski, Nowe metody wspomagania komputerowego w epidemiologii zakaźnych chorób szpitalnych, a dowód *prima facie* w postępowaniach sądowych o zakażenia szpitalne, in: ETYKA wobec współczesnych wyzwań, ed. A.Bobko, Uniwersytet Rzeszowski: Rzeszów, 2013

[6] A Rozanska, A Wojkowska-Mach, A Jarynowski, et. al. Is postpartum SSI surveillance possibile Wituchowo postdischarge surveillance? Results from multicentre surveillance in Polish hospitals, przyjęte do publikacji American Journal of Infection Control (Impact Factor>2)

[7] A Buda, A Jarynowski, Exploring patterns in European singles charts, ENIC - conference proceedings, IEEE**:** DOI 10.1109/ENIC.2015.27, 2015

[8] A Jarynowski, A Buda, M. Piasecki. Smoleńsk – Multilayer network analysis of Polish Parliament 4 years before and after Smolensk crash, ENIC, DOI 10.1109/ENIC.2016.17, conference proceedings, IEEE: 2016

[9] A Jarynowski, P Nyczka, Dynamic network approach to marriage/divorces problem, ENIC, conference proceedings, IEEE: DOI 10.1109/ENIC.2014.24, 2014

[10] A Buda, A Jarynowski, Durability of links between assets in financial markets. Minimal spanning trees and correlations, ENIC, conference proceedings, IEEE: DOI 10.1109/ENIC.2014.18, 2014

III. Reports, preprints, reviews:

[1] A Jarynowski, F Lopez-Nunez, H Fan, How network temporal dynamics shape a mutualistic system with invasive species? arXiv preprint arXiv:1407.4334, report, Umea University, 2014

[2] A Jarynowski, J Jankowski, A Zbieg, Viral spread with or without emotions in online community: http://arxiv.org/abs/1302.3086, 2013

[3] A Jarynowski, H Mondani, J Malmos, PhD review of Xin Lu, 'Respondent-Driven Sampling: Theory, Limitations & Improvements' KarolinskaInstitutet, 2013
[4] A Franzen, A Jarynowski, A Serafimovic, HPV in Sweden, report, Stockholm University, 2012
[5] I Silander, A Jarynowski, A Castro, Modeling Dynamic Competing Cooperating Networks, report, Umea University, 2011
[6] A Jarynowski, Wirtualne aspekty nauki i techniki, Racjonalista, s6648, 2009
[7] A Jarynowski, The establishment of paternity and Hugo Steinhaus results, report, Wroclaw University of Technology, 2008
[8] A Jarynowski, Zjawiska krytyczne w przyrodzie, report, Wroclaw University, 2007
[9] A Buda, A Jarynowski Exploring Arsenic danger awareness in the Polish Copper Basin via network simulation - preliminary results, arXiv preprint, 2016
[10] A Jarynowski, Optimal cervical cancer preventing strategies – model for Moldova, Scientific Dialogue, No 1, p169, 2015
[11] A Jarynowski, D Marchewka. Szacowanie ryzyka zakażenia patogenami szpitalnymi – macierzyństwo 2.0 (wersja beta), w recenzji rozdział w monografii Wydawnictwa Warszawskiego Uniwersytetu Medycznego
[12] A Jarynowski, D Marchewka. Choroby przenoszone drogą płciową w dobie Internetu i E-zdrowia – kalkulatory ryzyka, w recenzji rozdział w monografii Wydawnictwa Uniwersytetu Jagiellońskiego
[13] A Jarynowski, D Marchewka, A Buda. Internet - assisted risk assessment of infectious diseases in women sexual and reproductive health, w recenzji E-methodology
[14] A Jarynowski, A Rozanska, J Wojkowska-Mach, J Rosinski. Polska adaptacja kwestionariusza postrzegania Zasad Izolacji Standardowej – analiza pól znaczeniowych, Zeszyty Towarzystwa Doktorantów UJ – Nauki Ścisłe, 2017, numer 14/1, s 54
[15] A Jarynowski, A Grabowski, D Marchewka: Wdrażanie „Systemu Informatycznego Redukcji Szpitalnych Zakażeń (SIRS-Z)"w polskich szpitalach. Państwo i Społeczeństwo, 2016, nr 4 Supl. 1, strona 27
[16] A Jarynowski, D Marchewka, A Grabowski, Computer-assisted risk assessment of hospital infections: a preliminary implementation in Polish hospitals, Journal of Hospital Infection, 2016, 94S1, strona 128
[17] A Jarynowski, P Nyczka, Single parameter model of marriage/divorces dynamics with countries classification (in review) Physica A
[18] A Grabowski, A Jarynowski**, Influence of temporal aspects and age-correlations on the process of opinion formation based on Polish contact survey, (bioarXiv - 2016 arxiv.org/abs/1607.02588)**

IV. Conference papers:

[1] А Ярыновский, КАКАЯ ФУНКЦИЯ АКТИВАЦИИ СОТРУДНЕЧЕСТВА ОПИСЫВАЕТ ЧЕЛОВЕЧЕСКОЕ ПОВЕДЕНИЕ?, Современные проблемы математического моделирования и вычислительных методов, СЕКЦІЯ 3-4. 1, Ривнэ, 2015
[2] А Ярыновский, Распространение вируса в интернет-сообществе - модели эпидемического, наука, образование, культура. КГУ 7.1, 2015
[3] A Jarynowski, R. Morris, Oscylacje wapniowe. Rozpoznawanie typu sygnalu, w: BioMedinTech - materiały konferencyjne, Gdańsk, 2011
[4] A Jarynowski, A Rostami, Reading Stockholm Riots 2013 using Internet media, in: 6th Language and Technology Conference – conference proceedings, Poznań, 2013
[5] A Buda, A Jarynowski, DYNAMICS OF CORRELATIONS IN FOREIGN EXCHANGE, МЕЖДУНАРОДНАЯ НАУЧНАЯ КОНФЕРЕНЦИЯ, посвященная 20-летию экономического образования в Бельцком Университете «Алеку Руссо» 2015
[6] A Jarynowski. Collapse of cooperation and corruption in a mathematical model within game theory including Moldovan case study (Homo Sociologicus vs. Homo Economicus), МЕЖДУНАРОДНАЯ НАУЧНАЯ КОНФЕРЕНЦИЯ, посвященная 20-летию экономического образования в Бельцком Университете «Алеку Руссо», 2015

[7] A Jarynowski, M Klis, Socio-economic models of divorces in different societies, in: Applications of Mathematics In Biology and Medicine – conference proceedings, Krynica Morska, p. 60-64, 2012
[8] A Jarynowski, J Jankowski, A Zbieg, Stop Acta Natural vs. artificial viral spread within online community, E-methodology, 2, p71, 2015

V. Popular science papers

[1] A Jarynowski, Socjofizyka, czyli służba społeczna fizyki i nauk matematyczno-przyrodniczych, w: Nauka Prowadzi w Przyszłość, s.99, Kraków, 2014
[2] M Klis, A Jarynowski, Przykłady eksponowania szczątków ludzkich w muzeach, w: Historia w muzeum. Formy i środki prezentacji 1, s. 48-51, Bydgoszcz, 2013
[3] A Jarynowski, Modelowanie epidemiologiczne: klasycznie i na sieciach, SeMPowisko, Kraków, s. 25-28, 2012
[4] A Jarynowski, Testy samosprawdzajace, materiały Ars Docendi UJ, 2012
[5] A Jarynowski, Zadania matematyczne dotyczące Solca Kujawskiego, Muzeum Solca Kujawskiego, 2012
[6] A Jarynowski, Finansowanie naukowych wyjazdów zagranicznych, materiały szkoleniowe TD UJ, Kraków, 2012
[7] A Buda, A Jarynowski, Agent-based modeling, complex networks and system dynamics– practical approaches, PUET, Połtawa, 2015.
[8] A Jarynowski, A Grabowski, MODELOWANIE EPIDEMIOLOGICZNE DEDYKOWANE POLSCE, ISBN 978-83-942807-9-6, PORTAL CZM, 2015
[9] A Jarynowski, WYKORZYSTANIE MODELOWANIA MATEMATYCZNEGO WYWODZĄCEGO SIĘ Z NAUK PRZYRODNICZYCH W SOCJOLOGII ANALITYCZNEJ, ISBN 978-83-942947-5-5, PORTAL CZM, 2015
[10] A Jarynowski, A Buda, Hierarchical representation of socio-economic complex systems according to minimal spanning trees, PiJ 1, 37-45, 2015
[11] A. Buda, A. Jarynowski, The global phonographic market: record labels, artists and fans in the internet era, e-methodology 2, p96, 2015
[12] A Jarynowski, Human-Human interaction: epidemiology, in: Life-time of correlations and its applications, WN: Wrocław, 2010
[13] A Jarynowski, Anomalous interactions in network of Polish Football League, in: Life-time of correlations and its applications, WN: Wrocław, 2010

Introduction: bridge between science and social science

"Die Wahrscheinlichkeitrechnung (…) Bereich **sozialer** und biologischer **Vorgaenge** angewendet wurde, hat sich (…) ein ueberaus wechtiges Anwendungsgebiet erobert: die Physik"

M. Smoluchowski, Cracow 1918

Recently, the knowledge of complex system tools for sociology and medicine, such as complex networks, have undergone an accelerated growth, however all models of such systems are incomplete without real data, especially register-based. Simulations are relevant only if there is some data, of which the parameters can be calibrated to run simulations. The requirement to calibrate the parameters encourages cooperation between various registering institutions, which, in turn, exerts a pressure on collecting data for simple analysis by numerous researchers working on new models and using complex tools, often from other disciplines. In my thesis, I applied the concepts of physics to social science, but we cannot forget that physics had been highly influenced by social science as well. The topic of my PhD dissertation is the application of complex system tools in medicine and sociology to provide a bridge between theoretical physics, mathematical modelling and data exploration with social systems and epidemiology. The use of data and computer simulations to understand and modify social processes is an innovative project area. The development of IT tools, which can be used by decision makers in health and social services, has been listed among the EU research activity priorities.

Role of physics

To describe reality is the main goal of physics, as a base science of Aristotle or Comte. The contribution of physics is not limited to the material, but includes all aspects of life, the non-material, the essence. Even though physics uses mathematical, statistical and computational tools, it is at the same time tied to reality. This thesis, as well as in most computational research focused on social phenomena, will discuss some issues that are well-developed in physics:

NETWORK SCIENCE (social network analysis — empirical (Lazer et al. 2009), complex networks science — random (Barabasi et al., 1999), both influenced by spin glasses (Mezard, et al. 1987)

COMPLEX SYSTEM SCIENCE (emergence phenomena (Green et al., 2009), fractals, universal laws of nature, phase transition, critical phenomena (Jarynowski, 2007), anomaly detection (Jarynowski, 2010)

COMPUTER SIMULATIONS AND MATHEMATICAL MODELS (using the model of magnetism (Sznajd et al., 2001), Monte Carlo simulation, Metropholis algorithm)

DATA ANALYSIS (time series, hierarchical representation (Czaplicka, et al., 2013), statistical physics concepts like entropy, random matrices theory)

DYNAMICAL MODELS (fluid dynamics — crowd dynamics, diffusion (Rogers, 2005), particle allocation — wars and inequality)

Physics and social science: Short history of time

The application of physics in social science seems to be quite a new field of science or at least it has become officially recognised as a part of physics recently. It is not surprising that computer simulations and mathematical methods derived from statistical physics are used to model social behaviour. Computer simulations of physical systems have inspired scientists to play artificial societies. The current era of sociophysics begins with the work of Schelling (Schelling, 1971), and later Galam (Galam et al., 1982), however the philosophical background is older.

The term "Social physics" comes from Auguste Comte (first part of 19th Century) and, along with the trends of „mecanique sociale", it reflects the positivist approaches of world scientification. However, the development of modelling in social sciences was later suppressed by the boom of the humanist and neo-positivist philosophy of inter-war science (the 1920s and the 1930s) and weakened by the subsequent discovery of quantum mechanics, which partially changed the classical logic. Then, 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 resulted a "big divorce" of the Polish sociology and the techniques of natural sciences following (Ossowski, 1962) and (Sztompka, 1973). Perhaps one of 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 system transformation in Poland. However, thanks to the development of computer calculation techniques, the methodology developed by physicists was proved in application in other, sometimes highly remote, areas (like biology).

In 1918, when Marian Smoluchowski concluded that probability is the central problem of modern science, he introduced the methodology of social science into physics. If physics had failed to learn from sociology at that time, we do not know for how long physics would have remained in the utopia of determinism and reductionism. Moreover, the development of computer technology enables us to analyze large sets of data produced by people.

Nature provides us with many interesting examples of phenomena that can occur in humans. For example, the model of coordinated movement of animals: flocks of birds or swarms of fish. The animals that move in such patterns do not usually have a leader, and the knowledge is local, but in most cases the herd is moving in the right direction. This model can provide the gregarious properties of human communities if the variable trait represents a position of the animals.

In the 1950s, the physics of complex systems became the frontier of physics, thanks to the work of the (Bertalanffy, 1950) — "open systems" and later (Anderson, 1972) — "more is different". Bertalanffy criticised the attempts to describe living systems on the basis of closed thermodynamics systems and requested a completely new approach, taking into account their openness and exchange of energy and information with the environment.

The sociologists like Durhain argued that changes in the evolution of social systems occur not only in increments, but also in rapid way, which corresponds to far from equilibrium statistical physics. Thus, the fundamental question of social theory is the conception of human action (Granovetter, 1985), where not only internalised norms and values are taken into account (oversocialised — collective behaviour), but also the self-interest of the actors (undersocialised — linear changes) in the individual decision-making process (linear changes). The presented approaches to describing interactions correspond to Newtonian mechanics — undersocialised system and to Statistical Physics — oversocialised systems.

A real boom in field of sociophysics occurred around the year 2000, when a lot of new models were developed, aiming to recreate the entire spectrum of social phenomena. It is worth mentioning that Polish scientists have been some of the most widely represented in the process of nation building sociophysics and this was a thriving discipline in Poland, as evidenced by the existence of the Section of Physics in Economics and Social Sciences at the Polish Physical Society, the UNESCO Chair of Interdisciplinary Studies and S3 (Science meets Social Science) both in Wrocław.

Object of interest in complex social systems

Social reality has been a fascinating area for mathematical description for a long time. So far, however, most attempts at building mathematical models for social issues resulted in inapplicable simplifications. Since mathematical models became more realistic, they slowly began to be used in the prediction (social engineering). The most interesting feature of social science research subjects — people, organizations, societies — is their complexity (set of mainly non-linear interactions). For clarification, natural science is not monolithic, and so, for example, a branch of science dealing with the example described above is called: system science (by computer scientists), systems dynamics or non-linear dynamics (by mathematicians), or complex systems (by physicists). Sociologists have been wondering how to measure social mood, or other characteristics of the society, as well as how to predict and understand social change. Natural scientists and mathematicians have used the results of these measurements and included them into models usually based on an analogy with a known natural phenomenon. The great number of models and concepts developed throughout the 20th Century, resulted in independent fields of science: social cybernetics (Wiener, 1950) or younger, analytical sociology (Hedstrom, 2009).

Sociophysics formed at the interface of sociology (social science) and physics (natural science) is supported by mathematics and computer science, but it is based on the belief that there is an analogy between the particles, atoms, molecules and living organisms, humans, and even entire ecosystems and societies. The thesis postulates and research directive seems to be different from the sociological perspective, including the role of experiment, etc. (Ossowski, 1962), and for a very long time natural science methodologies were reduced to testing statistical hypotheses and deriving regressions in social sciences (Sztopka, 1973). However, a breakthrough in the practical use of new techniques for social sciences has only occurred in the 21th Century. Then, the volume and complexity of digitalised information on social activities forced a change of paradigms in social studies. The need for understanding social phenomena emerged as a natural step in detailed studies, within a broader, interdisciplinary perspective. Sociophysics allows building more comprehensive, multidisciplinary planes of social knowledge.

Computational Social Science was proposed to describe different modelling approaches combined with data exploration techniques. This reveals complex relations between agencies in various systems. Statistical and computational visualisation methods and quantitative techniques are currently fully exploited in research on social systems. It is not only basic quantitative methods like statistical regressions and data-mining procedures, but a range of various innovations that brings a better understanding of the processes and rules of the social systems evolution and, of course, the most important objective — prediction.

Main methods of investigation in complex social systems

Mathematical methods and computer simulations are becoming increasingly popular and successfully applied in the explanations of the phenomena observed in the real world social, economic and biological systems. In this chapter, I briefly introduce two computational methodologies, agent-based modelling (in NetLogo) and system dynamics (in Vensim), and one analytical tool — network analysis (in Pajek), which allow to represent complicated and complex non-linear social systems (Pabjan, 2004). Of course, many various approaches remained unmentioned, like decision trees or the game theory.

Differential equations (used by system dynamics) were applied to describe and predict those phenomena first (I solve some epidemic equations analytically, numerically or by simulations in (Jarynowski, 2016)), but recently agent-based models appear even more often. Sometimes one problem could be solved in two ways. Because computer simulation has changed the world of mathematical modelling, agent-based models do not always give better predictions and some hints for decision-makers, even parallel to the development of numerical methods for differential equations. On the other hand, differential equations allow us to understand the core process, which could be missing in the agent-based approach. As a result, both perspectives are common among specialists and depending on the theoretical or applied aspects, percentage representations differ.

System Dynamics (SD). There are several types of computer software used for SD, like Dynamo, iThink, Stella etc., but as an example, I choose Vensim — the environment of my papers on sexually transmitted infections I.1.HPV-Pl and (Jarynowski, 2016). The graphical notation allows a non-mathematician to build and solve sets of differential equations (Brouwers, 2009). The dynamical variables are represented as stocks and rates of change as flows [Fig. 1 left].

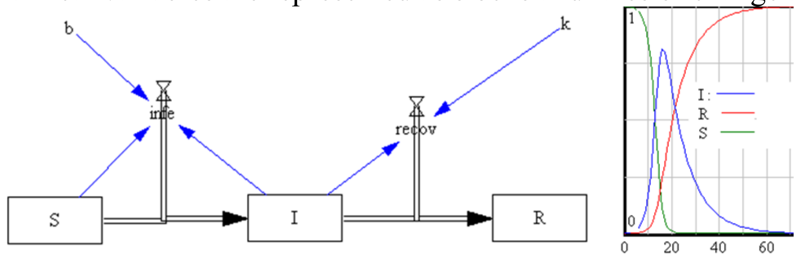


Fig. 1 [left] Susceptible, Infective, and Removed epidemic model (explained in details in the chapter on epidemiological modelling) in Vensim with parameters: *b*-infectivity, *k*-recovery rates and typical epidemic curve [right]

This type of diagram shows the relationships among variables [*Fig. 1* left], and the output presents change over time [*Fig. 1* right] with causal looks (representing more kinds of non-linear relations, like feedbacks). In this numerical approach, equilibrium conditions or systemic variables are traced in time.

Agent-based models (ABM). ABM is a computational technique used for experiments with artificial systems populated by agents that interact in non-trivial ways. I chose Netlogo and I made some simulations in this environment for example in my marriage/divorce project (Jarynowski, Nyczka, 2014), but other toolkits are also examples, like Swarm, RePast, MASON. In NetLogo, the turtle is a representation of an agent (an autonomous, interacting entity). A patch is the elementary spatial unit in the NetLogo grid [*Fig. 2*]. The goal is to imitate real patterns by running an (often computerized) ABM under different treatments and conditions.

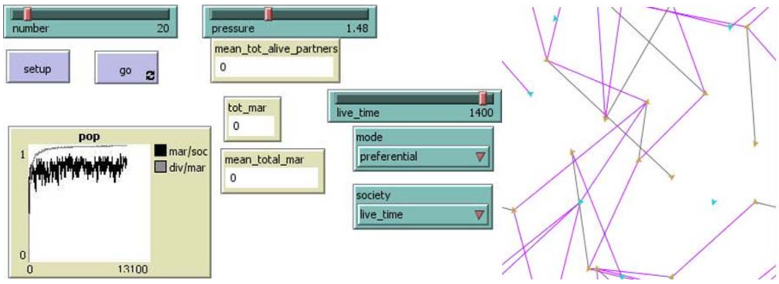


Fig. 2 Dialog window of the marriage/divorce model in society (from Netlogo), where links represent the network of marriages (Jarynowski, Nyczka, 2014)

Dynamical Systems and Chaos. The evolution of dynamical systems is ruled by an implicit relation of input and output. If the behavior of dynamical systems are highly sensitive to initial conditions it can be described in terms of chaos. I have provided analysis on the nature of calcium oscillation response in plants symbiosis with Lapunov test for determinism (Jarynowski, Morris, 2011).

Complex networks. We will focus on the most applied aspect of it: Social Network Analysis (SNA), which has its origins in the graph theory. Here, the topology is the most important: who is connected to whom [Fig. 3]. Each item is a node. The relations between them are connections (links). The degree is the links of given node. The shortest path length is the minimum number of connections to go through to get from one node to another. The clustering 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). The 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 identify the intermediate scale structure as different social groups. The preferential attachment property: a node is linked with higher probability to a node that already has a large number of links.

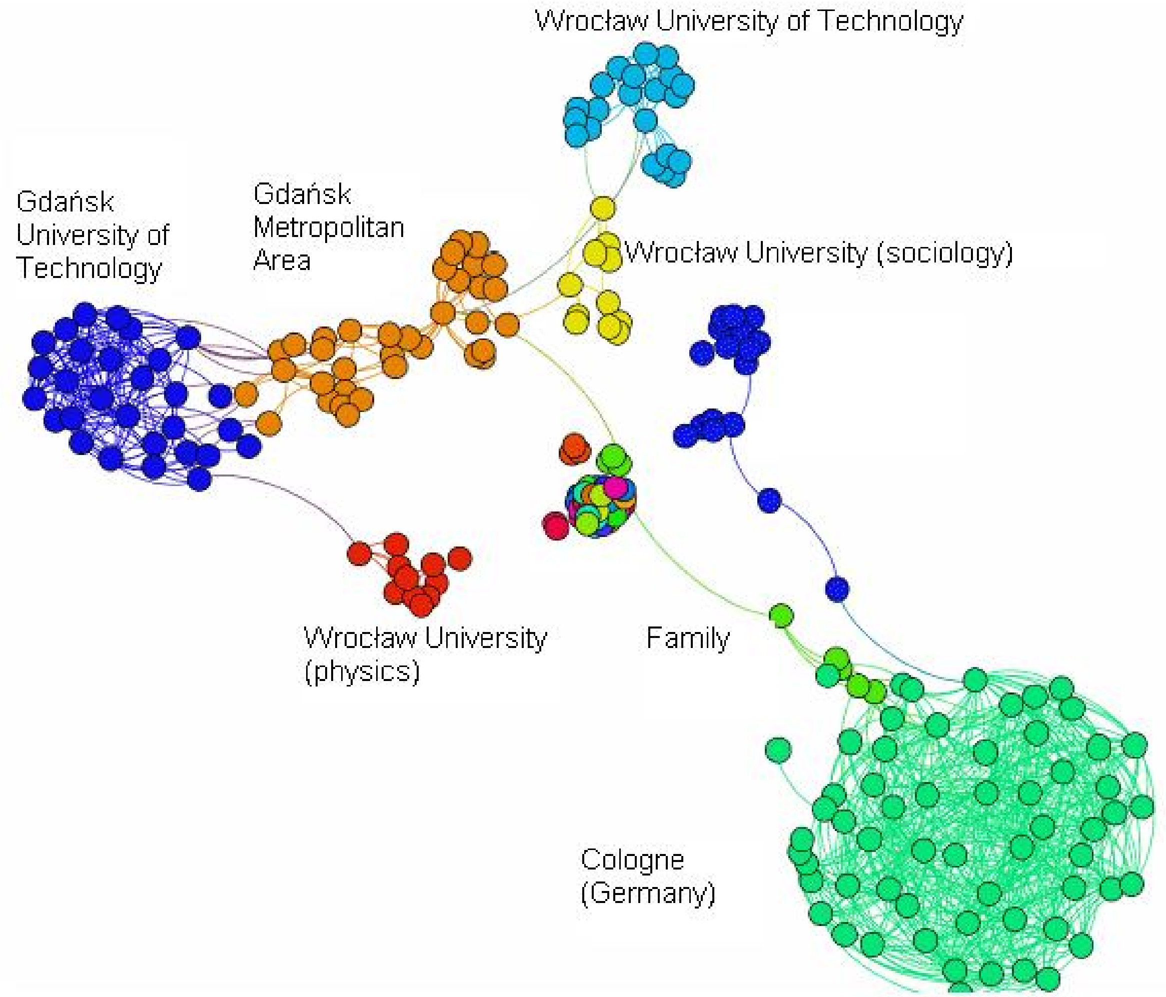


Fig. 3 Facebook network of thesis author with annotated communities

Statistical physics. There are many developed techniques to analyse analytically and some approximated methods of solution derived in physics and used by me in this thesis. Within game-theoretical papers (Jarynowski, 2015a), some analytical approximations was performed with mean field approach. In models form mentioned papers, main signal W – reputation of a player evolves in time according to stochastic rules depending of his previous strategy (cooperation or defeat with corresponding probabilities P). The general analytical form is not known, but some approximation can be done for averaged evolution:

$$\frac{d<W>}{dt} = \Delta W\_increase * P\_cooperation - \Delta W\_decrease * P\_defeat$$

Such a form allows me to use stabile point analyse to figure out phase plane of players characteristics. Presented mean field approximation formulation converges in this example to quasi - master equation.

Characteristic methods in complex system analysis

Control theory and cybernetics. The general goal is discovering patterns and finding principles that can be apply to prediction or control (Poczybut, 2006). Outside of central point of system theory – feedback loops, many another technics are used as neural networks, artificial intelligence, and artificial life, swarming and floating algorithms. There is a very important issue emergence - a process whereby interactions among basic entities exhibit properties different than simple sum of entities. In physics society it is called after Anderson "more is different". In two papers of mine, collective behavior is shown: in case of duration of norm change (Jarynowski, Kliś, 2012) and in division between Christmas and regular consumers (II.5.Music-StochProc). Information version of Entropy (H_I) can also describe some collective properties. It only takes into account the probability of observing a specific events p_i. General Shannon Entropy is defined:

$$H_I = -\sum_{i=1}^{N} p_i \ln p_i$$

Entropy concept was used by me in corruption detection in Polish Football League (Jarynowski, 2010a).

Critical phenomenon and self-organization. The critical phenomena theory (Jarynowski, 2007) has been applied to solving problems in behavioural, social and political sciences, although its roots are traced to mathematics and physics in the early 20th century. Tools developed primary to Earthquakes have been applied to my analysis on music industry (II.5.Music-StochProc) and calcium oscillation (Jarynowski, Morris, 2011). Catastrophes, self-organizing processes and chaos have been broadly overlapping in social and natural science within the complex system paradigm. A turbulent flow theory of fluids and gases can describe in many socio-economic systems, where equivalent of Reynold number is high enough. In the evolutionary game theory study, I have observed a phase transition in cooperative behaviour (Jarynowski, 2015a). The trajectories of analyzed dynamical system were also checked for deterministic chaos (Jarynowski, 2015d). The mean field approach was also applied and the master equation was derived (Jarynowski, Gawroński, Kułakowski, 2012).

