2017 COMPLEX SYSTEMS SUMMER SCHOOL

AT THE SANTA FE INSTITUTE

SANTA FE INSTITUTE   *Searching for Order in the Complexity of Evolving Worlds*   **www.santafe.edu**

2017 COMPLEX SYSTEMS SUMMER SCHOOL
GROUP PROJECT ABSTRACTS

Agent-Based Prediction Competition

Marco Pangallo[1]**, Katarina Mayer**[2]**, Yuji Saikai**[3]**, Chris Miles**[4]**, Yael Gurevich**[5]**, Cigdem Yalcin**[6]**, Uzay Çetin**[7]**, and Stefan Bucher**[8]

[1]Institute for New Economic Thinking at the Oxford Martin School and Mathematical Institute, University of Oxford, Oxford, UK
[2]Independent researcher
[3]Department of Agricultural & Applied Economics, University of Wisconsin, Madison, 53706, USA
[4]Center for the Study of Complex Systems, University of Michigan, Ann Arbor, 48109, USA
[5]Department of Molecular Biology and Ecology of Plants, Tel-Aviv University, Tel-Aviv, Israel
[6]Department of Physics, Istanbul University, 34134, Vezneciler, İstanbul, Turkey
[7]Istanbul Bilgi University, Computer Engineering Department, Istanbul, 34060, Turkey
[8]New York University, Department of Economics, New York, NY 10012, United States

ABSTRACT

Agent-Based Models (ABMs) have so far been used to describe complex systems. A qualitative agreement with stylized facts in real-world data has often been considered enough to validate an ABM. However, given their closer match with reality, there is growing awareness that ABMs should be able to outperform simpler models in quantitative prediction. Unfortunately, predicting with ABMs is an extreme conceptual and technical challenge. Recent research suggests generating synthetic data by ABMs of varying complexity and attempting to predict this data.

We independently propose an Agent-Based Prediction Competition. We generate synthetic data from a financial market ABM. We then try to predict the stock prices out-of-sample, by employing other ABMs, a dynamical system and statistical models. Our goal is to check whether prediction is at all possible, and whether more complex models (such as the other ABMs) outperform simpler or statistical models.

We show that in this case, a single realization of the ABM cannot be reliably predicted at all with any of the tested methods. A substantial level of noise casts a strong limit on the predictive power of the models. Considering the average realization of the out-of-sample stock prices, complex ABMs outperform simpler ABMs, dynamical systems and statistical methods. However, this result is not statistically significant due to the noise level. We conclude that for future research, it is necessary to design an ABM in which realistic assumptions increase the signal to noise ratio.
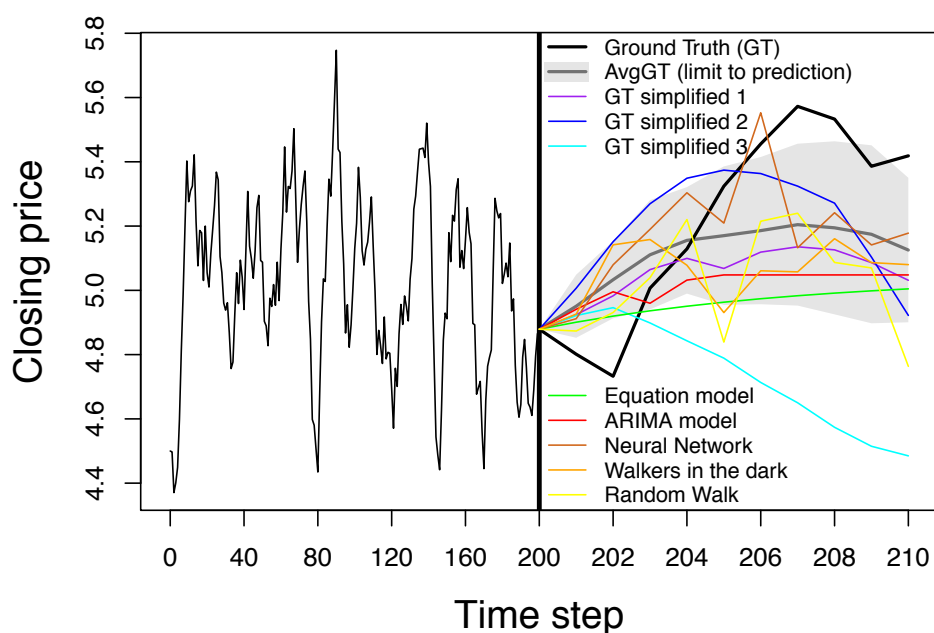
Figure 1. Forecasting performance of competing models. For visualization purposes the scale on the *x*-axis is different on the left and right panels.

Analyzing the Dynamics of Agent-Based Models with Delayed Coordinate Embedding

Stefan Bucher[1,+]**, Jose Alejandro Coronado**[2,+]**, Meghan Galiardi**[3,*+]**, and Markus Junginger**[4,+]

[1]New York University, Department of Economics, New York, NY 10012, United States
[2]The New School, New York, ¡postcode¿, United States
[3]Sandia National Laboratories, Albuquerque, NM 87185, United States
[4]Porsche, Germany
[*]mgaliar@sandia.gov
[+]The authors contributed equally to this work.