Time series analysis. By collecting data on human activity, linear and non-linear techniques of data analysis can be applied. Stochastic processes and data-mining allow us to investigate system properties. I analyse those data sets, (re)construct (simulate processes) with similar characteristics, like distributions etc., to predict future states. Series can be analysed from a hierarchical (Mantenga, 1998) or fractal (Kwapień, Drożdż, 2012) perspective to explore the features of processes. Aberrations of autoregressive processes for stationary and non-stationary analysis play a fundamental role in the record sales project (II.4.Music-NET and II.5.Music-StochProc), in which some questions are posed:
- Understanding the criticality of sales (why Hurst exponent is less than 0.5 - analogy to turbulence);
- The construction of a stochastic model describing the empirical time series (e. g. understanding and describing the seasonal customers' behaviour).

In author personal opinion, the most important application of time series analysis from human being perspective is medical applications. From individual patient perspective, two significant areas can be indicated: brain and heart (Makowiec, et al., 2006).

The most popular models applied for predicting future values of time series are mainly autoregressive models with error correction mode, integrated moving average and empirical nodes. The standard model to describe signal X_t at given time point is Autoregressive Moving Average (ARMA) by combining Autoregressive and Moving Average processes. We need series of $X_{t-1} \dots X_{t-p}$ are previous observations of variable X in time points t -1 up till t - p; $\varepsilon_t \dots \varepsilon_{t-q}$ are previous values of error terms in time points t up till t–q; set of model parameters Φ ad θ. The general ARMA(p,q) process is defines as:

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \ldots + \phi_p X_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \ldots - \theta_p \varepsilon_{t-q}$$

Choosing an adequate model for given data is one of the most difficult tasks in time series analysis. There are many model selection criteria and, different researchers choose different criteria to rely on. Akaike's Information Criterion (AIC) is one of the most commonly used model selection criteria. This criterion selects the model that minimizes the negative likelihood, including a penalty for over parameterization. AIC can be obtained by using the following formula:

$$AIC = -2 logp(L) + 2p$$

Following 'frequencies statistics' methodology has been used in my signal processing, e.g. in (Jarynowski, Morris, 2011). I have also applied 'Bayesian statistics' as Markov Chain- MC based estimation algorithms, mostly with shorter memory model (MC(1)) in I.2.MRSA. One would perhaps assume that Markov chains of second or several orders would produce better outbreak estimates with higher accuracy, as aggregated record sale shows periodicity. However, I claim, that higher order MC algorithm would not increase predictability significantly. I want to compare other predictive algorithms, such as dynamic Bayesian networks, neural networks and cellular automata, and to evaluate the feasibility of predicting aggregated sale with individual- agent based spread models.

Additionally, econophysics introduced concept of log-periodic oscillation to stock exchange time series analysis and even author of this thesis analyzed from this perspective Polish Stock Exchange (Jarynowski, 2007). It's just a simple mechanistic technical fitting procedure to a given function (without any fundamental meaningful interpretation), but it seems to work well, especially in rapid change of global trend. It is important mention, that Polish physicists predicted Lehman Brothers collapse crisis with weekly precision in 2007, sometime in advance.

Regressions and multivariateanalysis – artificial intelligence "in peoples home"

The logistic regression is a mathematical model which allows us to describe the influence of many variables X = (X1, ..., Xi, ...) on the depending variable Y(X). Multivariate analysis consists of those statistical techniques that consider two or more related variables as aentity and attempt to produce an overall result taking the relationship among the variables into account. Regression analysis is widely used for risk prediction and forecasting future status. In this situation, support for sexual and reproductive health decision-making is possible already with artificial intelligence (Duch, et al., 2000). For infectious diseseas our depending (explanatory) variable could be defined as probability of infection. I have develop open source infection risk calculators for patients and healthcare professionals as apps for hospital acquired infections (during child-delivery) and sexually transmitted infections (like HIV). After filling in a brief online survey, users will maintain numerical probabilities based on our algorithm. The algorithm based on the response to the sexual act (information about the type of contact) and the person with whom this relationship occurs (contact person information). These parameters determine the probability of infection.

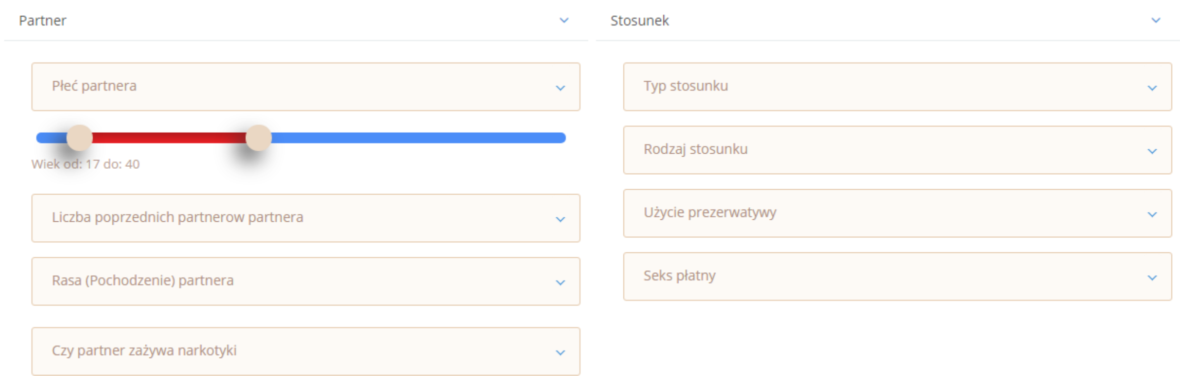


Fig. 4 Conditional probability regression model for infection risk calculator of sexually transmitted infections – part of questionnaire (Jarynowski, Marchewka, Buda, 2017)

Child-delivering app requires personalized hospitalization history of pregnancy described by questions represented by quantitative and qualitative variables. For risk of infection depending on the level of exogenic variables (qualitative and quantitative), was explained by tree logistic regression in a model of multifactorial covariance analysis.

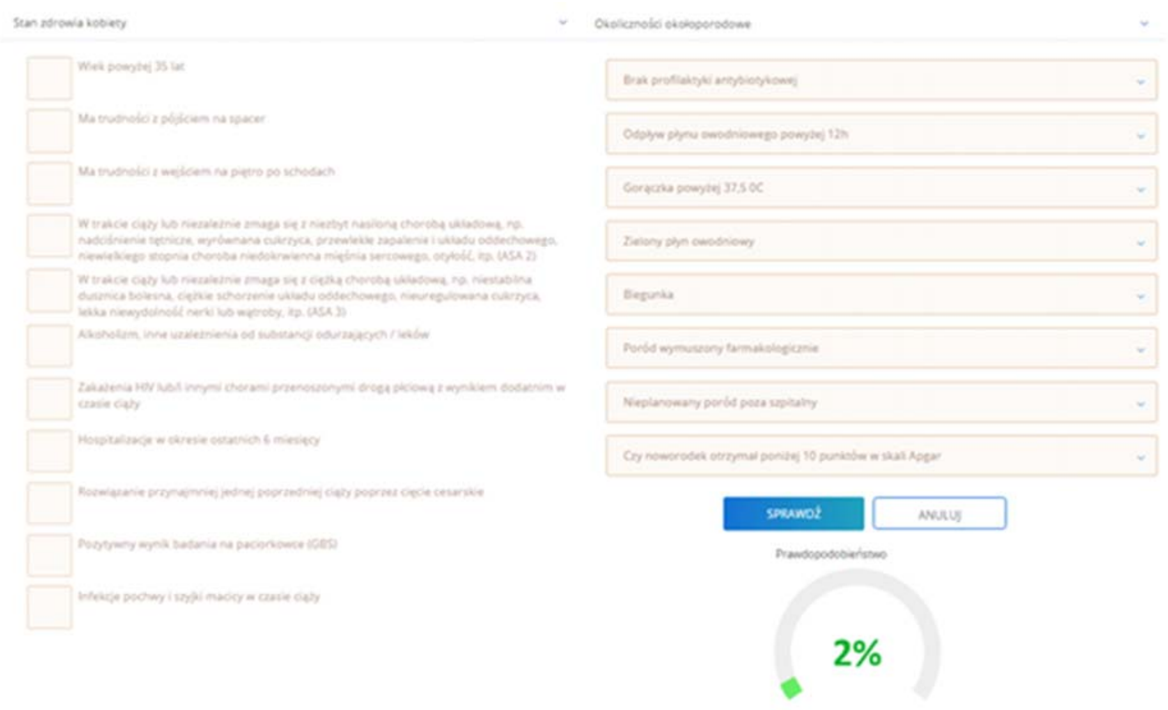


Fig. 5 Tree Regression model for risk calculation of hospital infections during childbirth – part of questionnaire implemented as a web-query (Jarynowski, Marchewka, 2017), firstly implemented in passive surgery site infection registry (Różańska, Wójkowska-Mach, Jarynowski, et al., 2017)

The algorithms that I developed are a subject to criticism from the clinician community (representing medical decision makers) that do not accept the democratization of e-solutions in the patient-doctor relationship. The Childbirth calculator may be a part of "pregnancy" apps. Proposed "gamification" that makes user guess the risk does not provide more knowledge than giving a system in which everyone can acquire, with a large degree of autonomy, competence for initiating processes of exploitation of own creative personality.

Missing data and data preparation. Data should preliminarily processed before modeling starts. In a kind of interdisciplinary research, as presented in thesis, problem with linking databases could appear. For example, there could be duplicated records, records with different id during two consecutive system integrations. Morover, the fields could be filled-in incidentally in a nonuniform fashion, so as a consequence, the coding must be applied. Both of those problems appear in collecting data wihin "SIRS-Z" project. How to treat missing data is just another issue. The most common solution is to delete record. However, if it seem to be Missing Not At Random or there is too many Missing Data, then other procedures could apply. In covariance analysis (Różańska, Wójkowska-Mach, Jarynowski, et al., 2017) lack of data was taken into account as a basic category for "MD" codes (missing data) from the original database. Digital epidemiology (technique developed in this thesis) and data processing is a big issue, because the ability to analyze large amounts of data (Big Data) at low cost may optimize medical processes. We live in an information society where individual medical data can be used to improve epidemiological safety, but if data or model are wrong, etical statement "*Primum non nocere*" (first, do no harm) is broken.

Game theory (the prisoner's dilemma). Game theory problems are well studied in economics and individual human decision can be understood this way (Axelrod, 1981). Cooperate or defect: these questions can be answered according to some mathematical rules. There are many types of games from single to repetitive (which can illustrate adaptable strategy - based on history). An interesting case of the problem initially considered by physicists is a minority game. The payoff decreases drastically when too many players choose the same strategy.

Let us explain the basic Prisoner Dilemma in details. Two criminals were caught by the police. If prisoner 1 confesses (defeat by betraying his colleague) and prisoner 2 denies accusing prisoner 1 (cooperate - will try to fair to his colleague). So prisoner 1 will be acting as a witness against prisoner 2 and go to prison for a short time (*T* years), while the prisoner 2 will get a full (S years) of sentence (and vice versa). If both deny (cooperate) and nobody confesses, they will be acquitted (R=0 years). If both defeat (betray each other's), both receive the same average sentence (*P* years). The order of payments in such a game would be in order from the most profitable: R=0 years $<T$ years $<P$ years $<S$ years. Such a system has an equilibrium point, in which cheating is the optimum strategy. Many other problems could be described in such a setup and in paper players are agents on the economic market and games represents transactions. I will focus on the optimisation of "social" goals, in which the agents pursue only their own material self-interest with exception to altruism.

Table 1 Payout matrix of standard Prisoner Dilemma

Players ½	C (cooperation)	D (defection)
C (cooperation)	R, R	T, S
D (defection)	S, T	P, P

However, I use the concept of the prisoner's dilemma in my own projects, but instead of the payoff matrix, the dynamic rules of changing reputation and altruism will drive the dynamics (Jarynowski, 2015d).

Social animal and ecosystems are example in which a network and modeling perspective is important to understand systems behavior. Such networks are defined by nodes-species or individuals and links-interactions like as seed dispersal or intraspecific competition. Many of these researches have been already applied to social system like Dunbar's number (biological limit of contacts what people can handle). Biomathematical modeling has much longer tradition, than mathematical sociology or sociophysics. Ecological networks allow us to study the structure and function of ecosystems and gain insights on species resilience/stability. The study of ecological networks is usually a snapshot focused in a limited specific range of space and time, prevent us to perceive the real dynamics of ecological processes. I developed a simulation platform that shows inference between temporality of ecological networks with competing mechanisms. (Jarynowski, Lopez-Nunez, Fan, 2014).

Inverse problem. There is a lot of attention to the topic of inverse problems e.g. for heat equation (finding initial and boundary condition from time evolution measurements in given sensor points). Due to the fact that its diffusion counterpart models so many phenomenons not only in physics, but also in social, solving corresponding inverse problem will always be in the demand. It is particularly interesting because it has a wide range of real world applications. An example here would be problem of identifying the source of infection in outbreaks. The idea of the inverse epidemiology is to some extent original and may throw some light on the general relationship between patterns of social behaviour and epidemic. A rigorous approach to analysing the data, using the methods of multidimensional network analysis, and yielding quantifiable results, is commendable. In my study, I have tried to find networks structure from infection datasets. I have proposed to use some of the best described epidemiological studies in history of civilization (in my mind) and run retrospective analysis of great epidemics: Wroclaw (smallpox, 1687-1691) and Warsaw (plague, 1624-1626), where both register and excavation based data allow us to proceed analysis (Jarynowski, 2014). This study is based on pathogen spread between people of given city, once infective persons arrive there. Such a model has sense only if there are some data to calibrate model parameters (like contact rates within investigated society). While Bernoulli's main objective (Wroclaw dataset) was to calculate the adjusted life table if smallpox were to be eliminated as a cause of death (from differential equation point of view). I try to learn age contact structure from my simulation [Fig. 6 right]. Hence Warsaw plague was well described by city major, who was a physician, much more information is available on geographical (evidences of infections by districts) or social structure (evidences of infections by professions) level. We propose an epidemiological model, where agents would represent citizens of Warsaw during the plagues linked to given location and characterized by their job attitudes. Network of contacts is represented by exponential random graph models. Preliminary results show geographical clusters of districts as well as professions particularly exposed to infection and burial in given cemetery. I suggest other potential application of my tools in 'visualization' of social complexity within museum exhibitions of post-epidemics mass grave excavations.

Evolutionary models explain many biological phenomena (such as the extinction of species) and genetic algorithms have successfully found their applications in computer calculations. This approach is also used for social processes, in which the time scale is tens or hundreds of years. My models on divorces (Jarynowski, Nyczka, 2016) or demographical change ran (I.1.HPV-Pl) for around 50 years. Biology explains the formation of different population structures, speciation through the rivalry of two forces of nature - natural selection and genetic drift. Depending on the model, different analogies are used, but for example, gene drift corresponds to the transmission of cultural and natural selection is also observed in society 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 (Axelrod, 1995). In addition to vertical transmission from parents to descendants, there is horizontal transmission (the media, schools, etc.), of which the 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 a metapopulation. The analogy to genetic drift in a human population can be inserting through social mechanisms in models.

In a simple model of the reproductive population, natural selection has an impact on its development with the presence of eugenics (Pękalski & Cebrat, 1999). Two populations are in a changing environment - the first, with the genetic pools without interventions, and the second, with two types of interventions. An intervention can be made by a eugenics program, which removes the ill-suited, and by in vitro fertilisation via the introduction into the system of previously eliminated individuals. In the model of eugenics, despite the increase in adaptation to the environment and a better functioning society, a disaster can happen in case of changes in the conditions of the environment (much less frequent than in the system without intervention). The introduction of the previously eliminated individuals causes a permanent weakening fit.

In another example, let me illustrate the model of the evolution of a certain desired characteristic, namely fertility. Historically fertility remained more or less at the same level (subject to a much higher mortality rate fluctuations) until the Industrial Revolution, when the trend reversed. Today, the Western civilisation is on the verge of extinction (fertility decreased below the level of society reproduction in most of countries). Extrapolating the fertility rate decrease, the Western society will die. In my papers on the demographic structure (I.1.HPV-Pl) and marriage dynamics (Jarynowski, Nyczka, 2016), I confirm that concern. By using the current rate of decline of the fertility rate, we can determine the endemic population decay time (the time, in which native inhabitants of the country will be considered as the minority) and e.g. for France it is already less than 50 years. Even if such discussion could be politically incorrect, I have already modelled such an extinction of native species, based on the observations from Azeri Islands (Jarynowski, Lopez-Nunez, Fan, 2014). However, many scientists, using evolutionary modelling, say it does not have to be like that. According to the Simmel theory, certain features of individuals are activated or not depending on external conditions. Due to the possible vertical and horizontal activation, fertility trend can change in the next generations. The differentiation of attitudes (hidden - biological, desire to have children) could affect the average value of that trait observed in the population and the trend could be reversed (Kolk, Cownden, & Enquist, 2014).

In my thesis research, I focus on the evolution of social norms about sexuality and divorces. I analysed the historical divorce rates to find a best fit among the available models (Jarynowski, Kliś, 2012). In next step, I have tried to parametrise the intensity of social norms (Jarynowski, Nyczka, 2016), which should correspond in principle to the phase diagram of water [Fig. 6 left].

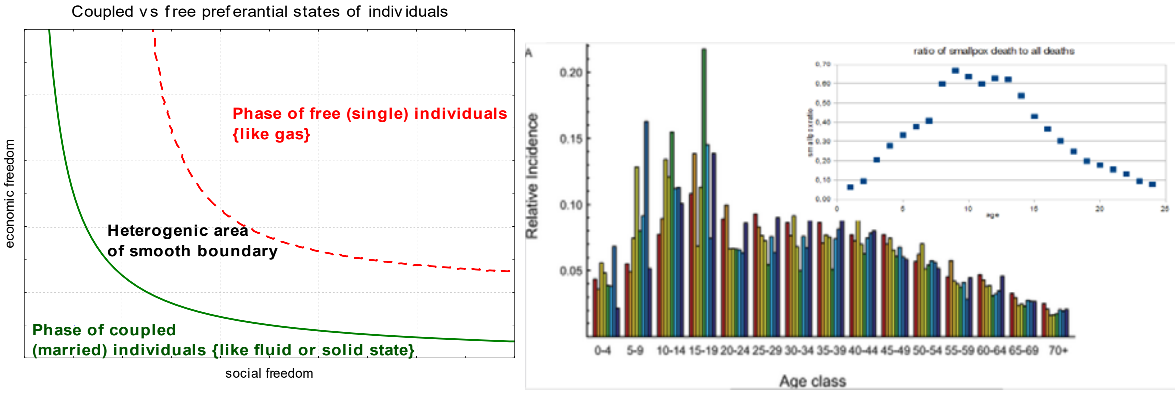


Fig. 6 (left) Idealistic phase diagram of marriage/divorce phenomena. (Right) Relative incidence of a new emerging infection in a completely susceptible population, when the infection is spread between and within age groups by the contacts as observed in correlated way for current polish society (Mossong, … , Rosinska, et al., 2008) and for Smallpox in XVIIIc. Wrocław (Jarynowski, 2014).

Natural Language Processing. NLP methods are quantifying text by reveaing stylistic, grammatical or lexical factors, eg. through meaningful vocabulary. In the initial analysis, researchers look at frequencies list of words separetly for various part of speech. Stylometry methods enable distinguishing individual authors on the basis of texts produced by them. A topic model by classifying items, could illustrate differences between text. In paper II.7.Fiction I look for the differences in understanding of different notions and narrative strategies on the level of characters and also lexical meanings (Jarynowski, Buda, Piasecki, 2016). The fields of NLP (Natural Language Processing) and Data Mining have already developed set of tools in Text Processing. Researchers in science and commerce have used them for a long time. If we want to built similarity measure fo set of texts, we can assumed that such differences can be observed on the level of semantic associations between words. The meaning of entities, and the meaning of grammatical relations among them, is related to Measure of Semantic Relatedness (henceforth MSRs), and can be specified as follows: Wo xWo -> R , where Wo is a set of words and R is a set of real numbers. For a pair of words, its MSR value is calculated on the basis of number of occurrences of both words across different contexts in a text corpus The contexts are typically characterised in terms of the co-occurring words and their syntactic relations. Good MSR should assign for words associated by some lexico-semantic relation higher values than for those that are not related. As the MSR values cannot be directly compared, pairs of MSR should be compared according to:

- how many words are shared among k-best list for the same word,

- how much are correlated k-best list for the same word considered as the rank lists.

Fig. 7 Possible use of MRS list using Polish instance of WorldNet. Source: http://plwordnet.pwr.wroc.pl/wordnet/

Comparison of different Measures of Semantic Relatedness gives us level of similarity between text/corporas. Network of concepts (words) which are connected according to their meaning by semantic relations in Polish WordNet [Fig. 7] have been studied form complex network perspective too (Bujok, 2014).

Structural equations modeling (SEM). The idea behind SEM is the possibility to find out causal relationships by systematic analysis (Konarski, 2014). Let consider a survey of hospital personel, where set of questions (manifest variables) correspond to statements related to theory of organization (latent variables) and include (Jarynowski, (…), Rosiński, 2017):

- causal modeling, or path analysis, which validate relationships among variables and tests the causal with manifest variables and latent variables [Fig. 8];

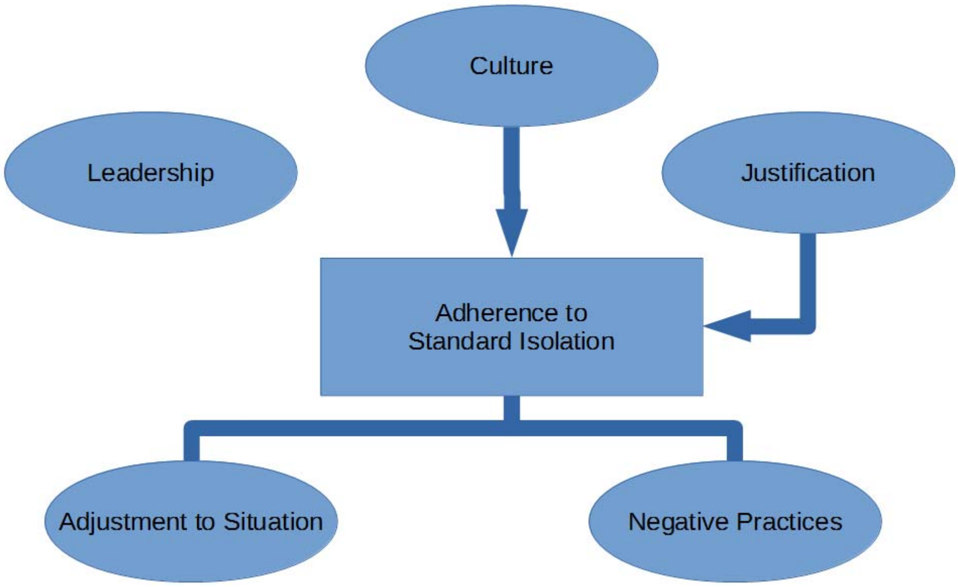


Fig. 8 Causual Model of Adherence to Standard Isolation (the most commom method to avoid Hospital Infections) from theory

- confirmatory factor analysis, an extension of factor analysis in which specific hypotheses about the structure of the factor loadings and intercorrelations are tested aginst data [Fig. 9].

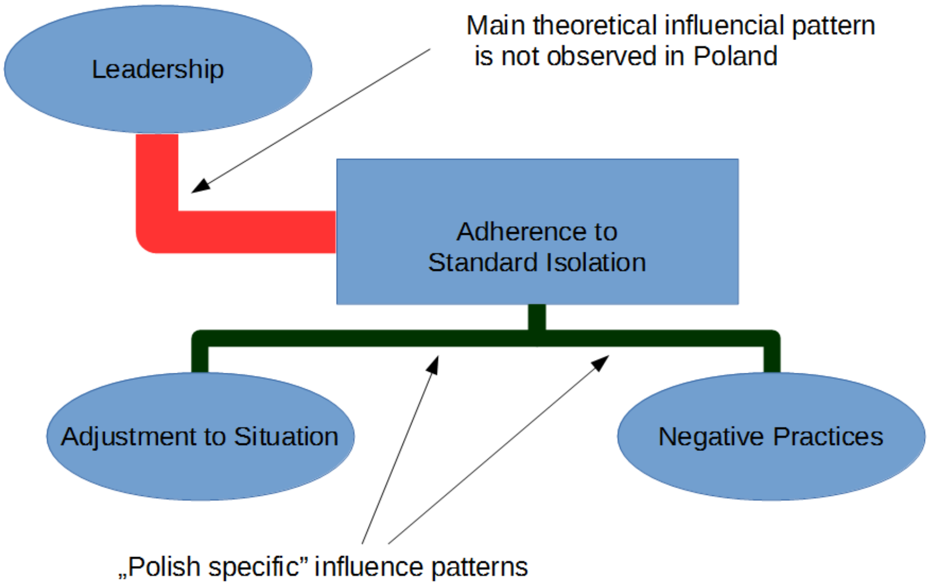


Fig. 9 Confimatory Model of Adherence to Standard Isolation (the most commom method to avoid Hospital Infections) after post-empirical model adjustment

Introduction to mathematical modelling

There is a huge variety of possible approaches to modelling (Pabjan, 2004). If we are interested in correspondence to reality, models can be precise, but very complicated. If we agree to lower the precision, models can be simple and easy to analyse. A model can be deterministic or stochastic; interactions can be implied by forces, energy, rules; variables can be discrete or continuous. Models may be further divided into two major, substantially different types, macroscopic and microscopic. In the case of macroscopic models, we want to answer the questions "how and how much". We do not care, what happens at the micro level of individual units of the analysis, only how the respective average values behave. Here, we are mainly dealing with all kinds of structural equations. This description is similar to the macroscopic description of complex systems, such as is the case in thermodynamics, which includes temperature, pressure, volume, etc. With this approach, you can answer a lot of quantitative questions; you can also generate more or less accurate predictions.

Macroscopic models and Malthus law case. The example is the Malthus model of population growth. Until mid-20th century, population growth on Earth was observed to be exponential. In the Malthus model, the world population keeps increasing (N) exponentially (curve "J"), while in the Verhulst model, it slows down, which is called a logistic curve ("S") [Fig. 10]. This also points to the very important issue of the reliability of models. The economists gathered around the Club of Rome predicted that human population would have exceeded 10 billion by now (2015), while it is totally not the case. Not going into details, in which there are problems and traps in modelling, I must confess that my H1N1 model presented in one of my master thesis was also far from reality (Jarynowski, 2010b). I tried to forecast the epidemic of swine flu outbreak size before winter 2009/2010. My prediction was perfect in terms of the peaking time and the duration of the outbreak, but I overestimated the total size of the outbreak more than two times.

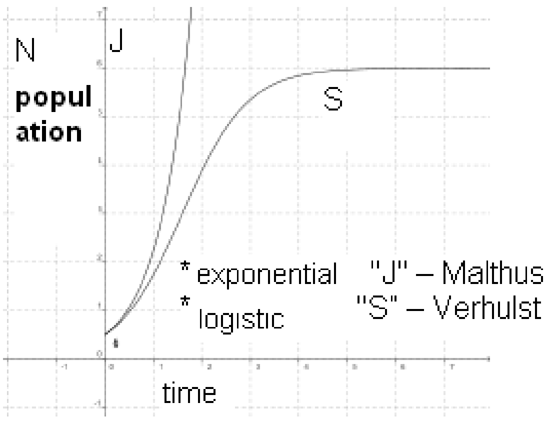


Fig. 10 The firsts models of population dynamics

Microscopic models. The main disadvantage of such models (macroscopic) is the lack of answers to the question about the causes for the occurrence of the phenomena ("why?"). Microscopic models were created in order to try to answer the question "why?". In the case of social and economic sciences, models are divided are divided as follows:

- Microsimulation, where the objects change their state due to deterministic or stochastic rules (Jarynowski 2010b);

- 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. You could say that building a named agent is a generalisation of the concept of particles, many body systems, etc. known from physics.

Modeling goals. The application of adequate theoretical methods and empirical approaches of rigorous sciences like mathematics and physics to economic and social issues has many faces. From the historical perspective, any quantitative science starts by collecting and systematizing empirical observations, then the search for regularities and patterns takes place, which finally results in theoretical formulation, which captures the observed behaviors and mechanisms. Although the current scientific community tries to make progress on all three stages, there are still methodological and conceptual issues that we think should be addressed in that context. Here, I present most important open or re-open methodological issues (Berman, Jarynowski, et. al., 2016):

- How to use statistical mechanics for approaching economic and social issues? This should be addressed at all levels of social systems, which sometimes lack micro-, meso- and macro-foundations.
- How conclusions from mathematical and physical models can be translated into practical policies that would affect the 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 a key to good science and should be trustworthy and broadly shared
- Universality versus specificity. The tendency in the hard sciences is generalization of phenomena and properties, while there is a tendency in social sciences to particularity and specificity. These natures should synergize in order to make progress and bring up new ideas on one hand and make them applicable on the other hand.

Software used in modeling and data analysis

Computational approach originated from the applications of engineering control systems needs software to be used. Inside this thesis:
- few langueages as R, C++ and Visual Basic;
- modeling enviroments as Vensim, Netlogo;
- Network Analysis Toolboxes as Pajek (De Nooy, Mrvar, Batagelj, 2011) and Igraph.

VENSIM
A short tutorial Vensim is presented here (Popov, Popova, Brouwers, 2010).

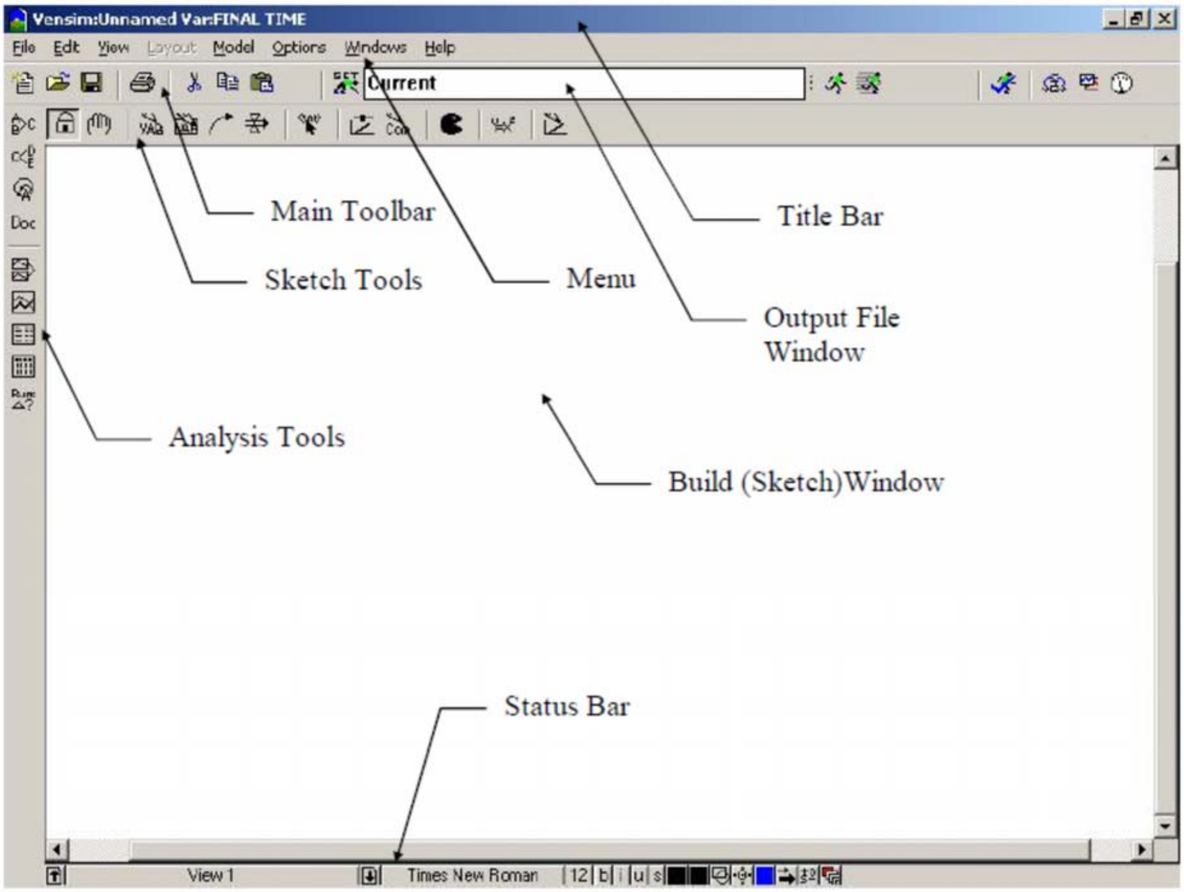


Fig. 11 User Iterface of Vensin. Source: (Vensim 5 Modeling Guide, 2003)

In Vensim you can create your models by drawing items on the Sketch Window. The most important item (blocks) are:

- Variable- Auxiliary/constant (creates variables or constants). You can click on it to create a constant or a variable (auxiliary in Vensim terminology)
- Stock variable (creates stocks or levels). You can click on it to to create stocks and work in the same way as for variables.
- Arrow tool creates arrows which will correspond to dependences
- Flow variable creates flows. For an inflow should click on the working area and drag to the stock variable. If the source for the flow is unknown, you start to draw the inflow by clicking in the working area and draging to the stock. The same procedure is for the outflow with the only difference that you start at the stock and end in the working area.

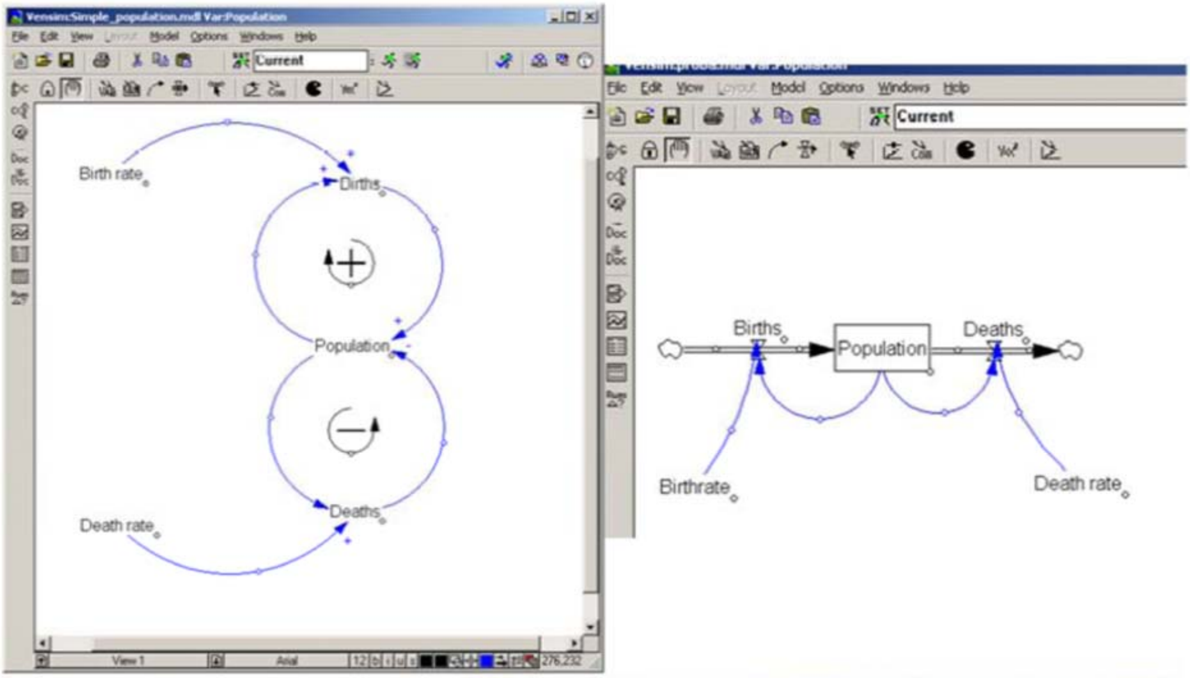


Fig. 12 Birth/Death process presented as Causal Diagram [left] and Flow model [right]

NetLogo

The aim is to integrate a wide range of networks in which physical space is a crucial factor. I have used NetLogo (Wilensky, 1999) for desctibing both spatial agents locations and movments with rule based characteristics of the model. Software cover the integration of empirical data and the automatic execution of multiple simulation runs, as well as the integration of ABM with spatial data (GIS).

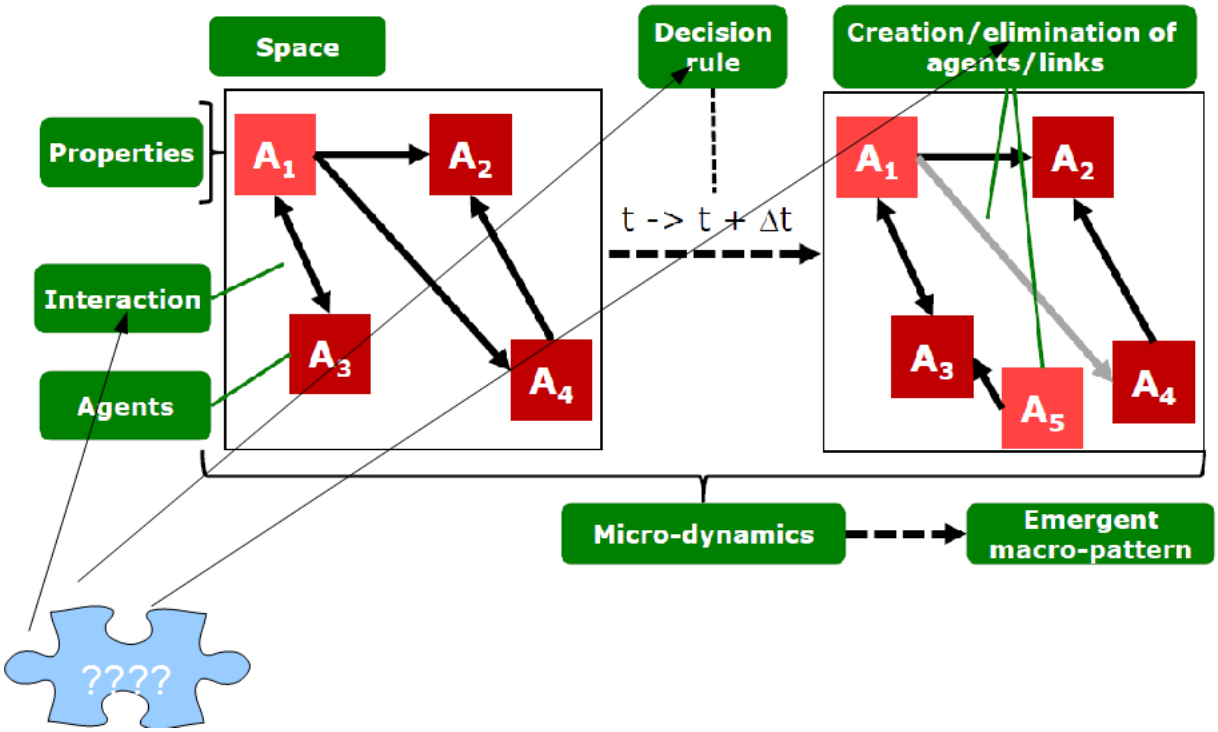


Fig. 13 General Agent-based methodology and the most difficult checkpoints wihin social phenomenon

In NetLogo GUI there is the three main tabs:

- Interface: where the model is run;
- Information: where the model is documented;
- Procedures: where the model is developed.

There are Four main types of user interface elements:

- The main model space (grey). It has origin and min/max x/y coordinates;
- Buttons to start/stop the model and to execute other procedures;
- Sliders/switches/choosers/inputs to enter parameters;
- Plots/output fields to retrieve information from the model.

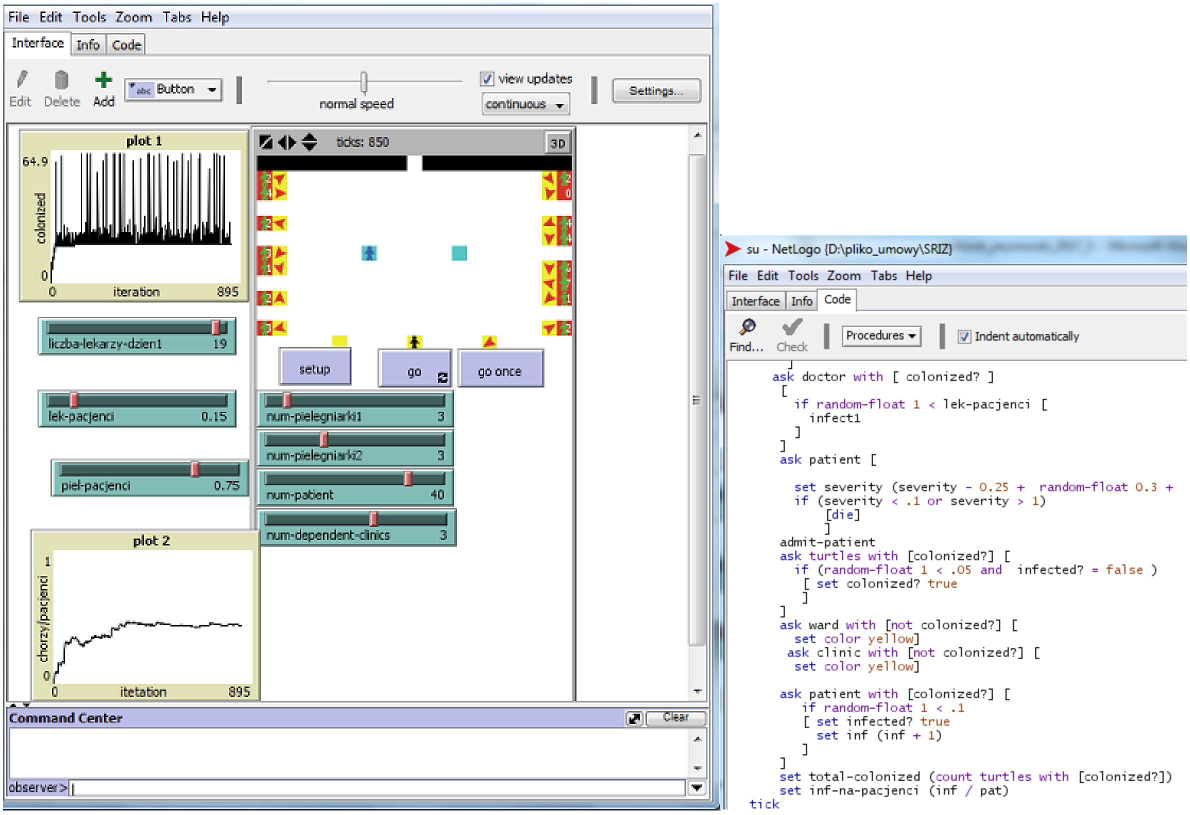


Fig. 14 GUI of Netlogo running a program and fragment of code from SIRS-Z project

NetLogo has a set of model agents with certain properties, which can run custom-written or built-in procedures ("primitives"). These agents are:

- Observer. An observer is an outside user of the model (user). The observer can run certain procedures directly, or "ask" other entities to run them.
- Turtles. A turtle is NetLogo's representation of an agent. It has certain built-in properties, but more can be added.
- Patches. A patch is the elementary spatial unit in the NetLogo grid. It has x/y coordinates, but also color etc. Other properties can be added.
- Links. A link is a connection between two agents.

In general, Netlogo is currently a tool widely used to simulate complex systems in social sciences. On the other hand, a there are plenty of languages or enveriments, to run stochastic processes whose also allow to manage spatial patters.

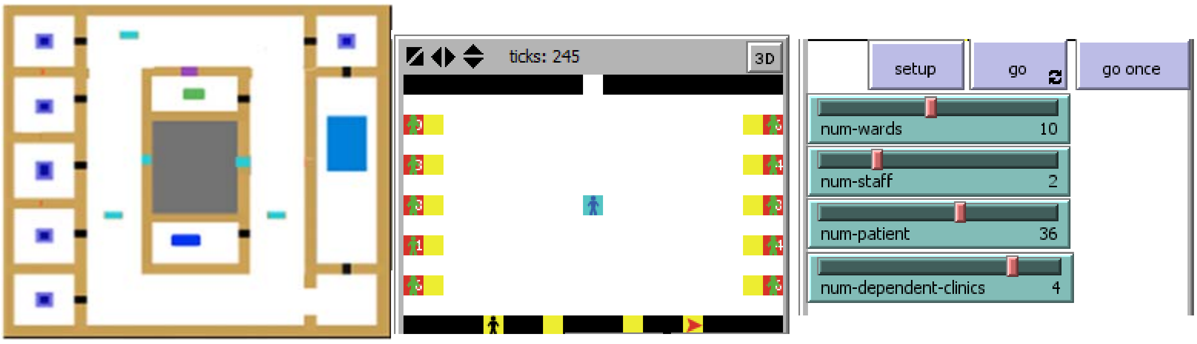


Fig. 15 Implementation of real buildings description and organizational habbits of personel in hospital infection project (SIRS-Z)

In currently run hospital infection project (SIRS-Z), I collect time-varying structure of contacts reconstructed from Polish hospitals. By interviewing staff, I obtain organizational habits, technical knowledge of infection control team have increased. First of all I plot CAD view of wards into NetLogo. Next, I ask question about transport paths, shift distributions, geo-localizing isolated/cohorted patients or stuff, etc. After filling in a brief online survey [Fig. 16], users will maintain numerical variable into our NetLogo application [Fig. 15]. In Netlogo user can study possible scenarios of organizational change as dedicating staff to patients, transport paths rerouting, shift distributions, geo-localizing isolated/cohorted patients, etc., with rererence to baseline introduces by survey. As far as I am concerned, some effects have to be entered manually and the model produce correct results in the order of magnitude.

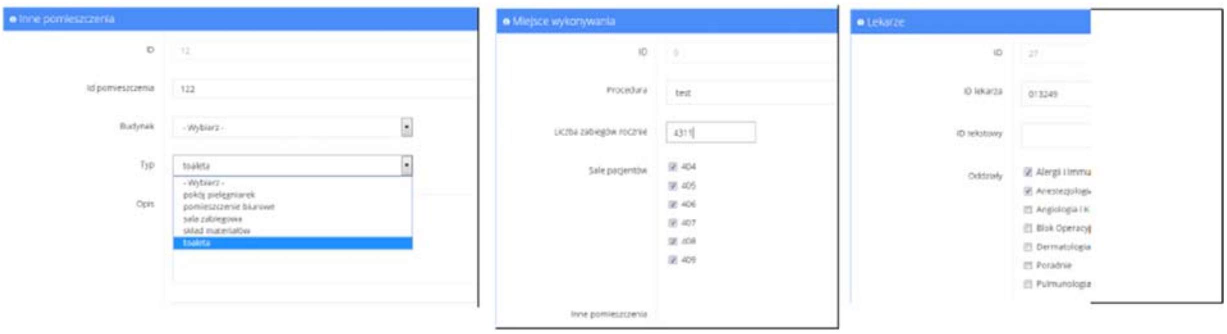


Fig. 16 Choosen questions for implementation of Netlogo program in hospital infection project (SIRS-Z)

I Epidemiological modeling

"All models are wrong but some are useful"

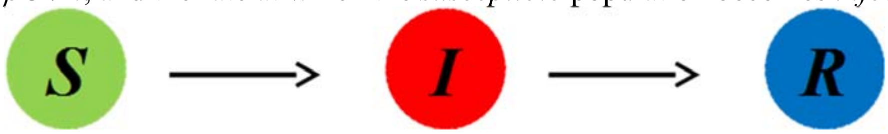
George E. P. Box

The area of epidemiological modelling is explored by researchers of different academic backgrounds: physicians, physicists, mathematicians, statisticians, computer scientists and sociologists. The mathematical modelling of the epidemiology of infectious diseases is an interdisciplinary science that supports public health institutions (Camitz, 2010). All of them have something to add to this issue with different approaches, beginning from mathematics, through computer science (simulations) and concepts from physics, and ending with sophisticated sociological and statistical analyses. The aim of my work was initially to collect a few perspectives (mostly mathematical) and to develop them, but in effect the final analysis is more empirical, which makes it more practical in use. This work is also very extensive, because of its interdisciplinary nature, and highlights some methods that were not used in epidemiology until this moment. The epidemiological models that treat transmission as "human-to-human" from the differential equation point of view do exist in earlier literature, but in more recent agent-based models they appear more often. Mathematical models and computer simulation start to play a significant role as the quantity of social interactions is enormous, but real data, especially register-based, is more important than simulations (Brouwers, 2006). My agent-based models on STI – Sexual Transmitted Infections (I.1.HPV-Pl), (Jarynowski, 2015b) and HAI – Hospital Acquired Infections (I.2.MRSA) were first calibrated on Swedish data, because that country has a huge dataset of citizen activities and gives an opportunity to analyse them. Unfortunately, at this point computer models are missing and there is no chance to help "field" epidemiologists. At present, outside the CIOP-National Research Institute (Grabowski et al., 2012), with little support of the National Institute of Hygiene, nobody does research on the modelling of the spreading of infectious diseases. It is sad that the fact that some Poles were involved in the beginning of the field is known only to small number of people and most of them are not Poles. The first mathematical model in history describes the epidemiology of smallpox in Wroclaw (Bernoulli, 1766). The author of that model, a famous mathematician named Daniel Bernoulli, presented his results in 1766 in Paris and concluded as follows, which is motto of my thesis: "I simply wish that, in a matter which so closely concerns the well-being of mankind, no decisions be made without all the knowledge that a little analysis and calculation can provide". In Poland, there is no scientific unit specialising in epidemiological modelling, and as a result there is no impact on the decisions makers in health care (1-2 scientific papers per year are published in Poland, while in the United Kingdom, a country with a similar number of citizens, there are hundreds of them). I have recently published a review paper in Polish about the most important empirical studies in Poland (Jarynowski, Grabowski, 2015).

SIR model formulation within differential equations

The mathematical description of phenomena needs more assumptions to be made in terms of understanding the real situation. In the first instance, let us consider the spreading of a non-fatal disease, to which no person is naturally immune. Let us suppose that the population can be divided into two groups: the Susceptible-Healthy and the Infectious-Infected (Murray, 2002).
Assume that at general time t:
$S(t)$ = Number of Susceptible
$I(t)$ = Number of Invectives

with $S(t)+I(t) = N$
The problem is to model spread of the disease.
Consider a single *susceptible* individual in a homogeneously mixing population. This individual contacts other members of the population at the rate C (with units' time^{-1}) and a proportion I/N of these contacts are with individuals who are infectious. If the probability of transmission of infection given contact is β, then the rate at which the infection is transmitted to *susceptible* is $\beta CI/N$, and the rate at which the *susceptible* population becomes *infected* is $\beta CSI/N$.



The *contact rate* is often a function of population density, reflecting the fact that the contacts take time and saturation occurs. One can envisage situations, in which C could be approximately proportional to N (which corresponds to mass action), and other situations, in which C may be approximately constant. Hence, terms like βSI and $\beta SI/N$ are frequently seen in the literature. For these, and for many instances, in which the population density is constant, the contact rate function C is subsumed into β, which is now no longer a probability, but a "transmission coefficient" with units' time^{-1}. To reduce the value of the coefficient, let us write: $r=\beta C/N$. Thus, the development of the model allows for a possibility of recovery (an individual becomes immune to the disease). Looking at the recovery term, we assume that it is proportionally related to the infectious. In the end, let us suppose that population is divided into three classes: the susceptible (S), who can catch the disease; the infectious (I), who can transmit disease and have it; and the removed (R), who had the disease and have recovered (with immunity) or are isolated from society. The transition pattern can be represented as follows:

$$\frac{dI}{dt}=rSI-aI,$$
$$\frac{dS}{dt}=-rSI,$$
$$\frac{dR}{dt}=aI$$

A key question for the given $r, a, S_0, I_0,$ is whether the infection will spread or not and if so, how it will develop in time and when it will start to decline. Since the initial condition for S- $S_0<a/r$ then $dI/dt<0$ in which case $I_0>I(t)$ and I goes to 0 with t going to infinity. On the other hand if $S_0>a/r$ then $I(t)$ increases and an epidemic appear. We have something like the threshold phenomenon S_c which depends on initial numbers. Concluding let me write:

$$R_0=\frac{rS_0}{a}$$

where R_0 is basic reproduction rate of the infection. This rate is crucial for dealing with an epidemic which can be under control with vaccination for example. Action is needed if $R_0>1$, because then an epidemic clearly breaks out.

Reduction of SI model. In first instance let consider the spread of a non-fatal disease, to which no-one is naturally immune. Suppose the population can be divided into two groups: Susceptible-Healthy and Infectious-Infected. Such a model referred to a phase plane. The concern is only with values of I in the interval [0,N], indicating that I'(t)>0 which means that number of infectives is increasing. By inspection of figure it is seen that the arrow approaches the equilibrium state I=N for all permissible values of I. This means that the number of invectives will tend to the equilibrium state N, no matter how many invectives are initially present. Thus I(t)=0 is referred to as an unstable equilibrium state.

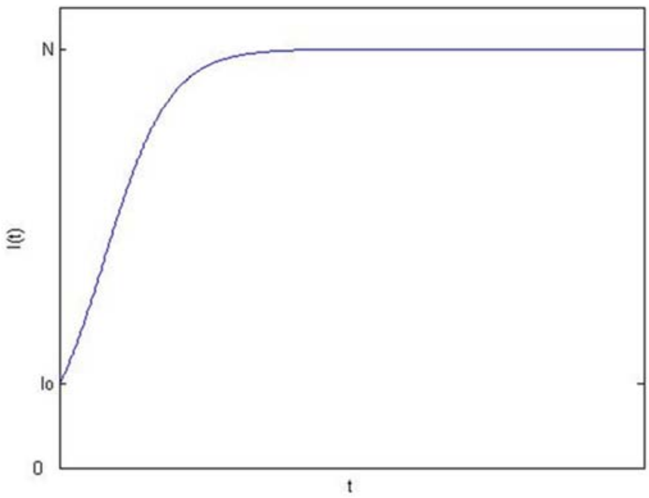
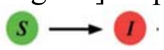


Fig. 17 Spread of Infection for SI model – where for a least on infected, whole population will be infected.