ABSTRACT

Complex systems models are frequently criticized for being opaque, because it is not always clear which assumption drives a result and it would be of tremendous help to understand and quantify its dynamics analytically. But since most complex systems are complex enough that they can only be simulated, going beyond reporting simulated time series is often times difficult.

DCE is a technique used to analyze non-linear time series data. DCE is used on a single observation from a non-linear time series and reconstructs the dynamics of the underlying process. We propose a different paradigm for applying DCE: analyzing the parameters and results of complex systems models. In this paradigm, we assume we have a model of an arbitrary system. There are often cases where data for this system is either unknown or unavailable. Therefore, modelers cannot calibrate the model to determine the appropriate model parameters or validate the results of the model against real-world data. Without this data, modelers have a set of tools they use to analyze the model. The most common tools in this tool set are sensitivity analysis and uncertainty quantification. Sensitivity analysis determines how uncertainty in the results is related to the input parameters. Sensitivity analysis usually performs a parameter sweep and analyzes how changes to the parameters affect the results. This allows the modeler to quantify acceptable ranges of parameters which produce acceptable results. Uncertainty quantification uses a variety of techniques and investigates the reliability of model results as well as particular model assumptions. We propose DCE as an additional tool. We view DCE as a similar tool to sensitivity analysis, but will give much more insightful results. By performing parameter sweeps and applying DCE to the results (time series) of the model, we can determine the underlying dynamics of the model. This will determine more than just the sensitivity of the results on the parameter, but will determine qualitative behavior of the results (ex - cycles or chaos). By sweeping all the parameters, we could in theory create a bifurcation diagram for our model.

Figure 1. DCE reconstruction

We begin by asking if the dynamics of an agent-based model (ABM) can be studied by performing DCE on the time series it produces. As a proof of concept, we study two ABMs for which the dynamics are known: a predator-prey model and an asset pricing model. We show that the dynamics reconstructed by DCE coincide with the true underlying dynamics for various parameter regimes and propose further work to extend these methods.
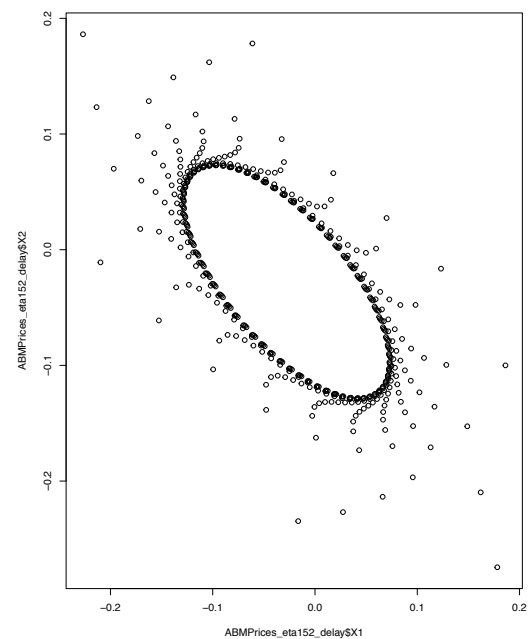
Behavioral footprints of depression from mobile-phone usage patterns

Ulf Aslak,[1] Hugo Barbosa,[2] Freya Casier,[3] Madison Hart,[4] Aaron Schwartz,[5] and Zhiya Zuo[6]

[1] University of Copenhagen, Copenhagen, Denmark
[2] Department of Physics and Astronomy, University of Rochester, Rochester, NY, USA
[3] The Danish National Centre of Applied Social Science, Copenhagen, Denmark
[4] University of Texas at Austin, TX, USA
[5] University of Vermont, USA
[6] University of Iowa, USA