In my model for Stop-ACTA (Jarynowski, 2015c), I considered the time evolution analyses based on Stop-ACTA data. I plotted [
Fig. *18*] empirical data (1) end solution of Susceptible – Infected (SI) model.



I solves 3 possible implementation of SI model with diffent assumption on total population size:

$I_{total}(t) = I(t) = \frac{NI}{I_0+(N-I_0)\exp(-rt)}$ (2) constant population size

$I_{total}(t) = \frac{N_{inf} I_0}{I_0+(N_{inf}-I_0)\exp(-r\prime N(t)t)}$, (3) correction on changing population size

$\frac{dI}{dt} = \frac{r}{N_0-at}(N_0-at-I)I - a\frac{I}{N_0-at}$ (4) system with dynamic population size

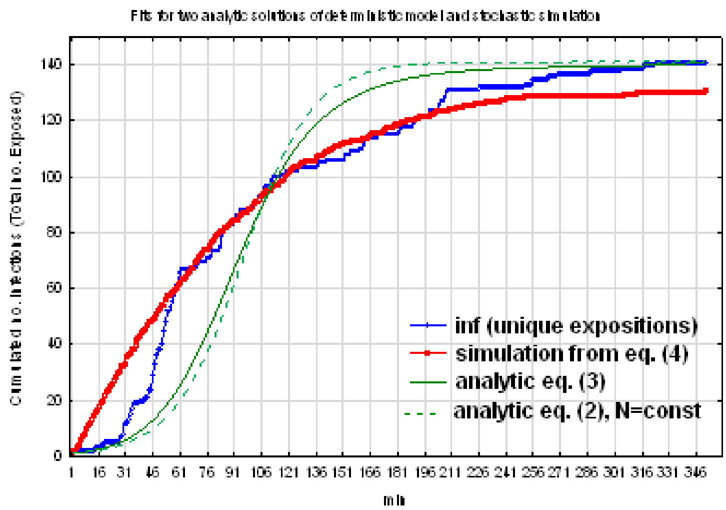


Fig. 18 Dynamics of Stop-ACTA campaign during fist day

Extension of SIR model. The model can be expanded by an additional sub-population (e.g. Exposed to the pathogen - E -), or various kinds of pulses/delay functions to be more realistic with the actual processes (Grabowski et al., 2004). Although those non-linear extensions make the equation closer to the real problems, an analytical solution often cannot be developed (usually only an asymptotic or a solution with certain constraints is possible). Moreover, even the simplest SIR model presented above has no unique analytical solution for the whole parameter and initial condition space (Jarynowski, Grabowski, 2015). Therefore, numerical methods are used - usually difference equations (a special case of the numerical procedure for obtaining the trajectory of the time).

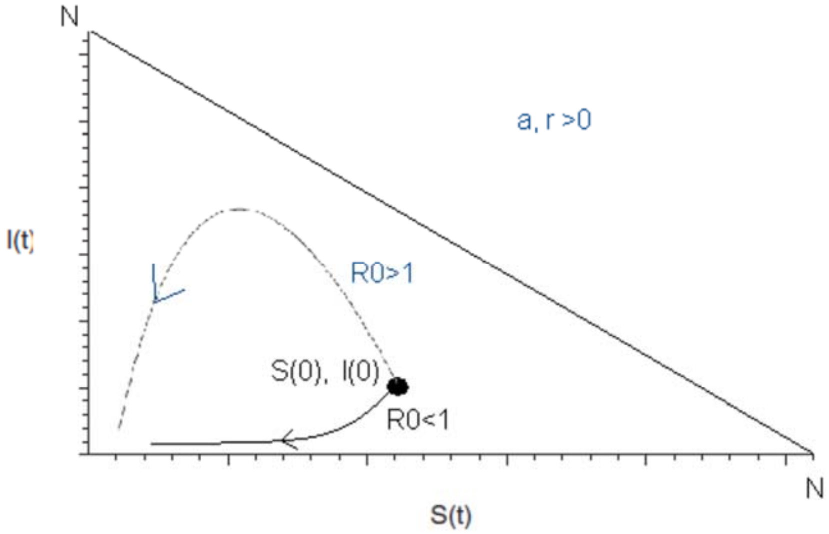


Fig. 19 The description of the stability of equilibrium points in simple SIR model, where no general analytical solution can be obtained

Other examples are partial differential equations in the geographic spread of diseases. The spatial component, e.g. the population density, can be applied in SIR models not only to describe the time evolution of the epidemic, but also of the transfer of the disease to new areas. For example, by slightly modifying the infecting equation to express the diffusion (where Δ is the Laplacian, and D is the diffusion coefficient), we obtain:

$$\frac{\partial I}{\partial t} = D\Delta I + rSI - aI$$

The dynamics of the spreading of the disease in a geographical space (diffusion) is used to map the epidemic of plague that started from the harbours of the Mediterranean Sea at the beginning of the 14th Century. It arrived to Poland five years later, went farther north and expired.

Stochastic model

A competitive approach to the deterministic modelling of an epidemic is the probabilistic method (Jacob, 2010). Instead of model parameters, it uses rates, because we are dealing with probabilities. Therefore, for example, the rate of infection (movement from *S* to *I*) is represented by a pure probability of changing state, not as a flow between stocks, in the deterministic approach. These two approaches were implemented parallel in the Stop-ACTA project, in which in one paper a stochastic notation was used (Jarynowski, Jankowski, Zbieg, 2015) and in another a deterministic notation (Jarynowski, 2015c) was used. In a stochastic model, you can use the concept of a branching process.
The main difference between the stochastic and the deterministic methodology is the meaning of R_0. In stochastic mode for $R_0<1$ it is already possible to start epidemic. Let us consider the Markovan discrete process with the notation ($I(t) = i$, $I(t + \Delta t) = j$):

$$p_{\Delta t}(i \to j) = \begin{pmatrix} ri(-i)\Delta t, & j = i+1 \\ ai\Delta t, & j = i-1 \\ 1-[ri(N-i)+ai]\Delta t, & j = i \\ 0, & \textit{other cases} \end{pmatrix}$$

where $p(i\text{->}j)$ is the probability of state changing (size of cohort I).

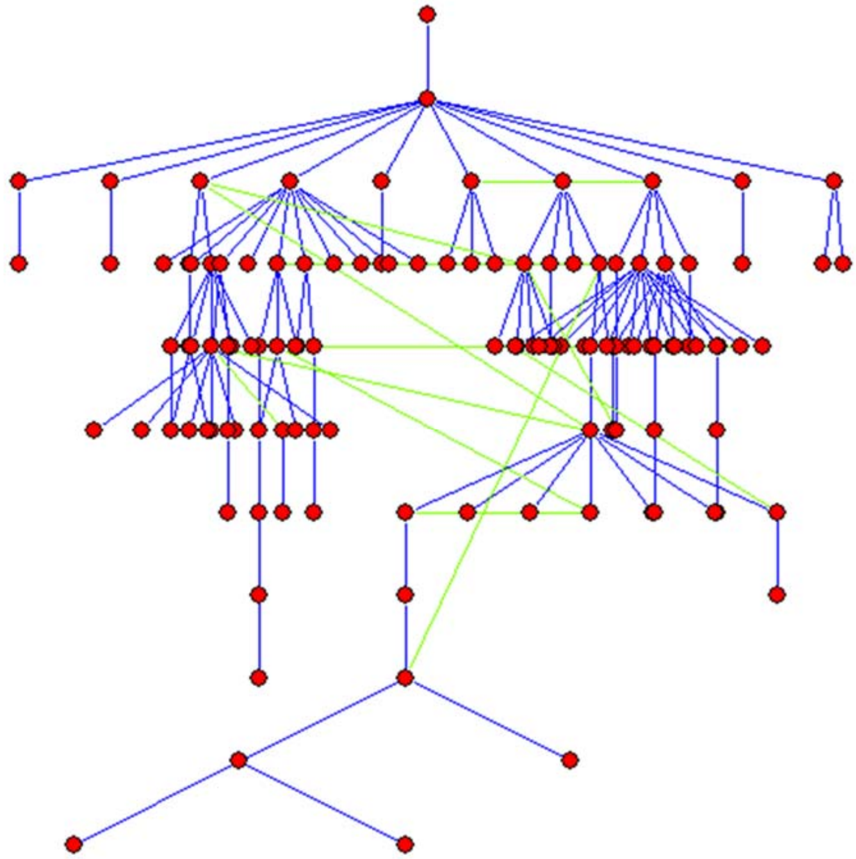


Fig. 20 The branching process of a real epidemic with 11 waves (generations). Blue is the primary and green the secondary (re) infections

In stochastic notation instead of rates, probabilities of changing state are more common. So β~r/N is probability of infection per time unit of of contact, and probability of transition between states I and R (recovery rate) equals γ~a. Epidemic reproduction rate can be written now in a new form: $R_0=\beta/\gamma$.

Cellular automata. In addition to the branching processes, cellular automata (grid cells with conditions and rules) are used in stochastic epidemiological modelling. The main objective of this approach is the statistical analysis of SIR models. Physicists also like those structures and many studies have been made of its topological characteristics, e.g. percolation. The epidemic is spreading through infected cells to neighbours in accordance with the established probabilistic rules. In addition to the standard models in the grid [Fig. 21] (most applications can be found in disease outbreaks in plants, where the system of flower beds literally has such a structure), there are all kinds of shortcuts corresponding to the vectors of infection (Dybiec, et al. 2009).

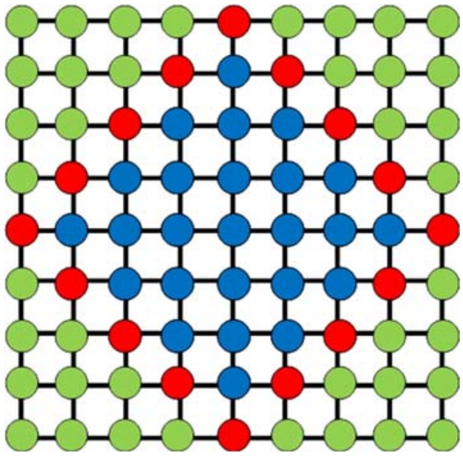


Fig. 21 Process of deterministic epidemic spread SIR started from single infective in central point of a grid. There are waves of infected - , after few steps of simulation change to removed- , and rest of population not yet affected susceptible - . It is CA with von Neumann neighborhood

Other propagation processes

The processes of information propagation as data flow are one of the core problems of complex systems and data science. Such pseudo-epidemic models can be of types like: diffusion of information (where mass is conserved), opinion formation, spread of influence or rumor spread. In physical systems, there are two different spreading mechanisms:

- With conserve observable, for example diffusion of gas or mixing of fluids. Here number of particles of different substances are constant (quasi-canonical or N- ensembles). Such systems can be described by various mathematical tools as Brownian Motions (explored more in II.5.Music-StochProc), Fokker-Planck-Smoluchowski equations and so on.
- Without conserve amount of observables, for example waves, percolation or chain reactions. However, conservation law still follows, but there does not depend on our observables directly.

Opinion formation. The evolution of opinions can be analyzed used in paper (Jarynowski, Grabowski, 2017) in Ising model frame – well analysed model of magnetism from various perspectives (Drzewiński, 2000). Here, in the thesis, I introduce single simple realization, while there are many other generalized models of opinion dynamics as voter model, Sznajd model, or the majority rule. Human equivalent of spins - Spinson (Nyczka, 2015) is defined by $S_i = \pm 1$ – an opinion of i-th individual. In a simple grid form 1D [Fig. 22] representation Ising model has analytical solution, but in higher dimensions many interesting from physical perspective phenomena appear.

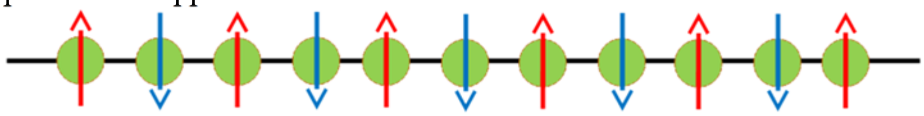


Fig. 22 Chain of 'Spinsons' (one-dimensional Ising model) with different opinions

Let introduce the local field h_i for each individual representing influence of peers and global variables. h_i is a function of interactions with k_i neighbours where weight of link between node i and j is w_{ij} and the external field (stimulation) I.

$$h_i(t) = -S_i(t)\left(\sum_{j}^{k_i} w_{ij} S_j(t) + I\right)$$

I applied Glauber dynamics to this model with system 'social temperature' T.

$$S_i(t+1) = \begin{cases} S_i(t); & \dfrac{\exp(-h_i/T)}{\exp(-h_i/T)+\exp(h_i/T)} = \dfrac{1}{1+\exp(2h_i/T)} \\ -S_i(t); & \dfrac{\exp(h_i/T)}{\exp(-h_i/T)+\exp(h_i/T)} = \dfrac{1}{1+\exp(-2h_i/T)} \end{cases}$$

This is the microscopic picture of opinion dynamics, considered within sociophysics. Such a model at the macroscopic can produce for example phase transition and many other interesting phenomena.

Rumor model. Rumors affect only for a very short time and agent can spread further content only in limited active state, so some differences could be expected with comparison with other diffusion, infection-spreading or percolation-type models.

Following the original terminology, there are three permitted states:
1) ignorants (IG) who have not heard the rumor and hence are susceptible to be informed;
2) spreaders (SP) who spread the rumor;
3) stiflers (ST) who know the rumor but are no longer interested in spreading it.

The dynamic of such a model can be driven by parameter β – aggressiveness of a rumor and γ – its time scale [Fig. 23]. I have investigated real features of empirical contact data with this model - I.3.Polymod.

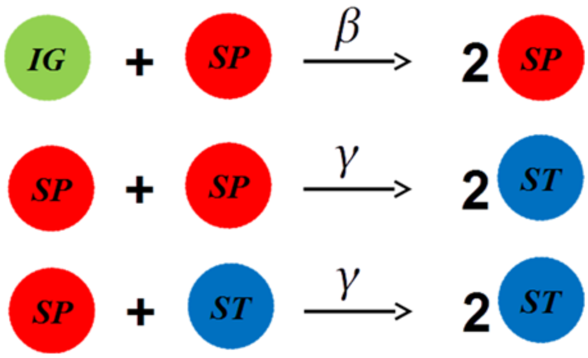


Fig. 23. The dynamics in rumor model

State of the art in epidemiological modeling

Mathematical models can help in describing phenomena of epidemiology spreading and give an answer how to fight with them. Some tools are dedicated to epidemiologist and easy applicable, some are very sofisticalted on theroretical level. Main goals of applied epidemiological modeling are to provide guidelines for controlling disease outbreaks. There are many problems in modeling because of big variety of epidemic types. Mathematical modelling of the epidemiology of infectious is an interdisciplinary science which supports public health institutions. The use of the data and the use of computer simulation, in order to understand and modify the social processes in an innovative project are to personalized health care (knowledge of the location of patient on temporal network of contact could have an impact on decisions concerning the medical treatment). In my thesis, I focus on Sexually Transmittable Infections and Hospital Infections specific characteristics of such kind of spread will be described. Firstly, set of usefull tricks will be presented here from some quick estimation methods through time series analysis up to real-life models.

R_0 post-epidemic estimation.

For SIR –like epidemics following approxitive calculation for empirical reproduction rate calculation have been used wihing epidemiologist. The reproduction rate can be approximated for the SIR model from empirical data:

$$R_0 = \frac{-\ln\left(\frac{S_\infty}{S_0}\right)}{1-\frac{S_\infty}{S_0}}$$

Where S_0 is initial number of susceptible, S_∞ is number of people remaining susceptible after the outbreak.

The formula derivation is easy to obtain (Jarynowski, Grabowski, 2015), but it is extremely powerful in its applications.

Early detection of outbeak - β estimation

The early growth estimation can be easily done by matching incidence trajectory to the exponential function, resulting in the approximation of the force of infection (β). With these calculations the basic reproduction rate of the epidemic (R_0), recovery/detection chance (γ) could be estimated. The early detection method in zeros approximation can be done by fitting the incidence curve to the exponential function, resulting estimation of the infectivity coefficient per time unit (β). According to simplified relation in SIR (Susceptible, Infectious, Removed) model $R_0=\beta/\gamma$, we check what range of possible paramerers fitting the data are critical to satisfy epidemic condition ($R_0=1$).

As a explamle, let us consider African Swine Fever (ASF) Virsus spread in 2017, resulting early stage estimation of the infectivity coefficient per month (β~1.25). According to simplified relation in SID (Susceptible, Infectious, Detected) model $R_0=\beta*det$, we observe that detection and elimination time $det=1/\gamma$ ~24 days gives $R_0=1$. It is enough to control epidemic with elimination time *det* shorter than 3 weeks (where two weeks are the time from infection to the first clinical symptoms).

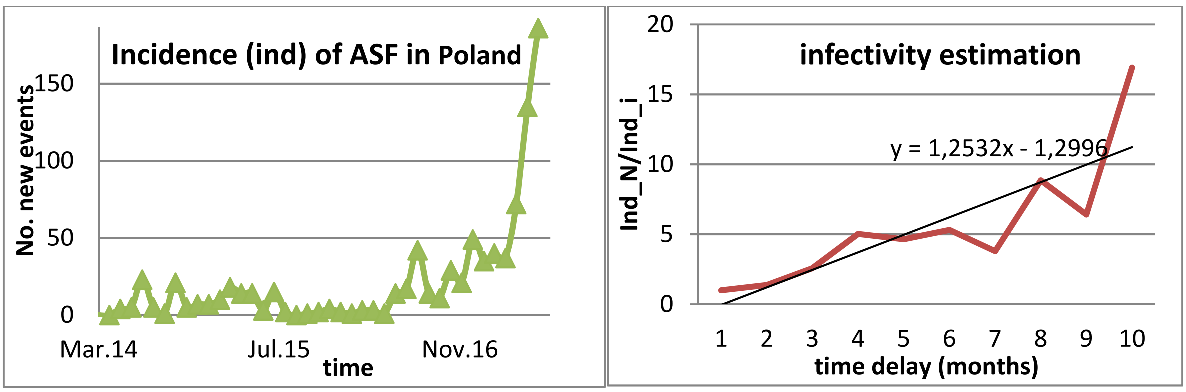


Fig. 24. ASF in Poland. Row incidence rates (Ind) [Left]. Fitting parameter β by incidence increments [Right].

Alert posting via time series analysis
Predictive modelling is the computational process by which a model is created or chosen to try to best predict the probability of an outbeak. In many cases the model is chosen on the basis of detection theory to try to guess the probability of an outcome given a set amount of input data, and for example new flu cases can be estimated based on previous season and current condition. There are many approach to modeling the incidence series with respect to seasonality and one of them has been implemented in R as a package – ‘surveillance’ (Hohle, 2017).

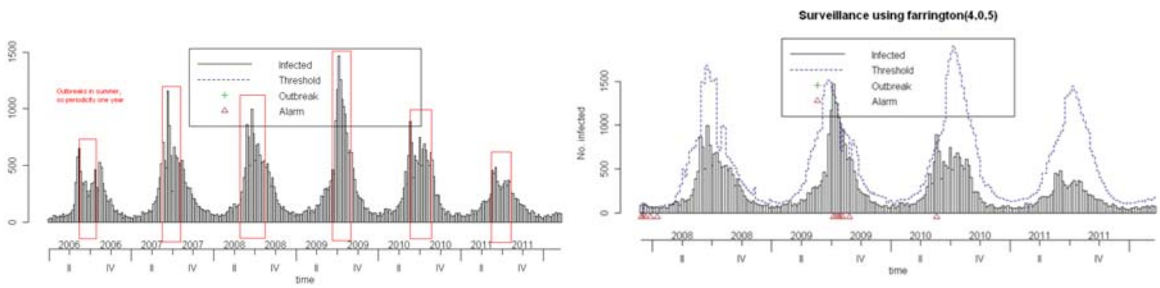


Fig. 25. Vomits queries per month in Sweden. Within ‘surveillance’ package view [Left]. Farrington algorithm for alarms [Right].

Presented package help epidemiologist to understand seasonal patterns and give some automatic alert if anomaly is detected.

Modelling coinfections. The spreading of co-infections (both mutualistic as well as antagonistic) in an empirical temporal network of contacts is an important challenge for modelers. Cooperative infections are very common in sexual infection and comordities are well risk factor for HIV. The same time cross-immunity phenomenon appears in HPV viruses. In bacterial hospital infection on the other hand, thery often rule “there can be only one” appears and longnitudal studies do not show permanent co-infections in too many cases.

Cost-effectivness and cost-benefit analysis
One of the the most important topics of epidemiological modeling is the authoritative analysis of the costs and losses of potential epidemiological control strategies (Kleczkowski, et al., 2011) and identification of potential problems that health care will have to face in the future. Let consider more precise monetary index and measument of population effects of intervention. To do so, epidemiologists introduce quality-adjusted life-year (QALY) - a measure of disease burden, including both the quality and the quantity of life lived (Jędrychowski, 2010). The QALY is based on the number of years of life and adjusted to health state (where 1 is a perfect health and 0 is death) that would be saved by the intervention. Due to infection, QALY is decreased. The same time treatment as well as social costs of disease should be calculated. If there are avaible vaccination, which protect form infection, then compulsory, universal vaccination will be cost-beneficial, because such kind of intervention reduce the losses in QALY for dedicated population. However it could be too expensive respectively to national GDP if cost if intervention will be higher of cost of the disease. To classify, which intervention in cost-effective, the incremental cost per QALY yearly is below GDP per capita of given country or it could be partly cost-effective if it’s below 3*GDP.

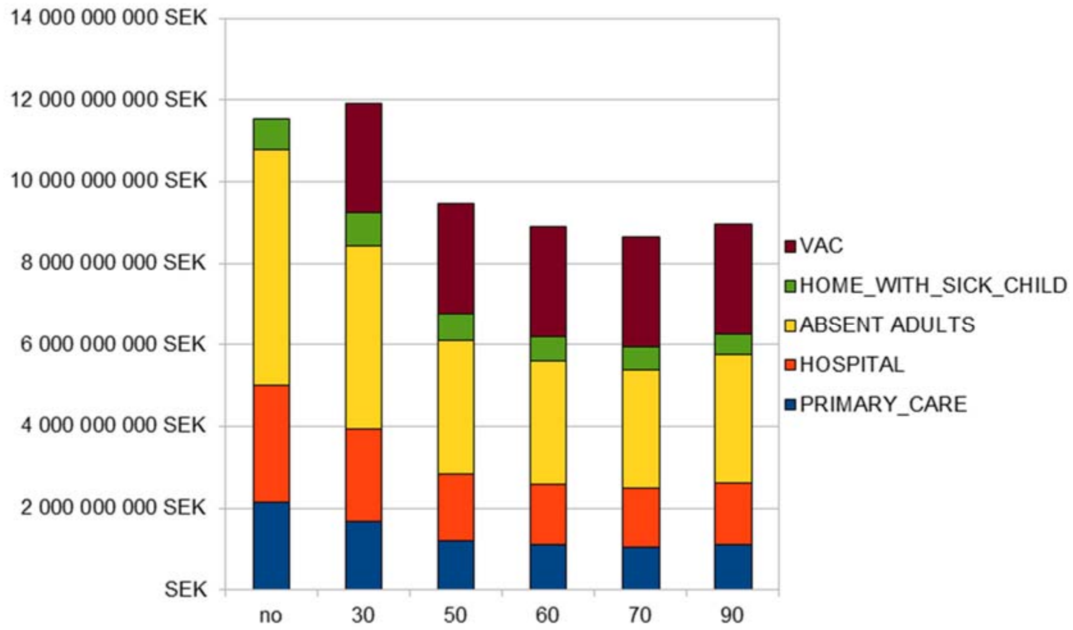


Fig. 26. Total cost in model for Sweden of H1N1 influenza during season 2009/2010 excluding cost of death cases in SEK for various coverage of vaccination (no, 30%, 50%, 60%, 70%, 90%) in medium scenario (Jarynowski, 2010b)

Landscape-based (spatial) propagation and pseudo-gravitational models

Within spatial modelling, researchers create a multi-modal network model of traffic at the intra- and inter-institutional level. Human mobility can be very often simplified by geographical gravity model, where flow of people between regions is proportional to multiplication of regions' population sizes. To illustrate such a very simple model, let consider African Swine Fever (ASF) Virsus spread in 2017 model, where I use forest coverage, pig population in poviats and the distance between centroids of poviats. I propose pseudo-gravitational models of short and long-range interactions referring to the socio-migratory behavior of wild boars and the pork production chain. I run set of simulations for selected subspace of parameters a- swine amount significance, b- disease vectors (wild boards) significance, c- pork production chain significance.

$$p_{ij} \sim \frac{a(P_i * P_j)}{1+d_{ij}} + \frac{b(F_i * F_j)}{1+d_{ij}^2}, \quad g_{ij} \sim p_{ij} * c$$

where
a, b, c – simulation parameters;
i, j – poviats;
P – normalized amount of pigs;
F – coverage of forests;
p_{ij} – probability of infection from a neighbor;
g_{ij} – probability of infection from a whole networks;
d_{ij} – angular distance between centroids of poviats.

the simplest SI (Susceptible, Infectious) model on the level on poviats was implemented, where with probability p_{ij} disease can be transmitted from already affected poviat i to an previously not affected poviat j, which is a neighbor to j. Long distance transmission can be done with probability g_{ij}.

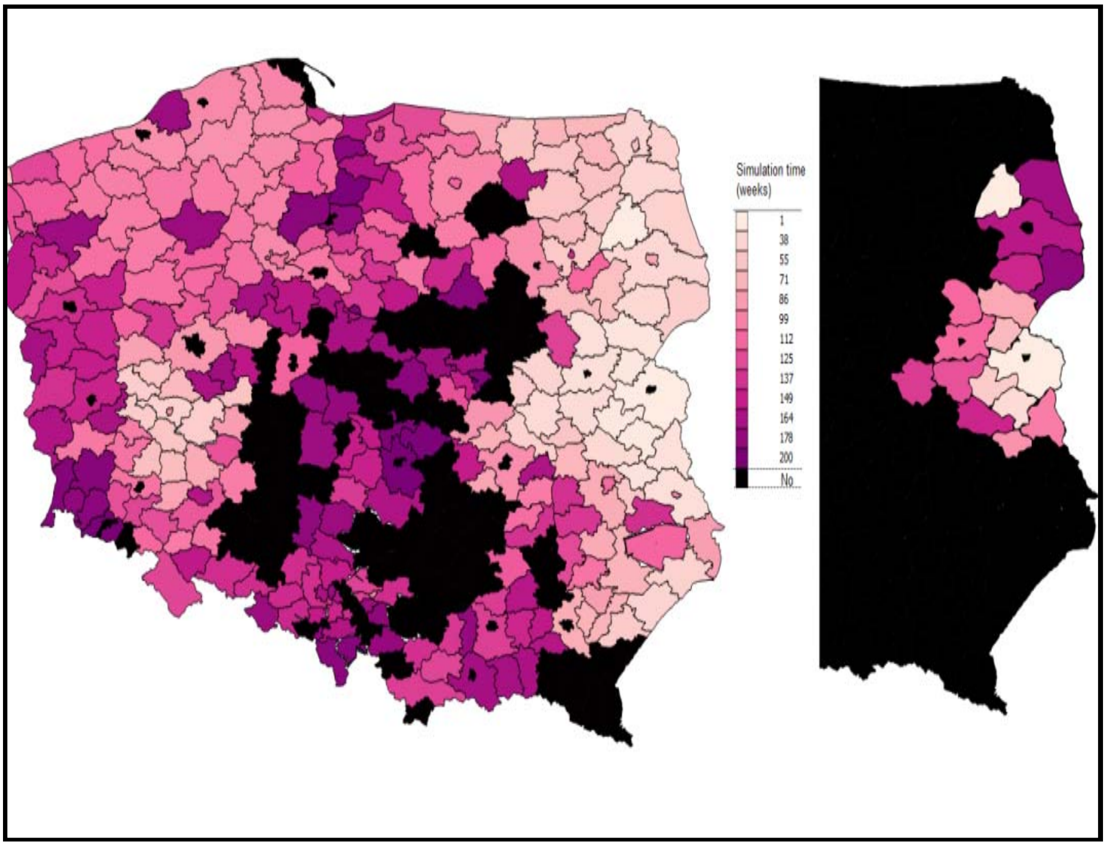


Fig. 27. Examples of simulation spread with seeds in poviats Mońki and Biała Podlaska for next 4 years with possible propagation barriers around invaded infected regions. [Left] Medium swine and vector significance (*a*,*b*), high pork production chain factor (*c*). [Right] Low swine and vector significance (*a*,*b*), no pork production chain factor (*c*).

Sexual transmittable infections (STI)
Modelling in the simplest instance can be done as by dividing population into two sexes and then comining stocks/agents in a bipartile graph. Knowladge of fat-tailed sexual networks, more described in chapter on networks, have also speed up modeling perspective in STI (Liljreos, et. al., 2001).
Computer simulations would answer some questions: when and with what probability should we expect a wave of such epidemic in Poland. Detailed epidemiological significance of pathogens with benefits from model and simulations:

- Chlamydia infection - an infection of the bacteria Chlamydia trachomatis is sometimes called the "silent epidemic" because occurs commonly, but few people know about his illness. The model will verify how likely is to have epidemic in Poland.
- Trichomoniasis - is the most common STI among sexually active young women in Poland cause by Trichomonas vaginalis. Trichomoniasis may cause a woman to deliver a low-birth-weight or premature infant or even its blindness. There was a lot of resistance to antibiotics among certain strains, so I want to check out potential consequences of such a resistance.
- HIV infection (Human Immunodeficiency Virus, HIV). The disease called AIDS leads to death (indirectly). This pathogen focuses the most of scientific attention, however, currently only few cases of endemic infections outside the highly risky group are observed in west European societies. An interesting issue to model is a development of the cross-effects among MSM with other diseases. Co-infection are very important issue for HIV positive.

- Genital herpes - herpes virus (Herpes Simplex Virus, HSV) causes the periodic appearance of bubbles in the genital area. The virus remains in the body of an infected person for life long. Its prevalence in Poland is around 10% mostly in asymptotic mode, so this could be a good indicator of sexuality.
- Gonorrhea - related bacteria: Neisseria gonorrhoeae. Disease is very dangerous to pregnant women. Gonorrhea is treated with antibiotics (potential acquisition of resistance)
- Syphilis-bacterial infection caused by Treponema pallidum, which is easy to overlook and effective treatment is available. Huge fraction of new cases in Poland is transmitted from mother to fetus during pregnancy or at birth, resulting in congenital syphilis. Greater than 90% of world-wide cases are notice in the developing world. The hypothesis is that model will show that the number of new cases would be decreasing from Western to Eastern province of Poland. From the simulation results, we can learn how quickly it will be happening.
-Viral hepatitis (HEPATITIS) B or C-virus infection of HBV and HCV may be transmitted during sexual contact, however other routes are more likely. The target is to quantify the impact of vaccination (HBV) on the number of new cases.
-Moniliasis-also known as candidiasis or thrush. It occurs when yeasts-saccharomycetes "get out of control" and begin to proliferate excessively. Despite the fact, the treatment is easy, simulation results would be complementary to the full spectrum of major pathogens.
- Infection with human papilloma virus (Human Papilloma Virus, HPV). The virus can infect the genital area causing genital warts (mainly types 9 and 11). HPV infection is associated also with the occurrence of cervical cancer (mainly types 16 and 18). This is pathogen, with which I have already started analysis, because its epidemiology has been widely described and modeled in recent years. Cost/benefit analyze for different vaccination strategies (whom? at what age? the girl or also boys?), various screening programs and preventive programs (using condoms) is needed as Poland introduced obligatory vaccination of young woman in 2017.

In this thesis, I model spread of Human papillomavirus (HPV) – a sexually transmittable virus infection, which is not only the main, but also necessary risk factor for developing cervical cancer. The main safety precaution with respect to hpv and cervical cancer (modelling scenarios of many researches) is going to be a combination of vaccinating and screening. Specific to such kind of disease spread states of women are:
D – detected (HPV positive or cancer), have been decteted by screening procedures. In addition, she can trigger a treatment with given sucess probability;
C – cancer stage. The final category, which is usually known from oncological registries;
V – vaccinated. Individuals in this class have been vaccinated. They cannot become infected;
O – removed due to age. This class includes individuals that have or not been through the disease but cannot infect anybody anymore;

Standard elements of the simulation environment:
-possible transition states of individuals;
-structure of the interaction topology;
-costs (preventive+treatment), where total cost is based on cost of treating cancer and cost of socio-political intervention as vaccination and screening.

Hospital acquired infection (HAI)
Health care institutions are associated with many hospital-specific pathogens such as opportunistic bacteria. The key role of patient contact networks in infectious disease transmission requires methods from sociological studies of social networks for preventive protection work. The contact patterns and network structures observed in the general population are also present in hospitalized patients and thus influence the dispersal of hospital-associated pathogens. In literature, there are many attempt to methicillin-resistant Staphylococcus aureus (MRSA) as a

model organismcase study, because it its related to hand hygiene, it is largely restricted to health care institutions (at least in Europe), and finally, MRSA evolves clonally. The MRSA is resistant to more than half of all antibiotics and it is known to the single largest care-related infection problem. For such infectious diseases, in which close contact is needed for the transmission to occur, an individual's position in the contact network is important for the person's risk to get infected. The awareness of the importance of contact networks has brought methods from sociological studies of social networks into the area of preventive infectious disease protection work. The contact patterns and network structures seen in the general population also apply to hospitalizsed patients and thus influence the dispersal of hospital-associated pathogens. In this case, it is assumed that individuals only meet in hospitals, and that sharing the same room is often taken as an approximation of contact as in my case. Close proximity interactions can be also measured by electronic sensors. These data are increasingly used to inform contact networks. However the high level of temporal resolution of the networks built in this way requires novel data analysis and modeling tools.

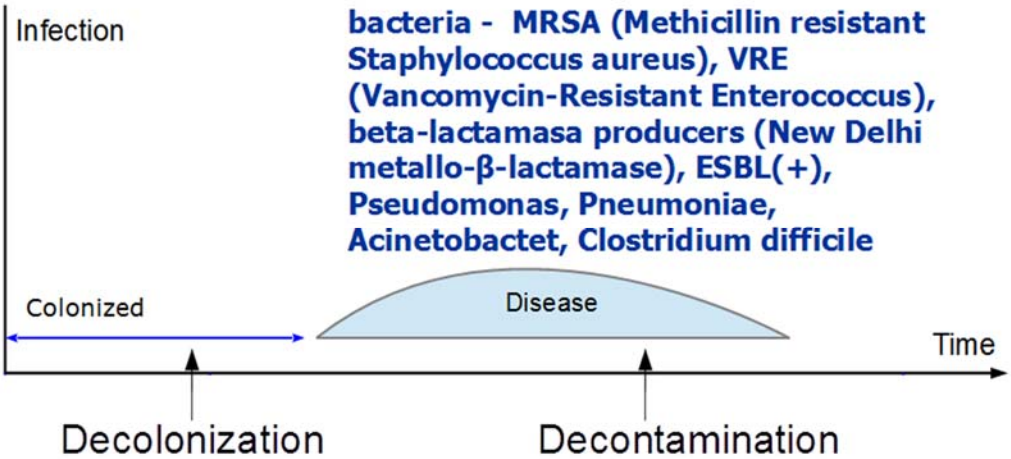


Fig. 28. Generalized schema of HAI development (based on Grundmann, et al., 2006)

Specific to such kind of disease spread states of patient or personel are:
C – colonized or carrier, patient of staff member have been a researvour of a pathogen. No clinical symptops are observed;
Iso – isolated or cohorted. Patient after dectection of possessing pathogen have been cutted off from contact network, to the limit of nessesery treatment for basic medical problems.

Very complicated framework of HAI development and propagation challenges researchers studying hospital infection with modelling by:
- Screening/Isolation cost-effectivness;
- The spreading of cooperative infections;
- Vertical vs horizontal nursing and physician teams organization;
- Analysis of the temporal network from available hospital data;
- Surveys within medical staff through an interactive web application;
- Incorporation of the intra hospital transmission model. E.g. Coupling matrix. Age-dependent. Although the detailed information may not be available, we could imply the colocations of patients based on their demographic and diagnostic characteristics
- The role of care givers;
- Analysis of the temporal network of animal trade;
- Linking networks of relations on the basis of incomplete and very sensitive data using statistical and mathematical techniques;
- Surveys evaluated by the structural model approach;

- Game theory with Hospitals.

A combined meta-population network modelling approach are currently on top, where intra-hospital transmission dynamics are described as a subpopulation within a meta-population of hospitals which are linked by a network. Meta-population modelling was developed in population ecology and depends on within and between patch (geographical regions) transitions. These models have been used in the past to analyse transmission of infectious diseases via airline travel (Belcan, et al. 2009). Meta-population models are perfectly suitable for hospital infection, as they describe exactly the situation we are facing with dynamics of pathogens within and between hospitals. There are well-defined subpopulations of patients and pathogens within hospitals and hospitals are connected via transition of patients from one hospital to another and are therefore linked with each other via a network. If designed using concepts from meta-population modelling, models are modular, so hospitals can also be analysed in isolation, while a lot of studies consider single hospital only. As a first step in developing the model, researchers design a relatively simple within-hospital models based on an SIS (susceptible, infectious, susceptible), as I did in thesis paper I.2.MRSA. This could be a deterministic/deterministi model that does computes the size of an outbreak, but only reconstruct the most likely paths. As an input, it needs the time of entry of a colonized/infected patient, and as an output it delivers the risk score of patients possible infected during the simulated outbreak. In multihospital models, the simple within-hospital module has been replaced by a more complex sub-model using detailed within-hospital information based on patient movements between and within rooms in different wards. With simulated paths of infections, one can measure network properties of patients. I obtained patient's risk of being infected, by calculating his/her in-centrality. Patient's risk of sending infections comes from his/her out-centrality (both analogous to Google Page Rank algorithm).

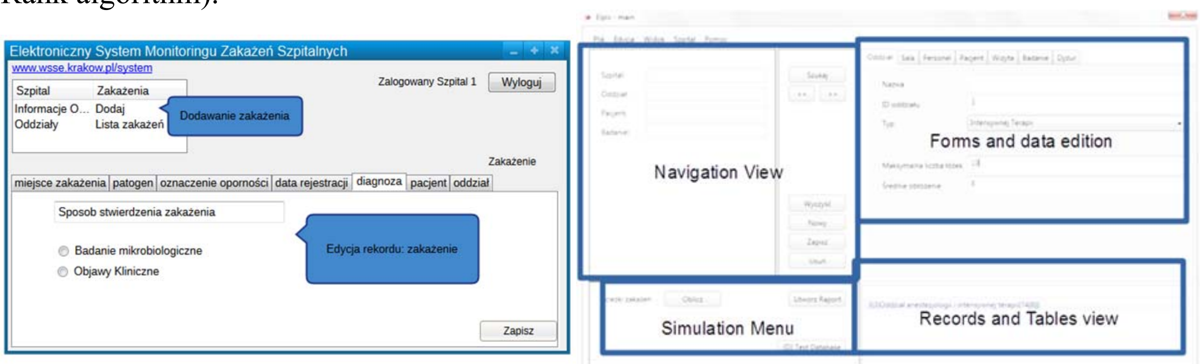


Fig. 29. Graphical interface of SIRS-Z application for institutional user [Left] and hospital user [Right]

The current trends in modeling are to increase the understanding of how diseases are transmitted on multilayer networks on hospital example. Methods to analyze the contact network of persons visiting the same hospitalcare have been already applied (Liljeros, et al., 2007). However, understanding in which way organizational – topological network structure affects the transmission of infections is still missing, because previous approaches have been homogeneous in terms of used methodology. Another possibilitity is to extend the computational model of inter-hospital disease spread (Belik, et al., 2016). In my procject "SIRS-Z", I have assumed incorporation of the detailed contact structure on intra-hospital level as obtained from Polish hospitals via surveys and co-location analysis. An additional software module to account for the population stratification and mixing is being developed.

II Networks and data modeling

"Sociologie=?Physique sociale"

August Comte, 1839

Carefully selected state-of-the-art methodology and tools of social network analysis and complex networks science will permit a multifaceted in-depth exploration of my complementary research foci

- epidemic and diffusive processes described in previous chapter;
- population and dynamical model in next chapter;
- the network itself presented in this chapter

Complex networks and Social Network Analysis (SNA)

The network theory is useful when it comes to the study of nature from a systems perspective, and there are several examples, in which it has helped understanding the behaviour of complex systems. Genetic regulatory networks, Internet transfer protocols, social interactions and financial market dynamics are some examples, in which a network perspective is important to understand systems behavior (Fronczak & Fronczak, 2008). 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' structure and the interactions between their individual elements. However, physics methodology proves helpful in many issues of complex systems properties including the 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 artificial concept developed by physicists and mathematicians and (at least in theory) they obey universal rules. Complex network analysis not only helps better understand social behaviour and determine the degree to which individual agents build functioning and working system, but creates quantitative 'machine learning' approach for collective intelligence. However, a social network analysis technique has been used to support social theories in qualitative way by social scientists. The very concept of a social network was developed back in the 19th c. by Durkheim, who compared the structure and functioning of societies to biological systems consisting of interconnected components. In Poland, Bronisław Malinowski (1924-44) combined anthropological study with knowledge from the borders of psychology, mathematics and economics, trying to get a better grasp of the functioning of the world. Moreover Jacek Szmatka his team at Jagiellonian University in Cracow actively participated in 1990th in the development of Social Network Analysis as the one of first lab of this kind in Europe. Currently, SNA serves as a methodology and set of tools enabling a multifaceted in-depth exploration of interacting systems. A graph formally consists of set of nodes (vertices) V - v_i and set of links (edges) E - e_{ij}.

In network terminology:

- Node = an individual components of a network e.g. people, power stations, words (Kulig, et al. 2013), neurons, etc. In my thesis papers nodes are patients - I.2.MRSA, music artists - II.3.Music-NET, countries - II.5.Music-diffusion, characters in novel – II.7.Fiction, hashtags in Tweets (Jarynowski, Rostani, 2013), Political Parties – (Jartnowski, Buda, Piasecki, 2016), etc. In standard graph theory it has notation v_i.

- Edge = a direct or indirect link between components referred to in social networking e_{ij} as a relationship between two agents), could be weighted.
- Path = a route taken across components to connect two nodes (it is a very important issue for projects on music II.5.Music-diffusion, in which we were looking at an influential culture spreader, and I.2.MRSA, in which infections paths are were our main goal of investigation). An example of a possible path is visualized on Fig. 9. Path searching has a long tradition in mathematical graph theory since 18th century Euler problem of bridges in Królewiec (Koeningburg).

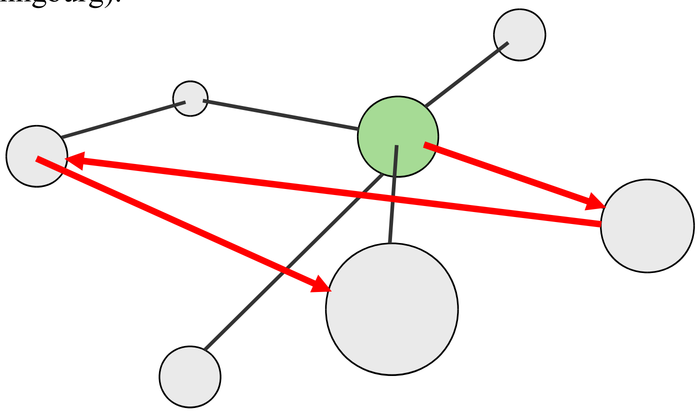


Fig. 30 Visualization of a path between nodes (arrowed links as a possible transmission channel

- Centrality = metric such as node degree, node centrality and betweenness centrality. By using some of these metrics it allow to measure both the number and strength of interactions and position of an individual. I obtained in I.2.MRSA the risk of getting infected, based on the patient's incoming connections, and the risk of spreading infections resulting from outgoing connections (both analogous to Google Page Rank algorithm).

To operate some practical properties, we must introduce some intermediate variables. The neighbourhood of a node *i* (N_i) is defined as its immediately connected neighbours. K_i is the number of neighbours of a node (degree). Let consider, we have algorithms for find the shortest path.

Typical network properties are also listed:

- Community's structure [Fig. 3]. Community detection algorithms serve to automatically identify sub-groups within the observed population. A partition of network is a classification of the nodes that each node is assigned to exactly one of selected communities. Community is a social structure with connectivity within sub graph is higher than with the rest of the graph. There are several ways to make partitions and it is a computationally difficult task. Algorithms for finding communities can be of different type as: Minimum-cut method, Hierarchical clustering, Girvan–Newman algorithm, Modularity maximization, Statistical inference, Clique-based methods. In my studies, I choose just two very common community detection algorithms. Both, the Louvain Method and VOS Clustering for community detection are methods to extract communities based on function optimization once modularity once VOS quality.

- Average (shortest) path length – important for the flow of information in the network. Average shortest path length (L) in big enough real social was estimated as L=6 in S. Milgram experiment (Milgram, 1967) and 4.74 in Facebook analysis.
- Degree distribution- distribution of connections of nodes. There are some characteristic distribution (mainly long-tailed), which fit very well to empirical networks [Fig. 10].

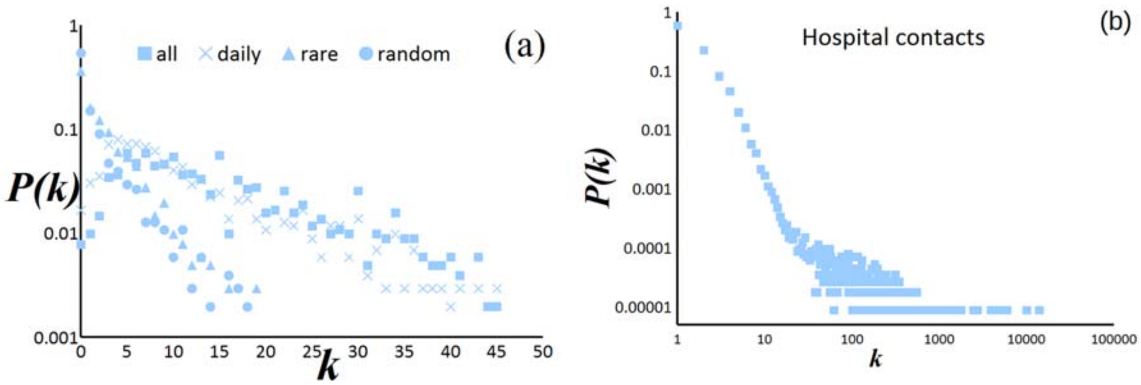


Fig. 31 Two characteristic degree distributions of in contact networks. (a) From POLYMOD contact survey for various contact loyalty used in study I.3.Polymod with exponential tail. (b) From Stockholm hospitals dataset I.2.MRSA with power-law tail.

- Assortativity: the assortativity coefficient is the correlation coefficient of degree between pairs of linked nodes.
- Clusters: A clustering coefficient counts the number of triangles [Fig. 32] in networks (e. g. I calculated clustering coefficients of musical artists' networks or respondent contacts in POLYMOD). Formally, a local clustering coefficient is defined by:

$$C_i = \frac{|\{e_{jk} : v_j, v_k \in N_i, e_{jk} \in E\}|}{k_i(k_i - 1)}.$$

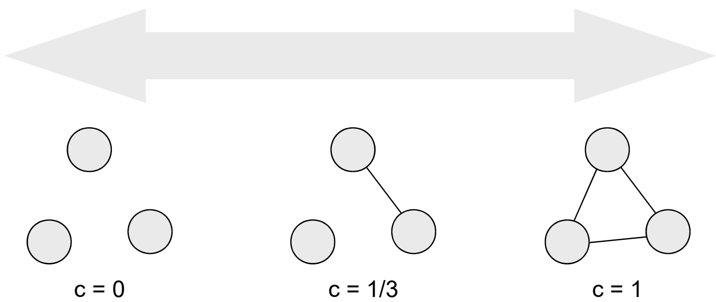


Fig. 32 Visualization for the clustering coefficient calculation

- Randomness: From a Grid/lattice network (ordered structure), via Small-world network (a mix of order and randomness), to Random networks (usually Barabasi-Albert: BA with power low degree distribution or Erdos-Renyi: ER with exponential degree distribution).

Classification and clustering. Partitioning as set of tools of exploratory data analysis that attempts to assess the interaction among patterns needs much more clarifying. If we do not know which nodes belongs to which cathegory such kind of partititioning is called clustering – wgere within clusters nodes are closer to each other, than those whose patterns belong to different clusters. “Groups” formed based on role of externally heterogeneous corresponds to classification procedure.

Dynamics and temporal aspects. However, those systems change in time at different scales and in different ways. The dynamics from this perspective have not been studied in detail and an integrative framework is missing. In this thesis, I propose a dynamic model that integrates contact (from empirical studies of HAI and STI) according to the network theory with rule-base modelling. I will apply and the test model with the quantitative data already available (described in details in the next chapters). We propose that my model represents a general and scalable approach to studying interactions in order to understand their structural and dynamical properties, allowing making predictions, which with some limitation can be linked to theoretical results. Temporal network (not translated officially to Polish until now) is a new concept, and from the point of view of the complexity science, only a theoretical basis exist (Holme et al., 2012).

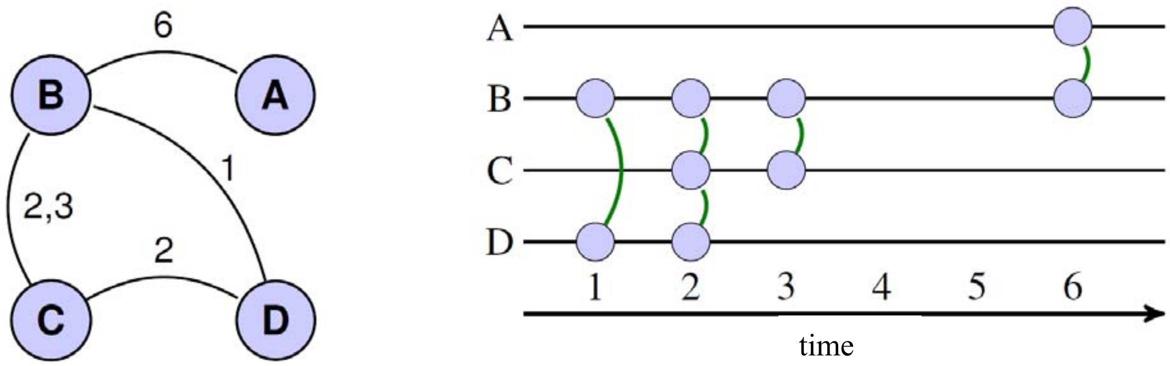


Fig. 33 Concept of temporal network, in which links are dynamic

Network theory is applicable to a wide verity of phenomena occurring in different scientific disciplines, such as particle physics, biology, economics, sociology and computer science. In order to create a dynamical model of a network, we need to have rate parameters, i.e., quantitative descriptions of how the components interact in time [Fig. 33]. Dynamical network models are important when we are dealing with complex systems, in which there are a significant number of nodes (Grabowski et al., 2010) and interactions that show non-linear behaviour. Models assuming interactions at the network level have been proposed [Fig. 35], however, an integrative framework of both theoretical and empirical approaches is missing.

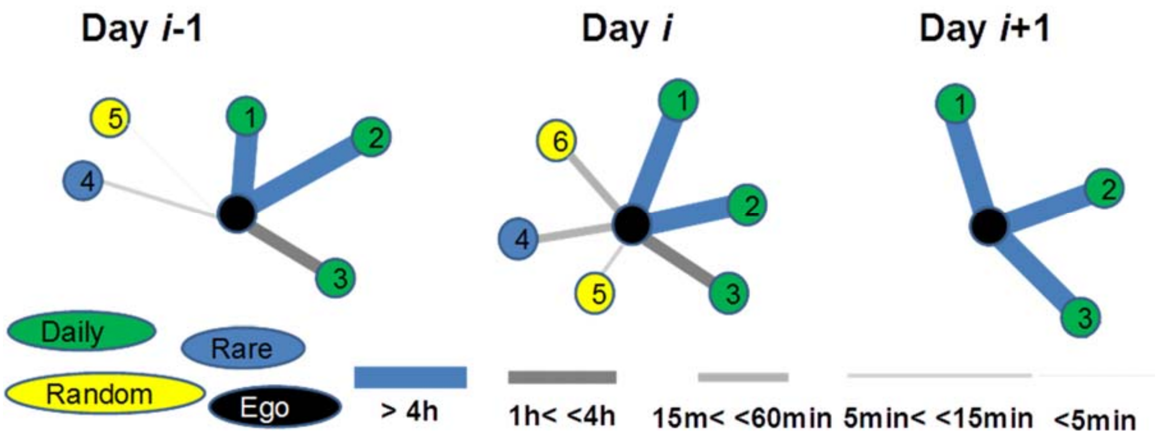


Fig. 34 The example of real evolving network. Time slices of ego network in POLYMOD study (I.3.Polymod) with different categories of intensity and loyalty of nodes

Overall, the use of an empirical network in a model offers a chance to enter the current research trends, especially it raises the issues that are usually under-represented, i.e. the comparison of the theory with empirical data (Leskovec et al., 2007).

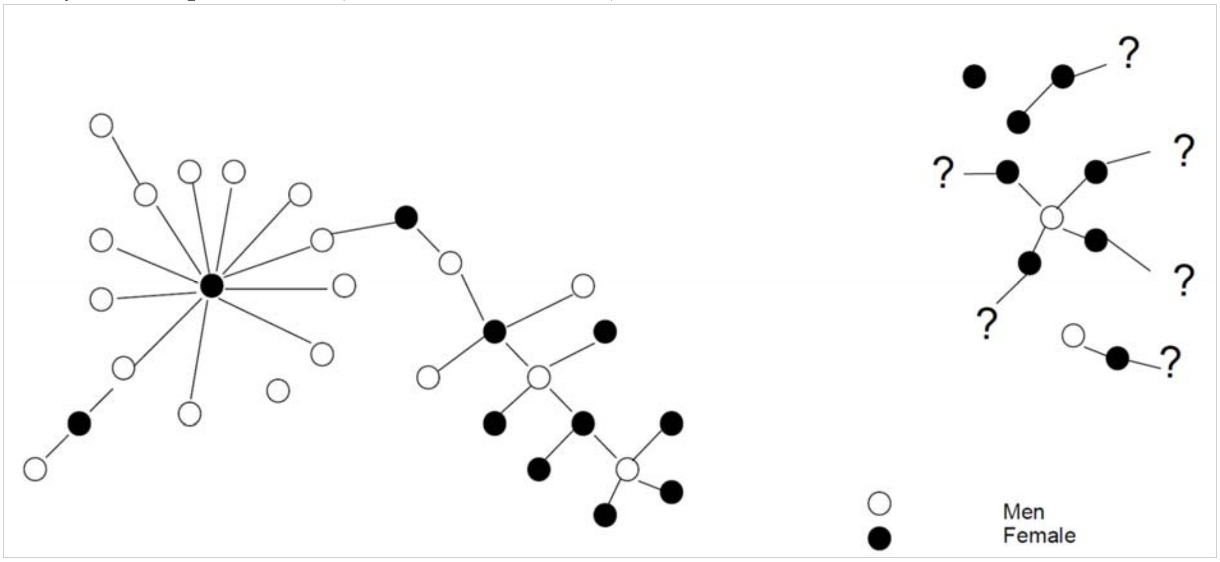


Fig. 35 Theoretical sexual networks (based on Liljeros, et al. 2001)

In my research, some of the most important network-based questions are:

- II.3.Music-NET: This is the type of network of relationships between music artists (stardom - cluster formation, creating new music trends, etc.)
- I.1.HPV-pl: Analyses of already collected sexual network data around the world and analyses of our own data and the application of temporal network methods for different scenarios for the most important pathogens spread
- I.2.MRSA (and further analysis): Is it possible to estimate transmission probabilities in contact networks based on empirical data on diagnosed cases and network data?

• Could these models help to localise unknown carriers (and spreaders) or places?

• Are there properties specific to people contact patterns or to positions within a contact network that are correlated with a higher risk of getting infected?
• Does a different pathogen display different transmission patterns?

I have investigated hierarchical structure in various complex systems according to Minimum Spanning Tree (MST) methods. To do so, I take for example from that matrix based on two vectors containing P_i and P_j: the time series of signal i and j. The matrix C_{ij} of correlation coefficients is a symmetric matrix and the $n(n - 1)/2$ correlation coefficients characterize the matrix completely. The distance between assets is transform to Euclidean metric. However, not only correlations can be used to create distances, but other kind of underlying signal D_{ij}, can be applied (like Manhattan measurement). The final matrix of distances can also be used in order to classify the artists into clusters on MST (for example by Kruskal algorithm). Presented trees have meaningful interpretation [Fig. 36 right], also in terms of tree life-time [Fig. 36 left]

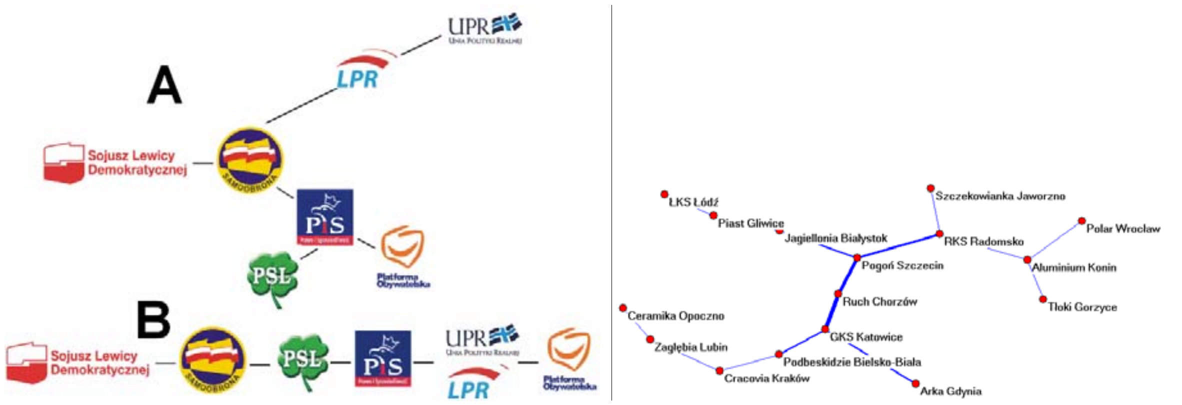


Fig. 36 (left) The MST obtained for Polish political parties according to opinion polls A) 2006-2010 (before the Smoleńsk crash), B) 2006-2013 (after the Smoleńsk crash). (Right) Hierarchical representation of First Polish League season 2003/2004 as an average of signal is ranking (place in table) and result in game (Jarynowski, Buda, 2015).

Network analysis from sociological perspective

"Żeby móc zastosować techniki matematyczne w praktyce trzeba spędzić przynajmniej 1000 godzin zagłębiając się w specyfikę danej dziedziny, przy czym ponad 100 godzin na rozmowach z jej przestawicielami, np. inżynierami, lekarzami, socjologami."
Hugo Stainhaus to his students at Wrocław University from Roman Duda memorials (paraphrase)

In this chapter I would like to fill the gap between sociology and physics. Unfortunately, most of physicists within the field of sociophysics are not really interested in the "state of art" in sociology. I would like to state here, that one of the founders of SNA was Jacek Szmatka – professor at Jagiellonian University in Cracow (Szmatka, 1997), but he died too young, so had I no chance to meet him. The main facts we already know and review were proposed in the most influential studies.
Complex networks have been well an established object in science and social science since late 1990s, because of new properties of Barabási-Albert (BA) – scale-free or small-world networks – Watts-Strogatz (WS), while previously standard random (exponential, p*) graphs like Erdős–Rényi (ER) were already in use in sociology. Social networks are found to be fundamentally different from non-social networks such as the World Wide Web, power grids and airline networks. In this chapter I will pick up few important specific and universal characteristics of social networks. Many studies on empirical networks have revealed that, in addition to common properties of real-world networks such as the small-world effect and skewed degree distribution; social networks often demonstrate a positive degree correlation, i.e., assortative mixing, and clear community structures (Lu, 2013). Physicists or mathematicians have played a huge role in the discovery and investigation of all mentioned network types. However, physicists working on this subject very often forget the impact of sociological studies on this area, mainly Granovetter and Coleman or Merton in mid-20th century. The sociogram, introduced by Moreno in early 20th century, were very broadly used in social studies (even I was taught how to study a social structure of a group of students during a course on teaching physics).

Discovery of new network properties

The avalanche of interest in complex networks starts with two very important papers (Barabasi et al., 1999), (Watts, Strogatz, 2000). Proposed network structures are canonical in the field, so I have also studied their properties in the BA, WS and ER networks within social norm change project (Jarynowski, Kulakowski, 2011) and in the marriage dynamic project (Jarynowski, Nyczka, 2014).

Scale free network. Barabási-Albert reconstructed the network structure of WWW (Barabasi et al., 1999). They constructed a computer robot (program), which crawled every possible webpage inside the Notre Dame Univ. domain and checked out all hyperlinks (more than 1.5 million links) and found a power low of degree distribution of webpages. They claimed that such a power law was a consequence of two mechanisms: network growth and preferential attachment. They observed that real networks developed due to stochastic processes, in which nodes were chosen randomly to build a link. Huge hubs were consequence of these processes: the more webpages pointing to a given WWW, the more people going there and the higher probability that their pointed to it on their own page. Preferential attachment is an example of a positive feedback cycle. This is also sometimes called the Matthew effect, "the rich get richer". The description of the constriction of such a network is well known by now and gets names from the authors: BA.

Small world network. Sociologist Watts and mathematician Strogatz differentiate networks according to their randomness (Watts, Strogatz, 2000). They introduced a model that showed properties of highly clustered networks with low average shortest path. Such a small-world network is classified between a regular and a random network. They started with a regular network with repeating-type topology (kind of chain with additional links between closed neighbours). There is no randomness in such a network (ran=0). They introduced the parameter p, which can be understood as the level of randomness. There is a full random network (ran=1) called the Erdős–Rényi (ER) graph in the opposite corner. Small-world network is in-between. It is based on “local links”, but some additional “long range links” were added (with probability p that a local link is rewired to a long distance one). The most interested effect is observed for small ran, in which there is only a little of “shortcuts”. In such a setup clustering confidence is still high, but the average shortest path is much smaller than in a pure regular network. It does so by interpolating the ER graph and the regular ring lattice. Consequently, the model is capable of at least partially explaining the "small-world" phenomena in a variety of networks.

Diffusion on networks

Processes in networks have become very important tools to the analysis and spreading of innovation and ideas (Leskovec et al., 2007). Human dynamics has a strong influence on the processes of spreading in social networks, e.g. the intercontact times with heavy-tailed distribution affect the spreading of virals. Let us recall some very old and important studies that cover the issue recognised by physicists only recently. Social diffusion could be modelled in a different way than physical diffusion in epidemiological modelling.

Diffusion of innovations. James S. Coleman and colleagues conducted a study on the diffusion of a new medical drug in 1956 (Coleman, 1957). They asked their respondents to list their interpersonal connections in order to investigate the effect of interpersonal network links on the new drug adoption. The result shows that the percentage of adoption of the new drug followed an S-shaped (logistic) curve, but that the rate of new drug adoption was faster than the rate of other innovations adoption. Early adopters were regarded as innovation traders by the rest of the medical society - late adopters. The researchers also found that the doctors that subscribed journals and gained their knowledge outside of the social network were characterised by a slower logarithm-shaped curve. One of the most important findings was that the doctors that had more interpersonal networks adopted the new medical drug quicker than those that did not. This means that interpersonal contacts are the communication channels that influence the adoption process. This study gives a theoretical basis for a divorce study of social change (Jarynowski, Kliś, 2012). I was testing the divorce rate increase speed in the second half of the 20th century using scaling exponents. My goal was to verify if the mechanism of change is caused by “Zeitgeist” (external fields) or herd behaviour (coupled network).

Social contagion. Nicholas Christakis and James Fowler suggest that obesity is “socially contagious” and spreading from person to person in a social network (Christakis, Fowler, 2007). Of course the transmission does not proceed by any pathogen, but by different kinds of social processes like: homophile, confounding and social contagion. Research was conducted on the longitudinal study of habits and health habits of Americans (Framingham Heart Study Network). The study showed that the social phenomena known as obesity, smoking, drinking, loneliness or happiness are connected to social networks of relations between peers. This means that people who have a lot of friends with similar characteristics (e.g. obese friends) were more likely to have this feature (are obese) and had an increased likelihood of getting such features in the future (will become obese). The authors found correlations between the properties of ego and its alters in the network of social ties. If an ego friend (alter) becomes obese, egos chances of becoming obese go up 57 percent. Surprisingly, the greatest effect is seen among close friends and not among people sharing the same household or sex. In conclusion, the authors mentioned that social effects were much stronger than it had been realised before. I am using the concept of social contagion in the modelling of viral spreading (Jarynowski, Jankowski, Zbieg, 2015).

Network brokers (Burt, 2004). Networks, like porous materials, may have holes. The physical theory of percolation has been already broadly studied to describe such phenomena as forest fires (disconnected chains of trees would stop spread of fire). In networks, holes could be bridged to develop an information spreading capability. In social networks they are called 'structural holes', formally defined by Ronald S. Burt, as 'when two separate clusters possess non-redundant [that is, complementary] information, a structural hole is said to be between them'. The author analysed networks around managers in a large American electronics company. A broker that bridges the holes is a very important agent. He could gain competitive advantage by engaging in information arbitrage. The good news is that the broker plays a powerful role in the network; the bad news is that he could be the single point of failure. The author also touched the problem of creativity and the spreading of ideas for managers with discussion partners. Connecting networks of entrepreneurs with their own strong and weak links by bridging structural holes may provide access to valuable new ideas, alternative opinions and practices, early access to new opinions, and the ability to move ideas between groups, where there is an advantage in doing so. To summarise: Burt's approach to understanding how an actor is embedded in its neighbourhood is very useful in understanding power, influence, and dependency effects. According to my own research on the spreading of influences in European music (II.6.Music-diffusion), the broker is Netherlands, while the biggest hub is the UK.

Multilevel networks. Influence dynamics between people and regions can be reconstructed as a diffusion process (Hedström, 2000). The spreading of unions in Sweden in the beginning of the 20th century was analysed. In this paper, the authors introduced standard geographical diffusion shortcuts. Those shortcuts were agitators traveling from town to town and establishing new local units of the Social Democratic Party. The speed of party development emerged out partially of the travel routes of political agitators. In this article, the authors have focused on two types of mechanisms: imitation and persuasion and on two types of actors: workers and agitators. Studying real-world diffusion dynamics can also be important to the unveiling of influence dynamics between agents, as it was shown by the authors. They study diffusion using the concept of internal and neighbourhood influence (micro-level) and social networks of agitation (meso-level): the study of the local or global properties of communities that could modify the dynamics of union diffusion within these communities. Such phenomenon was previously described by Merton and its main point of the middle-range theory. The proposed regression model suggests that meso-level agitation is significantly more suitable for empirical data. It was showed that for example the concurrence with free churches had a negative effect on socialist action. Analogically to traditional epidemiologic models, the spreading of party membership depended on both the micro and the macro level, as well as on the concurrence with other pathogens (like free churches) appears. The multilevel network approach used in this article as a pioneering technique is very popular nowadays, as in my paper on networks during riots (Jarynowski, Rostani, 2013), in which 2 layers of conflict were suggested.

Network topology and evolutions

Network structure and interdependence have a very important influence on processes in the network. In biological networks like gene-regulatory, the function of life comes from the expression or blocking of genes connected in very complicated ways. Moreover one type of networks can influence another and for example cause cascade effect like blackouts, the connection between the steering of the power plants network (which needs electricity to work) and power grids. Typical forms of networks were described in the previous chapter, however here I explain how they are formed and what their properties are.

Topological evolution. The formation and evolution of networks is observed by monitoring the addition or deletion of all links. Sociologists analyse quantitatively the tendencies used to create ties in these evolving online affiliation networks (Gallos, Rybski, Liljeros, et al., 2012). 5 main mechanism of link addition were proposed: 1) reciprocity (named social exchange), 2) friend of a friend ties or closing triangles (balance), 3) exploration of distant network areas that require at least 3 steps from the position of the person in the current network, 4) ties facilitating dissemination of information by linking to well-connected people (named collective action), and 5) links that act as bridges between two clusters that are not directly linked (structural hole mechanism). Following the rules was estimated on two Nordic datasets QX and POK to capture the statistical properties of real-world networks related only to node-to-node link formation. Even due to the fact that links can be formed for various reasons, snapshot analysis cannot often differentiate that. The authors attack the well know p* models – exponential graphs, very popular in other sociological studies (Robins, et. al., 2007), due to their lack of evolutionary interpretation. An example of the triangle was presented, in which the same link could be classified as Distant and Balanced, even in their complementary probabilistic interpretation. They also showed the variation of the mechanism due to sex (e.g. men are more prone to collective action, but women to exchange), age (younger users are more prone to exchange, but older to balanced). To conclude, the authors presented a generative model for creating social and affiliation networks, which could be used in many applications. I have investigated the evolution (increase) of a network in literary fiction (II.7.Fiction), but from a narrative perspective.

Link maintenance costs. Many researchers have begun to recognise the value of communities and theirs network perspectives (Holme, Edling, Liljeros, 2004). The mentioned paper is one of the first broad empirical studies on the time evolution of social network. It describes a methodology of tracking the progress of communities over time in a dynamic network. While the creation of communities around specific topics (dating in this case) is an important phenomenon, the authors used Social Network Analysis tools and technologies to describe it. They looked into friendships, messages, guestbook notes and romantic relations. Some questions were answered: how to measure the properties of a community? The first effect that they found was a little controversial (fat tailed degree distribution of social links without cutoff). Such a lack of cutoff is observed very rarely and in this social platform it comes from the low costs of link maintenance (it is not common in real networks). The main point of the article was about assortativity (more often called homophile).

Most assortative mixing coefficients were negative, suggesting an affinity to differentness. Another important finding was the formation of triangles and cycles (quasi-clustering structure). Parameters from the time-evolution of both assortative and quasi-clustering parameters to regimes could be classified. First, start up as non-equilibrium, and second, adult as equilibrium. The authors predicted correctly that the area of Internet communities would be investigated very widely, which we can observe now. I have investigated natural resource management networks of fishermen in Madagascar from this perspective (Jarynowski, Buda, Nyczka, 2014).

Bridges and percolation. Granovetter defines a tie (and its strength) as "a combination of the amount of time, the emotional intensity, the intimacy (mutual confiding), and the reciprocal services that characterise the tie" (Granovetter, 1973). While other scientists concentrate on strong ties, like family membership or friendship, social processes are driven by much weaker links (for diffusion across a network it is the weak ties that are the most valuable). The role of weak ties is crucial in many social phenomena. He raised the argument that social networks show the value of weak ties based on empirical data (e.g. based on a study on job search). He claims that in particular in case the ties are strong between two people AC and AB, an additional weak tie BC should also exists and sometimes can act as a "bridge". He briefly discussed the idea of an “absent” tie; this is a potential relationship (which, if it does not exist now, should appear in future). Some basic dynamics of these ties allow us to predict their evolution (like the prediction of a weak link BC). He argued strongly that “no strong tie is a bridge.” He also showed that job contacts were often signed via those weak links. Granovetter tried to connect micro-sociological analysis of a network of small groups or families with social structure. The small-scale interactions between people can give you a lot more insight into the macroscopic behaviour of the entire social network. The paper is extremely cited. I have confirmed his theory by analysing the spreading of a viral in the Stop-ACTA movement (Jarynowski, Jankowski, Zbieg, 2015).

Focus networks. Feld provided a good basic model for how general social networks are organised and grow from the focus choice (foci) perspective (Feld, 1981). The author moved the analysis of social networks by providing a framework to explain the formation and evolution of ties. He tried to support the foci theory with the Detroit Area Study, in which he showed pattern formations (if two nodes are linked, their alters are more probable to create links as well due to overlapping activities in the same focus). The multiplexity of relations was also taken into account (two people can share more than on focus, while they already have something in common). Individuals tend to group themselves around foci of interest: these can be “social, psychological, logical, physical entities”. Feld assumed that all efforts in ties were positive (positive feedback). The author claimed that the social organisation of ties does not indicate an individual thinking, but rather a collective one. He showed focus formation based on family, neighbourhood or profession relationships, but mechanism could be generalised in other forms of relations and actors could be not only people, but also nations, for example. This observation has a very important implication to the universal network analysis, because different types of contacts have different impact on social processes, which is usually neglected by physicists in their research.

Sexual networks. In another heavily cited short paper (Liljeros, et al., 2000), the power-low of sexual activity was discovered. Such a conclusion was stated based on the analysis of survey data from Sweden within adult population (with sample size: 4781 and responding rate: 58%). Two important questions were asked – about the number of sexual partners in life and during the last 12 months. Preferential attachment was proposed as a mechanism of occurrence of the multiple partner behaviour (someone that had a lot of partners in the past, he/she is more likely to have more in future). A heavy-tailed distribution of contacts has important consequences for the spreading of infections. There are hubs – people with enormous numbers of contacts – that should be observed by health authorities. An inconsistency between male and female reporting was observed, but it is a known artefact of the poll methodology (males overestimate the number of partners, while females underestimate that). The structures found are named webs of sexual contacts and under those names they entered the classical literature on networks and this example is often used to promote a universal physical law (scale-free property). Unfortunately, the holism of natural mechanisms does not work too well with sexual contacts, because further papers from British or Brazilian studies do not confirm the full power law and exponential cut-offs were needed to fit the data. However, this break-through article built a new field of STI epidemiology – network epidemiology. I used the same dataset to train model parameters of the STI project (I.I.HPV-pl), (Jarynowski, 2015b).

Temporal networks. Temporal networks (Holme, et al., 2012) touch many dynamic properties of genetic regulatory networks, Internet transfer protocols, social interactions and financial markets. In his paper, Peter Holme even pointed at a patient network in hospital as a possible transmission path of HAI (Hospital Acquired Infections). I worked on exactly the same dataset (I.2.MRSA). Like many other human behaviours, sexual networking (especially the temporal aspect like in Brazilian prostitution) is a complex phenomenon (this is also a subject I am working on now). Science has developed a set of mathematical tools (network and statistical analysis, modelling and simulations) that have already been successfully used in many fields from the graph theory to system dynamics. However, the concept of theoretical temporal network is basically new from the general perspective and Peter summarised the theory and qualified the empirical data.

Networks are everywhere: While many social phenomena can be describes by networks in "big data" time, we can extract networks from significant events like riots. I investigated Twitter and Internet Forum during the riots in Stockholm in May 2013 (Jarynowski, Rostani, 2013), using Natural Language Processing (NLP). Our aim was to provide a quantitative analysis of how people discussed Stockholm riots and what were their stimuli to such activity.

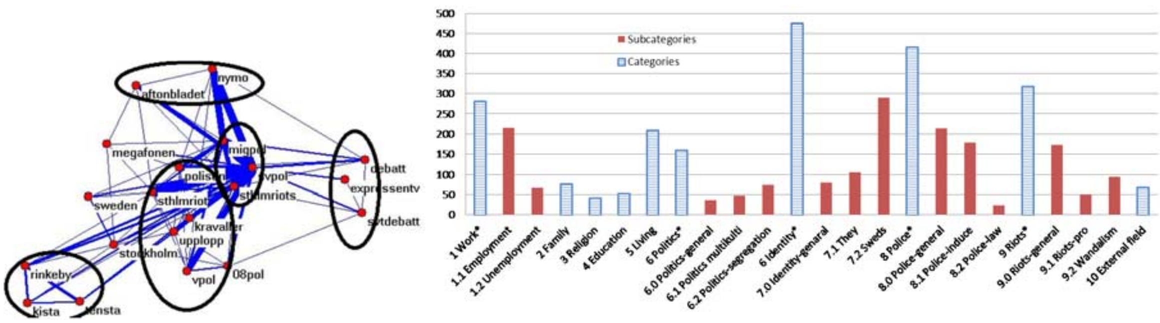


Fig. 37 [Left] Network of co-occurrence between hashtags in Twitter [Right] Categories and categories containing subcategories. * means sum of all count

There are at least 2 layers of conflicts in social media:

- the first layer, the one closer to rioters, should focus on the anger with the establishment represented by Police (Internet Forum);
- the second layer, the more abstract one, should provide antagonism in the world of politics (like Twitter).

Dynamics on networks

The network of contacts is usually the only layer, in which processes take place. Many social phenomena described in more detail in the thesis papers, like the Stop-Acta movement, could appear only because of networking platforms.

Threshold models are a basis of many further models. In very brief, a person's decision whether to do something or not (like to take part in riots, for example) depends on what everyone else is doing. Doing something different requires a critical number - a threshold (Granovetter, 1978). This threshold is assumed to be distributed in society; it is thus different for different people. The outcomes may diverge largely, even if the initial condition of the threshold differs only very slightly. He used the example of the riots to illustrate a phenomenon. There are 100 people, in which each person has a threshold different from everyone else, meaning that the range of thresholds' from 0 to 100. A cascade can be observed, in which someone with a low threshold will start rioting, and then will be joined another with a little higher threshold, and so on. It can be applied to other social phenomena and term can been changed to refer to processes like voting, the migration process of crowd behaviour and social movements. For the formation of norms, the author assumes a simple relation between collective results and individual motives. That is a model limitation, while collective behaviour can take place in a set of people, who share some common traits (e.g. young boys, who decide to steal a car). As a main conclusion of the paper, an equilibrium point can be predicted for how many people will choose one side of the binary decision (of course, the author did not fit any data and his investigation was only theoretical). Threshold models are also broadly used in system biology and I have proposed one, which describes calcium spiking in plants (Morieri, Jarynowski, et. al., 2013).

Criminal networks. This study used the network analytical approach to examine co-offending (Sarnecki, 2001). The aim was to test whether a network perspective could provide a fresh insight into the character of crime in the metropolitan area. Professor Jerzy Sarnecki (criminologist from Stockholm University) proposed a data set from special approved homes. Theoretical perspectives see the causes of crime as partially or completely associated with an individual's ties to different types of social networks.

Small-world property of real social networks by Milgram. In his famous experiment (Milgram, 1967), he tried to learn more about the probability that two randomly selected people would know each other (to find the average path length between any two nodes). The procedure was as follows: a source person in would be given a letter to deliver to a target person in Massachusetts (with basic information about the target); the source would then be instructed to send the letter to someone he knew on a first-name basis, who could proceed further and following recurrently until get to target. Out of 296, 64 letters eventually reached the target contact. Among these chains, the median path length was 5. Such small-scale and more meso-level social structures were unknown previously, while from different mathematical approximations the measured path in US varied from 3 to 100. Milgram mentioned that there are nodes more central like the US President, for whom all paths were supposed to be shorter. The example is an Erdős number in science, the distance in co-authorship of paper to Erdős (my number is 4). The small-world question is still a popular research topic today, with many experiments still being conducted. It has application not only to sociology, but even also to computer science (computer networks) or biology (animal networks).

Prostitution networks. The paper about sex workers in Brazil gave some unknown features of sensitive aspects of human sexuality (Rocha, et al., 2010). The investigation of sexuality is not an easy task. There is no hard data, precise research methods or fully verifiable results anywhere in the world. As in most similar studies, not the sexuality itself was targeted, but the health issue of sexually transmitted infections. Prostitution networks are built on the basis of the posts hosted by Brazilian online forums on paid sex. The authors studied both buyer and seller activity from the network and geographical perspective. On the basis of experimental data on interactions between humans, they have investigated the process of epidemic spreading in a social network (bipartite). Some of the most important findings were long-tailed properties of partner distribution (fatter for sellers). Another was the fact that a real network was more resistant to disease spreading than random one (geographical diversity). This way of studying statistical properties can be supplemented by modelling (the authors run artificial outbreaks on a given network). Some theoretical aspects of temporal networks were also studied. However the project is a little controversial from the ethical point of view, the results are relevant to a very important issue.

Propagation processes on networks from own perspective

This chapter correspond to “Other propagation processes” sub-chapter within epidemiological modeling part of thesis, but here network are on the focus. It is also extention of “diffusion processes” sub chapter, but here more recent view of the problem is presented.

Modeling Stop-Acta popularity propagation. I model the behaviour of an online community (spreading of virals) exposed to external impulses in two modes of experiment: with or without award, and on one natural (spontaneous organization against ACTA).

$S \xrightarrow{pe} E \xrightarrow{p} I$ $\quad$ $S \xrightarrow{pe} E$ $\quad$ $S \xrightarrow{pc} I$

Fig. 38 The analysed virtual world with a viral – avatar and model description: Susceptible (S) – Exposed (E) – Infectious (I)

The decisions related to the participation in viral diffusion are based on many different factors that are difficult to observe and monitor. We observed that emotional messages [Table 2] are deliberately targeted (high pe for Stop-ACTA) and less blindly forwarded (high *pc* for Stop-ACTA). However, the effectiveness (*p*) of an emotional viral is lower than in an incentivised case, but there is no significant difference between *p* for Stop-ACTA and both artificial networks.

Table 2 Model parameters calculated for campaigns

Event	proportion of unique invitations	probability of resending invitation per invitation	probability of resending invitation if get invitation
Epidemiological notation	probability of being exposed per contact (*pe*)	probability of infection per contact (*pc*)	probability of infection if exposed (*p*)
Non-incentivized	0.31	0.07	0.24 (std=0.23)
Stop-ACTA	**0.87**	**0.33**	**0.38 (std=0.28)**
Incentivized	0.14	0.07	0.48 (std=0.26)

Effects of termporal aspects on propagation with physical contacts.

The temporal contact patterns and network structures seen in the general population thus influence the dispersal of pathogens. In I.3.Polymod, it is assumed that individuals only meet in within definition of network obtainted from POLYMOD study and it is taken as an approximation of contact as in my case. Close proximity interactions can be also measured by electronic sensors. These data are increasingly used to inform contact networks. However the high level of temporal resolution of the networks built in this way requires novel data analysis and modeling tools (Obadila, et al., 2015; Valdano, et al., 2015), with the biggest respect to implemention the temporal epidemic threshold calculation (Valdano et al., 2015). We observe many phenomena related to the temporal aspects, such as the burstiness or fidelity. The availability of temporal contacts (mostly digitalized) is providing critical information to understanding propagation. A real temporal structure of contacts affects processes in many various studies in different way (Holme, et al., 2016) and motivated by theoretical studies of temporal networks, I explore set of models for possible parameter space to find out if process slow down/speed up for real temporal patterns.

Music popularity

The global phonographic market is a commodity market that contains several unique properties like periodicity of record sales or strong historical dependence and features competitions between record labels or artists, sofisticated critics polls in recommendation systems, customers heterogeneity and product lifecycles. The motivation of the present study is to find a kind of more general arrangement between artists – labels, critics and fans. I focus my activity on interdisciplinary research in the area of behavioral economy, sociology of music, social anthropology, but by using mainly: theory of complex systems, nonlinear dynamics, analysis of empirical data, and modelling of socioeconomic systems. So, I chacaterize the main properties of such a system:

- Phonographic market as a predictable system – in complexity and in particular for a single artist in II.4.Music-NET;
- Record industry might be considered as a programmable system in II.5.Music-StochProc;
- Even unpredictable recordings spread by the Internet do not break the equillibrium on the market (for example: Gangman style);
- The system might be unstable in case of artists death only;
- There is a strong seasonality, there are fans that buy records every week and occasional clients who buy music just before Christmas or St. Valentine's day;
- The violent increase of record sales characterize the phonographic markets all over the world;
- The trajectory of record sales decreases exponentially with some additional peaks caused by promotion including tours, singles, videos, etc;
- In the past, the peak of record/single sales has appeared a few weeks after the release, and sales have decreased slowly, but currently the peaks of record sales appear in the first week after the release;
- Popularity speads both via cultural and geographical paths II.6.Music-diffusion;
- Instead of pop, there is celebrity sector of genres.

Conclusions

"Omnia sunt physica, nihil est sine physico"

Paraphrases of Paracelsus

Researchers working in the physics paradigm postulate that in order to capture the complex and dynamic nature of social phenomena there is the need to merge non-structural factors into universal laws of universe. However, simplistic socio - or econo – physics models fail too often in predicting or even reproducing known sociological observation. For example many physicists claimed universality of power law degree distribution of social networks in the beginning of XXI century and it is on consider as a ground truth any more based on careful empirical studies (Jarynowski, Buda, Nyczka, 2014). On the other hand physics of complex systems increased our general understanding of the mechanisms of social dynamics and impact of social interaction. My exploratory research aims at investigating issues which have thus far remained virtually unexplored and applying a computational approach and a methodological complex system apparatus in a combination never before successfully carried out in nature. Thus, I used computational physics methods in some extend with complementary tools as regressions, structural equation modelling, social network analysis, etc. Moreover, the suitability and feasibility of the conceptually novel approach and applicability of the methodology to still new domains at least in Poland as sociology and epidemiology is promising and some results are very exiting.

My thesis is strongly interdisciplinary in its nature and I focus on solutions of real-live problems. Thus, for the purpose of the model building, I take advantage of various techniques of nonlinear physics, physics of complex systems, mathematical modeling, stochastic processes, intelligent systems, economic theories, innovation spreading, and psychological and sociological studies. Personally, I am the most proud of paper **(Jarynowski, 2015b),** because I provided clear recommendation to decision makers in matter of the best cervical cancer program in scientific sense. One of the major tools of influencing society is by regulation and policy. I have showed that the official profilactic program for cervical cancer in Moldova could be optimised in terms of costs and medical efficiency and I proposed a program, which was optimal the in simulation and contained:

- vaccination of targeted groups (the population effects are similar to high frequency screening schemes with EUR 1-1.5M savings yearly);
- screening only women in age 26-65 (with the prevention cost of EUR 5-12k per QALY - quality-adjusted life year), which could provide savings perspective in the range of EUR 150-300k per year in 10-15 years.

The knowladge of best strategies is very important to understand problem and to prevent misuse of resources. In propagation studies of Stop Acta (**Jarynowski, Jankowski, Zbieg, 2015)** and Music popularity spread (**II.6.Music-diffusion),** I posed novel research questions on influence patterns. I have also provided some numerical analysis in new field of temporal networks within series of papers on POLYMOD survey with example of **I.3.Polymod**. Another important contribution is marriage/divorce *ab initio* model **(Jarynowski, Nyczka, 2016),** which reproduces well real-life properties with only a single parameter.

The most promising research is in my opinion hospital infection project and in thesis only some preliminary insights have been presented – **I.2.MRSA**. While material resources in hospital environment are limited, creativity and the ability to create increasingly efficient ways to use resources is unlimited. Data used in my model already exists in hospital dataset, but it is wasted, while only administration and insurance companies can analyze it. Human health and life is the most precious gift. Data used in our model already exists in hospital dataset or human and animal health authorities. If data protection and e-ignorance wall falls, patients could be safer due to epi-intelligence.

Result obtained wihin this thesis ars summarized in the following SWOT analysis which also gives an author view about the strengths and weaknesses associated with this research.

Table 3 SWOT analysis of thesis results

STRENGTHS:	**WEAKNESSES:**
• Strong interdisciplinary collaboration between theoretical modelling with epidemiology and social science • Combination of different methods of modelling • Feasibility for epidemiological modelling proven in pilot study of STI for Moldova • Wide area of topics covarage	• More explanatory than hypothesis driven analysis • High heterogeneity of data from different sources (lack of interoperability of data) • Restriction of the modelling approach to specific networks or spread • Too simplified models and no analytical solution available • Results mostly on the border of significance – diffuclty in claiming satisfactory conclusions
OPPORTUNITIES: • Development of a generalizable hospital infection platform for investigating transmission dynamics of various pathogens (SIRS-Z project) • Several prizes for possible epidemiological applications (2x Polish Section of PhD Candidates in Medicine [Fig. 39, 41], Wrocław Days of Public Health [Fig. 5], Warsaw Days of Health Promotion [Fig. 4], Małopolska Incubator of Ideas [Fig. 14], OBSR ICT in Health [Fig. 26] and The First European Network Intelligence Conference [Fig. 6])	**THREATS:** • Complexity of the data and data gaps do not allow proper model validation • Using existing tools to solve problems instead of developing models ab initio • Problems in comunitating results to different kind of audience with various background and expertise • Over-intepetation of results without proper sensitivity analysis

Main findings of thesis papers

> “The USA is not richer than Poland, because Poland can afford raising and educating good theorists without having any profits from their work. USA cannot afford that.“
>
> Hugo Stainhaus to his students at Wrocław University in Aleksader Weron notes

In previous chapters, I have introduced the methodology and current state of art in sociophysics and computational social science within two non-linear concepts: epidemiological modelling and social network analysis. I have defined the methods, problems with the most spectacular achievements in applying techniques from computational physics into social science. Here comes the time to present my own research for a better understanding and predicting of social phenomena, as well as the optimum intervention strategies. It is organised as follows – I introduce the project by raising the main problem or the main idea. After that, I point to the main scientific result, which could be a suggestion, a recommendation or an explanation.

I.1.HPV-Pl Studying possible outcomes in model of sexually transmitted virus (HPV) causing cervical cancer for Poland

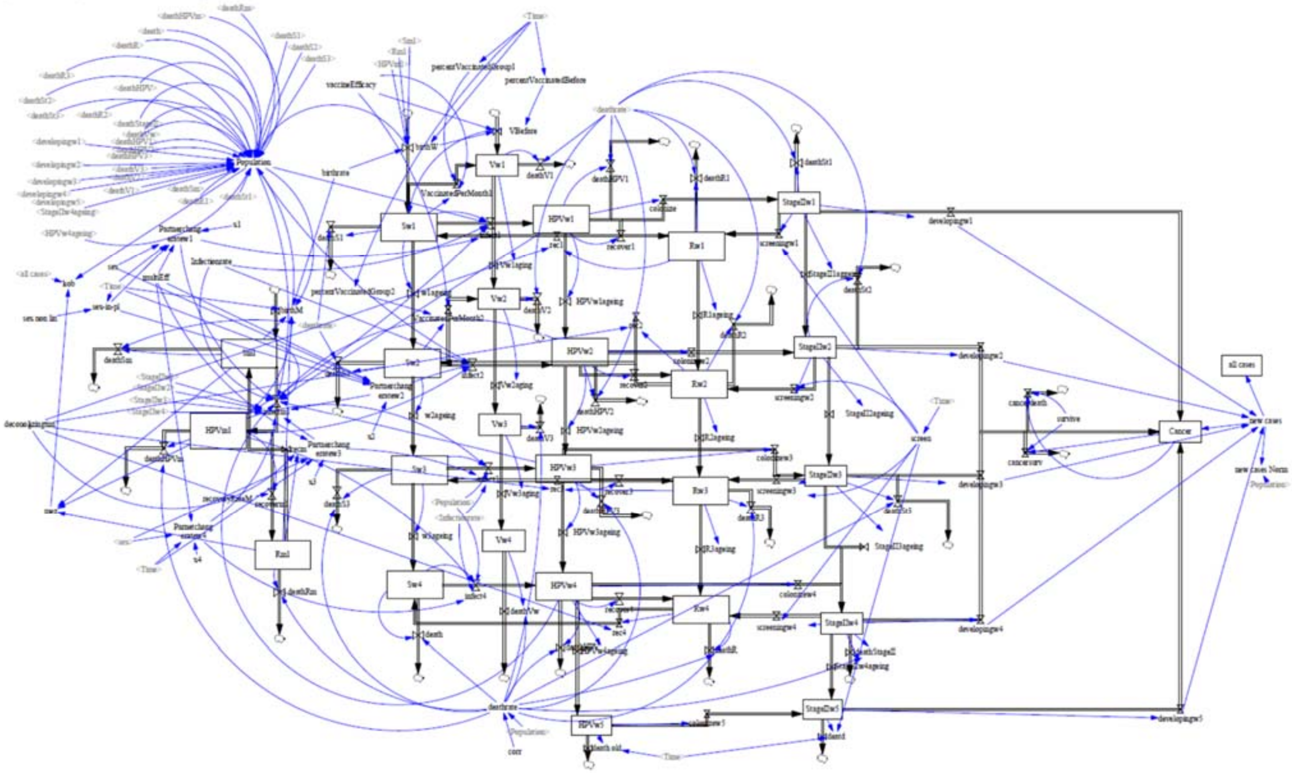


Fig. 39 Vensim model of the spreading of HPV in connection with development and preventions scenarios

PROBLEM and MODEL: HPV (sexually transmitted virus) is a necessary factor for the development of cervical cancer (second most common type of cancer in women in Poland. There is known optimal control strategy: a combination of vaccination and screening. In my model, I test different scenarios of: demographical change, increase of sexuality, screening freq.

DATA PROCESSING: No processing for demographics (Statistics Poland) and cancer cases (Institute of Oncology in Gliwice). In case of paremeters (some factors from literature review with the quotations):

- If more than one factor found in litererure – I calculated mean value;
- If range of accessibility different from value found in literature, then extrapolation is done. E.g. in literature cancer developing time form first cytological change is defined for women above 30 and below years old, but I have divided population in 4 age groups.

For sexuality patterns (Swedish/Finnish suvey), I assume that empirical histogram of partners numbers has the same distribution as in original data, but average partner numbers is adjusted (divided by around 3).

RECOMENDATION: he results should alarm authorities that some dangerous processes (taking place in Poland), like the ageing of the society and the increase of sexual activity, could recall an epidemic (the increase of incidence for some scenarios without proper preventative activities [Fig. 40].

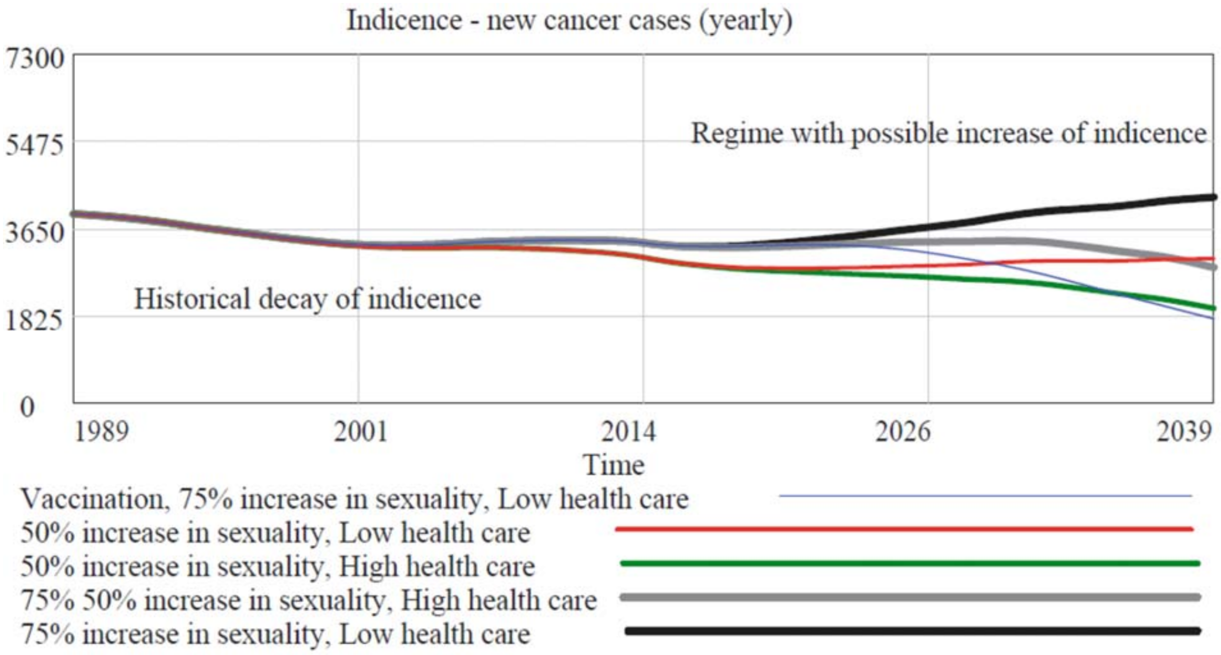


Fig. 40 Projection (daily Incidence) of new cancer case estimated for historical data (1989-2012) and possible projections till 2039

I.2.MRSA Contact networks and the spreading of MRSA in Stockholm hospitals

PROBLEM: The bacterium Methicillin Resistant Staphylococcus Aureus (MRSA) is resistant to more than half of all antibiotics and it is known to the single largest care-related infection problem.

For such infectious diseases, in which close contact is needed for the transmission to occur, an individual's position in the contact network is important for the person's risk to get infected. The awareness of the importance of contact networks has brought methods from sociological studies of social networks into the area of preventive infectious disease protection work.

DATA PROCESSING: No processing for MRSA positive cases (Infection Registry). In terms of patients visits (Hospital Information System), time span tables contain only those patient stay or come to hospital during at least one overnight and patiens with unknown codes were deleted.

SOULUTION: I studied the matrix of disease transition in hospital population. This matrix P was our first goal. In rows Infected and in columns People, who could transmit infection. The elements of the matrix are probabilities of what Infection would be transmitted by the indicated person. Unfortunately 1/4 of all infected are patients, who had no contact with any other infected person.
I had adjacency matrixes $\boldsymbol{P_t}$, $\boldsymbol{Q_t}$ and $\boldsymbol{P'_t}$ of contacts for each year and we estimated the parameters of infective strength (s, k, m).
The formula for an individual's change of state in one time step can be written as:

$$p_i(t) = 1 - exp\left(\sum \sqrt{P_t(i,j)s} - \sum \sqrt{Q_t(i,j)k} - \sum \sqrt{P'_t(i,j)m}\right)$$

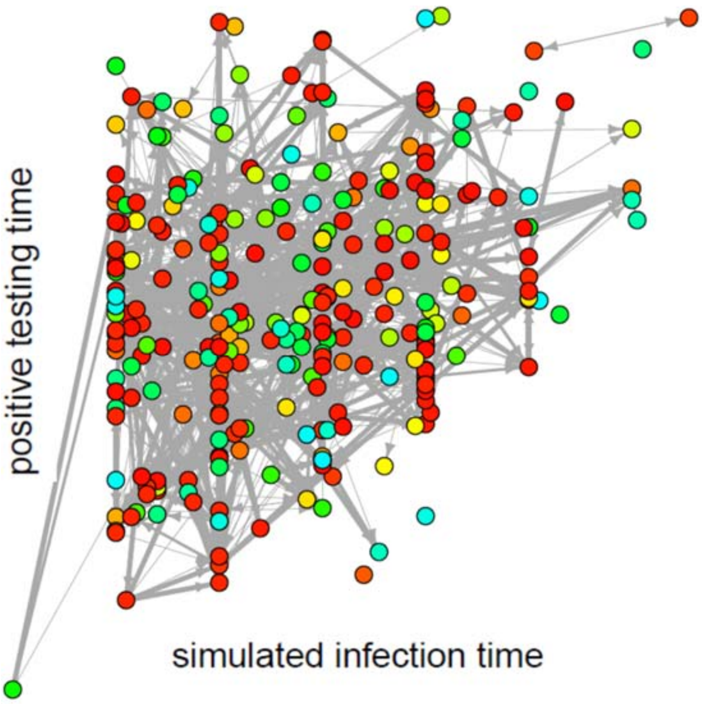


Fig. 41 Simulated paths of infection transmission vs time of positive tests on MRSA (Supplementary information not provided in the article)

My goal was to know, who could infect patient *i* and we produced the most likely paths. The presented paper is an early work on the epidemic model that can help understand the spreading of MRSA and the author of the thesis is currently the principal investigator in the SIRS-Z project (IS-2/195/NCBR/2015) developing a better model in Polish hospitals.

I.3.Polymod Rumor propagation in temporal contact network from Polish polls

PROBLEM: Real network temporal properties do not simply speed up or slow down contagious processes (Holme, 2012). We present effect of temporal as well as homophily feature of real networks (Mossong, et al., 2008). We used the same network data set in previous papers: (Grabowski, Rosińska, 2012) and (Grabowski, 2014), where we run epidemiological models and (Grabowski, Jarynowski, 2016) for Ising model of opinion formation. Here, we analyze rumour spread on real Polish contact data from Polymod survey.

SETUP: Ego-centric networks form survey contact (no data processing) are aggregated according to 3 dimensions of procedures:

1) Static vs Dynamic : rewiring rare and random contacts
2) Uniform vs Weighted: duration of contact as a link weight

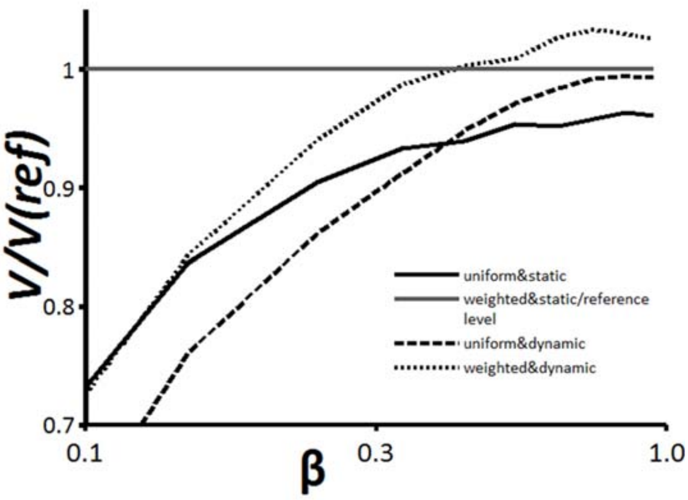


Fig. 42 The relationship between Total Outbeak Size (V) for different types of networks against reference size.

CONCLUSIONS: There is no general theory how temporal and homphily aspects influence the real networks, so we test realistic phase space of model. We gather results into the series of papers of wide defined contagious phenomena (Epidemic, Opinion formation and Rumor spread): Epidemic (SIR) (Grabowski, Rosinska, 2012; Grabowski, 2014),

- real dynamics – speed up
- real duration (weights) – slow down
- age structure – slow down

Opinion formation (Ising) (Grabowski, Jarynowski, 2016)

- real dynamics + NO real duration (weights) – favor spread
- NO real dynamics + NO real duration (weights) – inhibit spread
- age structure – no monotonic influence

Rumor (I.3.Polymod)

- real dynamics – slow down,
- real duration (weights) – speed up

II.4.Music-NET Network structure of phonographic market with characteristic similarities between artists

IDEA: I find a topological arrangement of assets traded on the phonographic market by using the Social Network Analysis of correlations between record sales of artists.

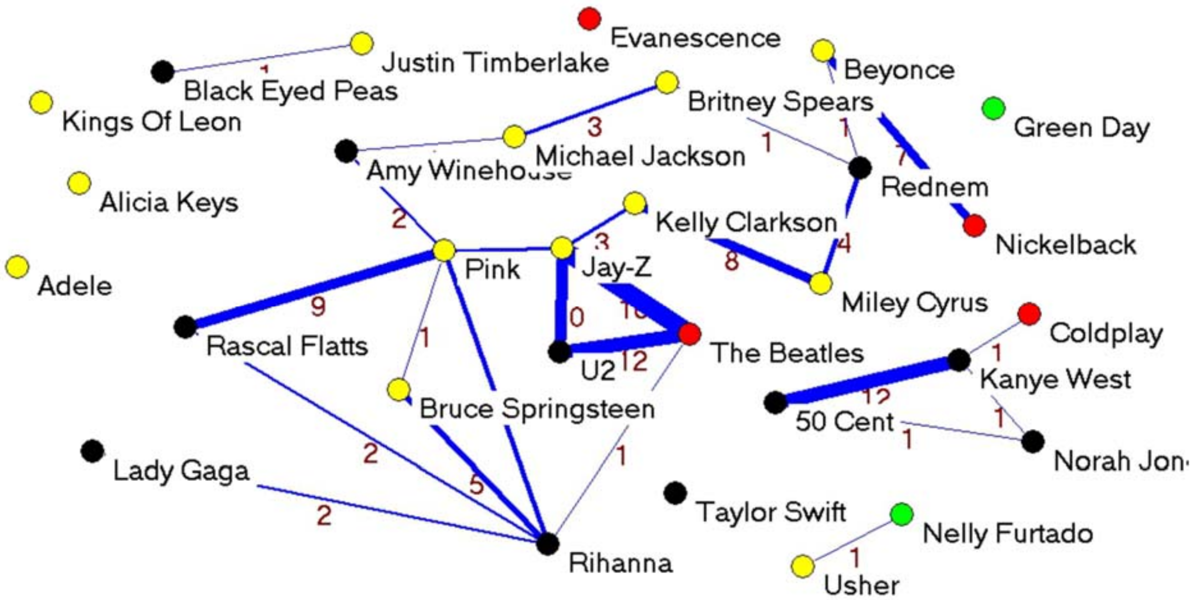


Fig. 43 Network of scaled positive correlations between artists (colors stand for record labels)

CONCLUSIONS: The Social Network Analysis of correlations between artists does not always fit to the music genres classified by the Billboard and other music magazines devoted to the music industry. The SNA reveals sectors that belong to rap (Kanye West, 50 Cent, Usher), rock (Green Day, Kings Of Leon, The Beatles, U2, Evanescence or Coldplay, Nickelback, Kelly Clarkson), soul and r'n'b (Adele, Britney Spears, Alicia Keys, Beyonce, Michael Jackson), but does not show a sector for pure pop music. Instead of pop, the main community, we have a celebrity sector [Fig. 43] that contains Lady Gaga, Rihanna, Bruce Springsteen, Pink, Jay-Z, The Beatles, Black Eyed Peas and Justin Timberlake.

DATA PROCESSING: Data of albums sales between 2003 and 2013 (IFPI registry) have been cleaned by (the same data set have been analysed in II.5.Music-StochProc):

- different instance of the same artist as well as solo works of band memebers have been aggregated to single category;
- all albums different of the same artist have been aggregated to single weekly sale value;

Suveys of expert view of similarity have been collected by own web questionnaire. Critic polls have been collected via data-mining from book (Buda, 2006). Allmusic.com have been extract by web crawler following/followed by structure.

II.5.Music-StochProc Dynamics of popstar record sales on the phonographic market—the stochastic model

PROBLEM: We investigate weekly record sales of 30 world's most popular artists (2003-2013), as in the paper of the main thesis. We modelled such a time series by discrete mean-reverting geometric jump diffusion (MRGJD) and Markov regime switching mechanism (MRS) between the base and the promotion regimes. The Brownian motion part is responsible for a relatively small (proportional to s multiplied by current value of the process) fluctuations around the long-term mean a/b. The first long-tailed Q is responsible for a promotional effect of the album releasing (log-normal in our case). The second one q is obtained from unitary distribution and is related to first single, which is promoting the whole album. $j(t)$ and $i(t)$ are Heaviside functions of the state (base or exited).

$$dX_t = (a(t) - bX_t)dt + sdW_t + j(t)(dQ_t + dq_t) + i(t)dq_t$$

RESULTS: The provided models reflect the market behaviour for the best-selling artist, assuming their homogeneity [Fig. 44]. A non-stationary mechanistic model replicates the phenomenology of the collective behaviour of the market (without Christmas time), although interaction terms were not introduced directly.

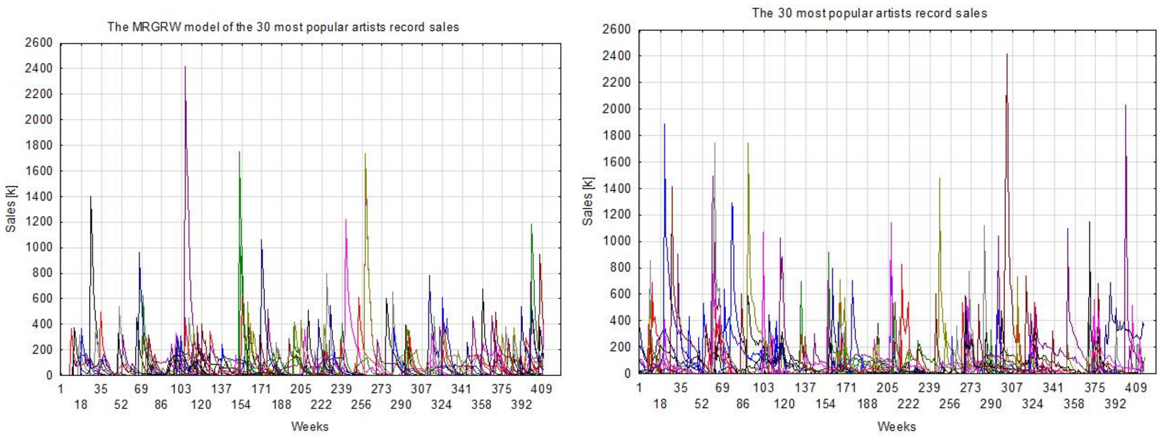


Fig. 44 Visual comparisons of [left] the stationary model, [right] Empirical data

II.6.Music-diffusion Exploring patterns in European singles charts

PROBLEM: European singles charts are an important part of the music industry responsible for creating popularity of songs. After modelling and exploring the dynamics of global album sales in previous papers (II.4.Music-NET, II.5.Music-StochProc), we investigate the patterns of hit singles from weekly charts (polls) in 12 Western European countries. The diversity of gaining popularity in European countries is significant and has become strongly investigated in the Internet era in spite of globalisation.

DATA PROCESSING: We have selected 100 the most popular singles in the period 1966-2015 from more than 17000 available, that have reached the weekly published Top 15 in the national hit singles charts – one per every year by meta-analysis (Buda, 2006; own research) and next 50 – these which reached whole network. Cultural similarity (World Value Survey) was not processed.

Table 4 Correlations between networks of distances between European countries. Music delays in 3 time regimes- Analog, 2- Digital, and 3- Internet, with median and mean delay times. Geographical and Cultural distance stand for reference (Buda, Jarynowski, 2016).

	Analog mean	Analog median	Digital mean	Digital median	Internet mean	Internet median
Geographical	**0.56**	**0.63**	**0.12**	**-0.28**	**0.24**	**0.34**
Cultural	**-0.12**	**-0.32**	**0.26**	**0.50**	**-0.09**	**0.33**

In fashion studies (Kułakowski, et al., 2016), average value of the overall characteristics of a metapopulation (as current trend) depends on external conditions and contingencies, which—according to e.g. Simmel's theory—will determine whether certain features of countries will be activated or not. In addition to the vertical transmission of culture (positive correlation with cultural distance – Table 3) there always exists horizontal transmission (positive correlation with geographical distance – Table 3), which was modeled here [Fig. 45].

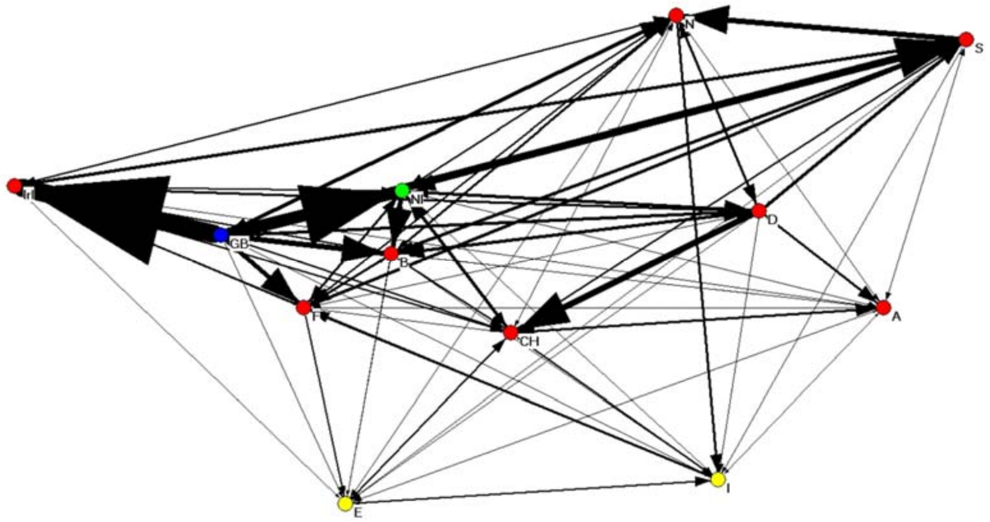


Fig. 45 The popularity spreading network: hub – the GB, bridge – the Netherlands, outliers – Italy and Spain

RESULTS: Spatial and cultural relations in Europe were already known to scientific community, but not from singles popularity perspective and interplay between horizontal transfer (from country to country) and vertical (external field, e.g. influence of USA) was unclear.

SPECULATION: In our model Great Brittan is hub much more in out-degree as in- degree. In case of "Brexit", European Union seems to harm more, because of huge demand of British music.

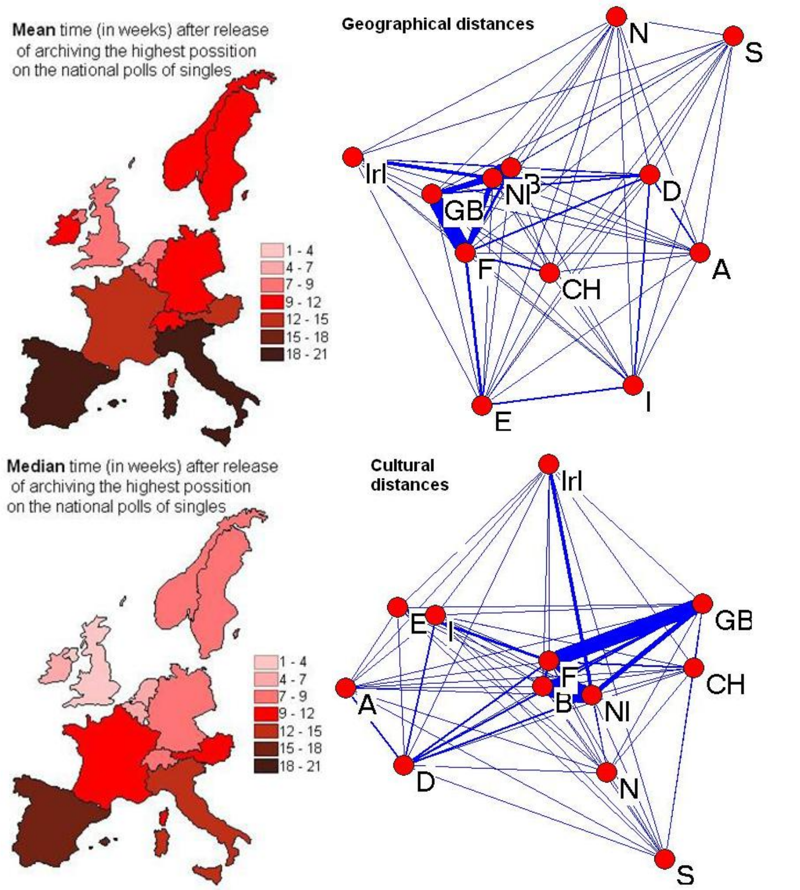


Fig. 46 (left) The map of Western Europe according to the spreading of music popularity for hit singles (1966-2015). (right) The map of Western Europe according to geographic and cultural distance. The links are based on the flows between counties (inverse gravity force)

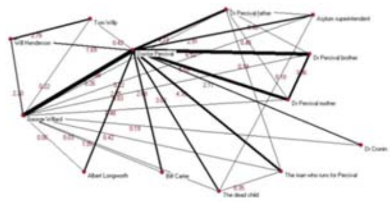
II.7.Fiction Social Networks Analysis in the Discovering the Narrative Structure of Literary Fiction

PROBLEM*:* We used the tools of the network theory to understand and explore the narrative structure in literary fiction (interactions between characters), an approach that is still underestimated. The project aimed at investigating the different ways, in which social interactions are 'read' in texts by comparing the networks produced by automated algorithms-natural language processing (NLP) with those created by surveying more subjective human responses. There are 2 human methods:
- Task 1, the respondents were asked to enter every interaction they recognised with importance in a scale;
- Task 2, the respondent entry lower-matrix, which represent links between characters, also in a scale.
There are also computer algorithms to find the networks structure of narrative, in which one is based on a paragraph and another on a sentence as a unit of calculation.

DATA PROCESSING: We collect links via own survey engine from reader. Task 2 gave some links which do not appear in Task 1, what could be a human error. My idea for correction is thus: If a value is far away from the mean then assume it should be 0 (there is no 'real' link). This value is probably due to human error. We propose two ways of correction, 2 and 3 sigma (well known from statistics).

- Correction 1 (3sigma): Delete link if distance from mean is greater than 3 sigma (standard deviations). This algorithm delete extreme outliers only.
- Correction 2 (2sigma): Delete link if distance from mean is greater than 2 sigma (standarddeviation). This algorithm delete all suspected errors, but also some potential in-directed relations.

Table 5 Correlations between Task 1 and 2 for whole population with final human-based network of Task 2

	Paragraph – based algorithm	Sentence – based algorithm	
Task 1 (entries)	**0.84**	**0.91**	
Task 2 (matrix)	**0.70**	**0.58**	

RESULTS: The different levels of correlation between the respective computer methods and various reader's tasks show the ambivalence and sensitivity to external factors. No universal method is available, which was a paradigm of the NLP. For communication entries (task 1), shorter text window algorithms work better, but for a general perspective (task 2), longer text windows should be applied.

Bibliography

Anderson, PW. (1972). More is different. Science, 177(4047), 393-396.

Axelrod, R, et al. (1981). The evolution of cooperation. Science, 211 (4489), 1390-1396.

Axelrod, R. (1997). Dissemination of culture: a model with local convergence and global polarization. J. of Conflict Resolution 41.

Barabási, AL, et al. (2002). Statistical mechanics of complex networks. Rev. Mod. Phys. 74: 47–97.

Balcan, D., Colizza, V., Gonçalves, B., Hu, H., Ramasco, J. J., & Vespignani, A. (2009). Multiscale mobility networks and the spatial spreading of infectious diseases. Proceedings of the National Academy of Sciences, 106(51), 21484-21489.

Belik V., Hövel P., Mikolajczyk R. (2016). Control of Epidemics on Hospital Networks In Control of Self-Organizing Nonlinear Systems, Schöll E, Klapp SHL, Hövel P, eds. Springer International Publishing, Cham, Switzerland

Bertalanffy, L. (1950). The theory of open systems in physics and biology. Science, 111(2872), 23-29.

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. (1766). "An attempt at a new analysis of the mortality caused by smallpox and of the advantages of inoculation to prevent it". Mem Math Phy Acad Roy Sci Paris , (the English translation by Sally Blower).

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

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

Buda, A. (2006). Historia rocka, popu i hip-hopu według krytyków 1974-2006. Wrocław: WN.

Buda, A., Jarynowski, A. (2016). Exploring patterns in European singles charts, ENIC - conference proceedings, IEEE: DOI 10.1109/ENIC.2015.

Bujok, M., Fronczak, P., & Fronczak, A. (2014). Polish and English wordnets--statistical analysis of interconnected networks. Acta Phys. Pol. B Proc. Suppl. 4, 245-256.

Burt, RS. (2004)"Structural holes and good ideas" American sociological review, 110 (2): 349-399

Camitz, M. (2010). Computer Aided Infectious Disease Epidemiology - Bridging to Public Health. Doctoral thesis, Karolinska Institutet.

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.

De Nooy, W., Mrvar, A., & Batagelj, V. (2011). Exploratory social network analysis with Pajek (Vol. 27). Cambridge University Press.

Czaplicka, A, Sloot, P, et al. (2013). Noise enhances information transfer in hierarchical networks. Nature Sci. Rep. 3, 1223.

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

Dybiec, B., Kleczkowski, A., & Gilligan, C. (2009). Modelling control of epidemics spreading by long-range interactions. Journal of The Royal Society Interface, 6(39), 941-950.

Drzewiński, A. (2000). Coexistence of excited states in confined Ising systems. Physical Review E, 62(3), 4378.

Feld, SL. (1981) "The focused organization of social ties" AJS 86 (5): 1015-1035.

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

Galam, S, Gefen, Y. Shapir J., (1982) Sociophysics: A new approach of sociological collective behaviour, Journal of Math. Sociology 9, 1-13.

Gallos, L.K., Rybski, D., Liljeros, F. Havlin, S. and H.A. Makse. (2011). How people interact in evolving online affiliation networks. Physical Review X, 2(3), 031014.

Grabowski, A., & Kosinski, R. (2004). Epidemic spreading in a hierarchical social network. Physical review. E , 031908.1-031908.7.

Grabowski, A., & Kosinski, R. (2010). Life span in online communities. Phys. Rev. E , 82, 066108.

Grabowski, A., & Rosinska, M. (2012). The relationship between human behavior and the processof epidemic spreading in a real social network. Eur.Phys.J D 83.

Grabowski, A., Jarynowski. A. (2017) Influence of temporal aspects and age-correlations on the process of opinion formation based on Polish contact survey, (submitted) Acta Physica Polonia A.

Granovetter, MS. (1973) "The strength of weak ties" American Journal of Sociology 78 (6): 1360-1380.

Granovetter, MS. (1978). Threshold models of collective behavior. American journal of sociology, 1420-1443.

Granovetter, MS. (1985) "Economic action and social structure – The problem of embeddedness" American Journal of Sociology 91(3): 481-510.

Green, D., & Bossomaier, E. (2000). Complex systems. Cambridge University Press.

Grundmann, H., Hellriegel, B. (2016). Mathematical modelling: a tool for hospital infection control. The Lancet infectious diseases 6(1), 39-45

Hedström, P., Sandell, R. and Stern C. (2000). "Mesolevel Networks and the Diffusion of Social Movements: The Case of the Swedish Social Democratic Party". American Journal of Sociology 196:145-172.

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

Holme, P., Edling, CR., Liljeros, F. (2004). "Structure and time evolution of an Internet dating community." Social Networks 26(2): 155-174.

Holme, P. Saramäki, J. (2012). Temporal networks. Physics Reports, 519(3), 97-125.

Höhle, M. (2007). An R package for the monitoring of infectious diseases. Computational Statistics, 22(4), 571-582.

Jacob, C. (2010). Branching Processes: Their Role in Epidemiology. International Journal of Environmental Research and Public Health , 7, 1186–1204.

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

Jarynowski, A. (2010a). Anomalous interactions in network of Polish Football League. In Life time of correlation. Wrocław: WN.

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

Jarynowski, A. Morris, R. (2011) Oscylacje wapniowe. Rozpoznawanie typu sygnalu, w: BioMedinTech - materiały konferencyjne, Gdańsk.

Jarynowski, A, Klis, M. (2012). Socio-economic models of divorces in different societies, in: Applications of Mathematics In Biology and Medicine – conference proceedings, Krynica Morska, p. 60-64,

Jarynowski, A, Gawroński, P, Kułakowski, K. (2012). How the competitive altruism leads to bistable homogeneous states of cooperation or defection. LNCS, 7204, 543.

Jarynowski, A. (2013). Modelowanie epidemiologiczne przy wykorzystaniu analizy tymczasowych sieci społecznych. W Postępy inżynierii biomedycznej. Rzeszow: Inprona.

Jarynowski, A., Rostami, A. (2013). Reading Stockholm Riots 2013 using Internet media, in: 6th Language and Technology Conference – conference proceedings, Poznań, 353-358

Jarynowski, A, Nyczka, P. (2014). Dynamic network approach to marriage/divorces problem, ENIC, DOI 10.1109/ENIC.2014.24.

Jarynowski, A (2014), Models of great epidemics. Case studies: Wroclaw (smallpox, 1687-1691) and Warsaw (plague, 1624-1626), poster - Advances in Machine Learning for Social Media Analysis (ENIC), Wroclaw.

Jarynowski, A., Lopez-Nunez, F., Fan, H.(2014) How network temporal dynamics shape a mutualistic system with invasive species? arXiv preprint arXiv:1407.4334, report, Umea University.

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

Jarynowski, A., Grabowski, A. (2015). Modelowanie epidemiologiczne dedykowane Polsce, Portal CZM 9(6).

Jarynowski, A., Buda, A. (2015). Hierarchical representation of socio-economic complex systems according to minimal spanning trees and diagrams. PhD Interdysciplinary Journal, (1). 37-45.

Jarynowski, A., Jankowski, J., & Zbieg, A. (2015). Stop Acta Natural vs. artificial viral spread within online community, E-methodology, 2, 71.

(Jarynowski, A) Ярыновский, А,. (2015a). КАКАЯ ФУНКЦИЯ АКТИВАЦИИ СОТРУДНЕЧЕСТВА ОПИСЫВАЕТ ЧЕЛОВЕЧЕСКОЕ ПОВЕДЕНИЕ?, Современные проблемы математического моделирования и вычислительных методов, СЕКЦІЯ 3-4. 1, Ривнэ.

Jarynowski, A. (2015b). Optimal cervical cancer preventing strategies – model for Moldova, Scientific Dialogue, No 1, 169.

(Jarynowski, A) Ярыновский, А. (2015c). Распространение вируса в интернет-сообществе - модели эпидемического. наука, образование, культура. КГУ 7.1

Jarynowski, A. (2015d). Collapse of cooperation and corruption in a mathematical model within game theory including Moldovan case study (Homo Sociologicus vs. Homo Economicus), МЕЖДУНАРОДНАЯ НАУЧНАЯ КОНФЕРЕНЦИЯ, посвященная 20-летию экономического образования в Бельцком Университете «Алеку Руссо».

Jarynowski, A., Nyczka, P. (2016). Single parameter model of marriage/divorces dynamics with countries classification (in review) Physica A.

Jarynowski A., Marchewka D., Grabowski A. (2016). Computer-assisted risk assessment of hospital infections: a preliminary implementation in Polish hospitals, Journal of Hospital Infection, 94 (S1), 128.

Jarynowski, A., Buda, A., & Piasecki, M. (2016,). Multilayer Network Analysis of Polish Parliament 4 Years before and after Smolenńk Crash. In Network Intelligence Conference (ENIC), 69-76.

Jarynowski A., Marchewka D. (2017). Szacowanie ryzyka zakażenia patogenami szpitalnymi – macierzyństwo 2.0 (wersja beta), submitted, monografia Wydawnictwa Warszawskiego Uniwersytetu Medycznego

Jarynowski A., Marchewka D., Buda A. (2017). Internet - assisted risk assessment of infectious diseases in women sexual and reproductive health, sumbitted, E-methodology.

Jarynowski A., Różańska, A., Wójkowska-Mach, J., Rosiński, J. (2017). Polska adaptacja kwestionariusza postrzegania Zasad Izolacji Standardowej – analiza pól znaczeniowych, work in progres.

Jędrychowski, W. (2010). Epidemiologia w medycynie klinicznej i zdrowiu publicznym. Wydawnictwo Uniwersytetu Jagiellońskiego.

Kleczkowski, A., Oleś, K., Gudowska-Nowak, E., & Gilligan, C. A. (2011). Searching for the most cost-effective strategy for controlling epidemics spreading on regular and small-world networks. Journal of The Royal Society Interface, 10216.

Kolk, M., Cownden, D., & Enquist, M. (2014). Correlations in fertility across generations: can low fertility persist?. Proceedings of the Royal Society of London B: Biological Sciences, 281(1779), 2561.

Konarski, R. (2014). Modele równań strukturalnych: teoria i praktyka. Wydawnictwo Naukowe PWN.

Kulig, A, Kwapień, J, et al. (2012). Complex network analysis of literary and scientific texts. International Journal of Modern Physics C, 23, 1250051.

Kułakowski, K., Kulczycki, P., Misztal, K., Dydejczyk, A., Gronek, P., & Krawczyk, M. J. (2016). Naming Boys after US Presidents in 20th Century. Acta Physica Polonica, A., 129(5).

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

Leskovec, J., Adamic, L., & Huberman, B. (2007). The dynamics of viral marketing. ACM Trans. Web , May.

Liljeros, F., Giesecke, J., & Holme, P. (2007). The contact network of inpatients in a regional healthcare system. A longitudinal case study. Mathematical Population Studies, 14(4), 269-284.

Liljeros, F., et al. (2001). The web of human sexual contacts. Nature , 411 (6840), 907-908.

Lu, X., Wetter, E., Bharti, N., Tatem, A. J., & Bengtsson, L. (2013). Approaching the limit of predictability in human mobility. Scientific reports, 3.

Lazer, D., Pentland, A. S., Adamic, L., Aral, S., Barabasi, A. L., Brewer, D., ... & Van Alstyne, M. (2009). Life in the network: the coming age of computational social science. Science, 323(5915), 721.

Makowiec, D., Gała, R., Dudkowska, A., Rynkiewicz, A., & Zwierz, M. (2006). Long-range dependencies in heart rate signals—revisited. Physica A: Statistical Mechanics and its Applications, 369(2), 632-644.

Mantegna, R. N. (1999). Hierarchical structure in financial markets. The European Physical Journal B-Condensed Matter and Complex Systems, 11(1), 193-197.

Merton, R. (1968). The Matthew effect in science. Science,159 (3810), 56-63.

Mezard, M., Parisi, M., & Virasoro, A. (1987). Spin glass theory and beyond. Singapore: World Scientific.

Milgram, S. (1967) " Small-World Problem" Psychology Today, 1 (1), 61-67.

Morieri, G., Martinez, E., Jarynowski, A., Driguez, H., Morris, R., Oldroyd, G., Downie, J. (2013) Host-specific Nod-factors associated with Medicago truncatula nodule infection differentially induce calcium influx and calcium spiking in root hairs, New Phytologist, 200(3), 656-662.

Murray, J. (2002). Mathematical Biology. I. An Introduction (chapter 10: Dynamics if Infectious Diseases: Epidemic Models and AIDS),. Springer (Polish translation: Urszuli Foryś).

Obadia, T., Silhol, R., Opatowski, L., Temime, L., Legrand, J., Thiébaut, A. C., ... & I-Bird Study Group. (2015). Detailed contact data and the dissemination of Staphylococcus aureus in hospitals. PLoS Comput Biol, 11(3), 1004170.

Ohst, J., Liljeros, F., Stenhem, M., & Holme, P. (2014). The network positions of methicillin resistant Staphylococcus aureus affected units in a regional healthcare system. EPJ Data Science, 3(1), 1.

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

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

Pękalski, A., & Cebrat, S. (1999). Model of population evolution with and without eugenics. The European Physical Journal B.

Poczobut, R. (2006) System-struktura-emergencja, [w:] M. Heller, J. Mączka (red.), Struktura i emergencja, Kraków-Tarnów: PAU-OBI, 11-38.

Popova, I., Popov, O., & Brouwers, L. (2010). Modeling and Simulation Training for Undergraduate Students. In: Eurosim 2010: proceedings of the 7th EUROSIM Congress on Modelling and Simulation, Prague

Robins, G., Pattison, P., Kalish, Y., & Lusher, D. (2007). An introduction to exponential random graph (p*) models for social networks. Social networks, 29(2), 173-191.

Rocha, LEC, Liljeros, F, et al. (2010). Information dynamics shape the sexual networks of Internet-mediated prostitution. PNAS, 107 (13), 5706-5711.

Rogers, E. (2005). Diffusion of Innovations. New York: Free Press.

Różanska, A., Wójkowska-Mach, J., Jarynowski A., et. al. (2017) Is postpartum SSI surveillance possible without postdischarge surveillance? Results from multicentre surveillance in Polish hospitals, accepted, American Journal of Infection Control.

Sarnecki, J. (2001). Delinquent networks: Youth co-offending in Stockholm. Cambridge University Press.

Schelling, T. (1971). Dynamic models of segregation. Journal of mathematical sociology, 1(2), 143-186.

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

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

Sznajd-Weron, K., & Sznajd, J. (2000). Opinion evolution in closed community. International Journal of Modern Physics C, 11(06), 1157-1165.

Sztompka, P. (1973). O osobliwościach nauk społecznych raz jeszcze. Studia filozoficzne, 8 (105).

Valdano, E., Poletto, C., Giovannini, A., Palma, D., Savini, L., & Colizza, V. (2015). Predicting epidemic risk from past temporal contact data. PLoS Comput Biol, 11(3), 1004152.

Vensim 5 Modeling Guide. (2003). Ventana Systems, Inc.

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.

Wilensky, U. (1999). NetLogo.

Errata to thesis papers

In paper I.2.MRSA Fig.3 should be replaced by presented below Fig. 47 Right

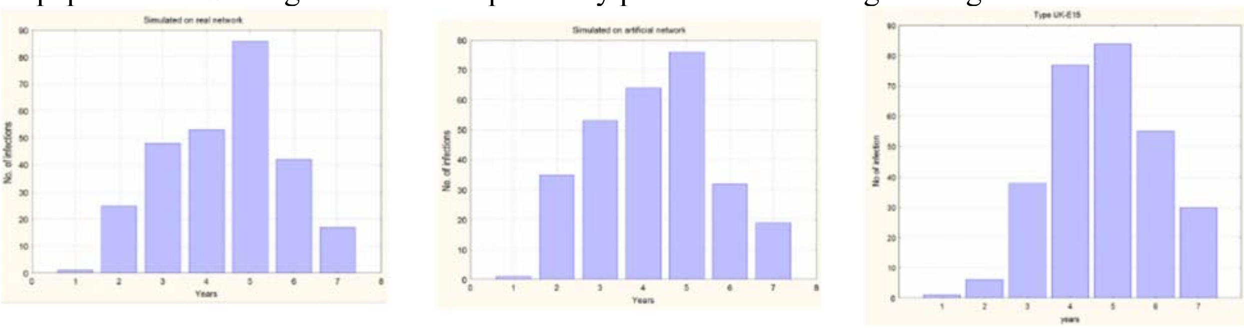


Fig. 47 Correct graphs from paper I.2.MRSA. Simulated outbreak – real network [Left]. Simulated outbreak – artificial network [Central]. Real outbreak [Right]

In paper II.6.Music-diffusion Equation (1) shoud be replace by: $d_{ij}=|t_x-t_y|$

List of tables

List of Figures

Thesis papers

I Epidemiological modeling

8) A Jarynowski, A Serafimovic, Studying possible outcomes in model of sexual transmitted virus (HPV) causing cervical cancer for Poland, *Advances in Intelligent Systems and Computing*, Volume 229, pp 129-141, 2014 (I.1.HPV-Pl) [Indexed by Web of Science Core Edition, 15 points]
9) A Jarynowski, F Liljeros, Contact networks and the spread of MRSA in Stockholm hospitals, ENIC - conference proceedings, IEEE: DOI 10.1109/ENIC.2015.30. 2015 (I.2.MRSA)) [Indexed by Web of Science Core Edition, 15 points]
10) A Grabowski, A Jarynowski, Rumor propagation in temporal contact network from Polish polls, ENIC - conference proceedings, IEEE DOI 10.1109/ENIC.2016.19, 2016 (I.3.Polymod)) [Indexed by Web of Science Core Edition, 15 points]

II Networks and data modeling

11) A Buda, A Jarynowski, Network structure of phonographic market with characteristic similarities between artists, *Acta Physica Polonica A*, vol. 123, no. 3, 2013 (II.4.Music-NET) [List A, 15 points]
12) A Jarynowski, A Buda, Dynamics of popstar record sales on phonographic market--stochastic model, *Acta Physica Polonica B* (PS) 2 (7). 2014 (II.5.Music-StochProc) [List B, 9 points]
13) A Jarynowski, A Buda. Diffusion paths between product life-cycles in the European phonographic market. *Control & Cybernetics* 45.2. 2016 (II.6.Music-diffusion) [List B, 14 points]
14) A Jarynowski, S Boland, Rola analizy sieci społecznych w odkrywaniu narracyjnej struktury fikcji literackiej, *Biuletyn Instytutu Systemów Informatycznych* 12, 35-42, 2013 (II.7.Fiction) [List B, 2 points]