Depression is increasingly prevalent in society. Alone, it affects over 16.1 million American adults and is today the single most pervasive mental illness in the world. Many efforts trying to address this problem leverage technological advancements such as smartphones and smart tracking devices, to facilitate patient self-reporting for use in analyses that can aid medical experts. However, patients frequently "drop out" or do not self-report, likely a behavior caused by their perpetual depression. This begs the invention of reporting mechanisms that do not rely on active participation from the patient. In this paper, we investigate the relationships between behavioral indicators, measured from smartphone activity, and depression in a large population of highly interconnected young adults. We find that depression correlates strongly with features that indicate social activity, communicational responsiveness and academic attendance. Using this insight we construct a simple classification model that predicts, from their set of behavioral indicators, whether an individual is more or less depressed than the population median. Our results have implications to the future of depression treatment. For example, using a classification model based on the features we highlight here, alerting systems could be built to help counteract the onset of depression in individuals.
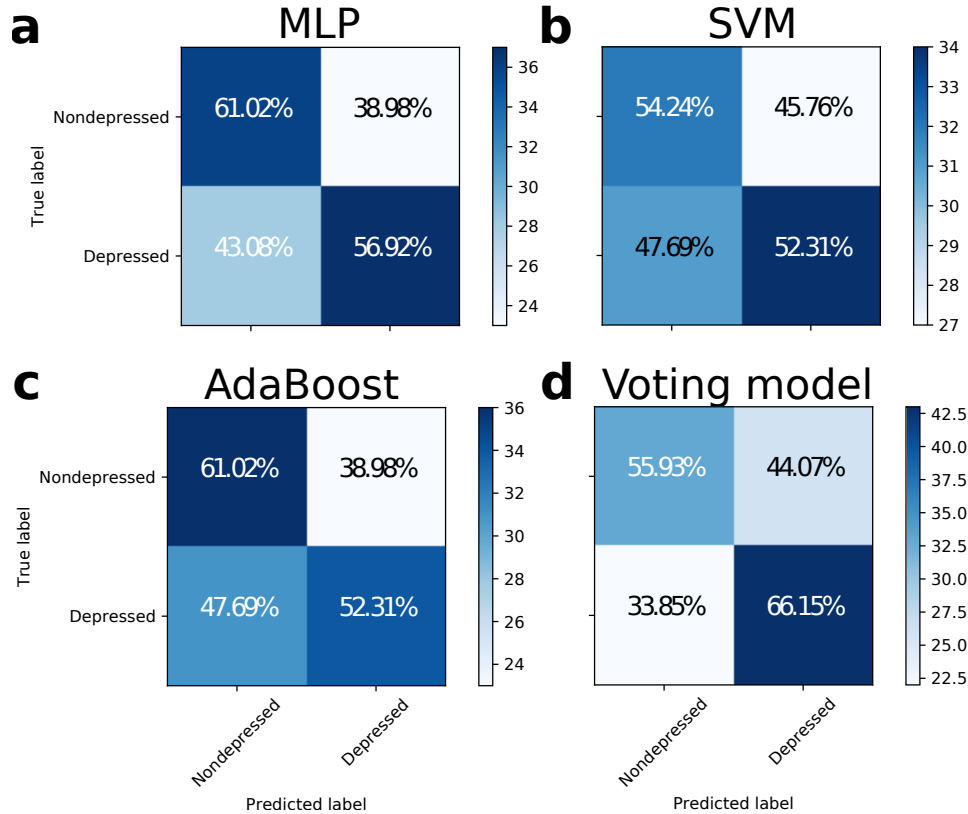
FIG. 1: **Model performances - two weeks of averaged data before and after each survey prompt.** (a) The MLP once again exhibited the best performance of all three individual algorithms. However, there was a subtle performance improvement in the SVM (b) when trained with the aggregated data in comparison with the previous experiment. A similar performance gain was also observed in the AdaBoost algorithm (c), adding more evidence that the aggregated data can better capture behavioral signs of depression, which is corroborated by the performance gain also observed in the voting model (d).

Can bacteria play hide and seek? The unseen spread of carbapenem resistance

Alicia N. M. Kraay[1], **Jiří C. Moravec**[2], **Ximo Pechuan**[3], **Jake L. Weisman**[4], **Hilje M. Doekes**[5], **and Makoto Jones**[6,7,*]

[1]Department of Epidemiology, University of Michigan, SPHII M5326, 1415 Washington Heights, Ann Arbor, Michigan, USA 48109
[2]Statistics and Bioinformatics Group, Institute of Fundamental Sciences, Massey University, Private Bag 11 222, Palmerston North, New Zealand
[3]Department of Systems and Computational Biology, Albert Einstein College of Medicine, Bronx, USA
[4]Department of Biology, University of Maryland College Park, MD, USA
[5]Theoretical Biology and Bioinformatics, Utrecht University, the Netherlands
[6]Veterans Affairs Salt Lake City Health Care System, Salt Lake City, Utah, USA
[7]University of Utah School of Medicine, Salt Lake City, Utah, USA
[*]makoto.jones@hsc.utah.edu

ABSTRACT

Motivation: Carbapenems are the antibiotics of choice to treat infections caused by multidrug-resistant bacteria because they are generally safe and effective. However, the recent and often silent spread of plasmid-borne carbapenemases between people and bacterial genera has increased the prevalence of carbapenem resistance and is forcing use of more toxic antibiotics.
Problem: Inaccurate resistance tests with low sensitivity increases the probability of false-negative results, which may contribute to spread of carbapenem resistance within hospitals. We model how these false-negative test results could fail to prompt appropriate clinical response and investigate possible mechanisms of these erroneous results.
Approach: To account for complex interactions between hospital transmission, antibiotic treatment, laboratory methods, and plasmid replication, we constructed a set of models spanning multiple scales. Hospital transmission was modeled deterministically using a modified SIR framework incorporating antibiotic-resistant and susceptible infection classes and alternative screening protocols and plasmid transfer regimes. The probability of false-negative tests for carbapenemases were modeled using the biology of *bla*KPC for context and a combination of deterministic and stochastic models.
Results: Within feasible parameter ranges, the false-negative rate for testing contributes more to the spread of carbapenem resistance than increases in infectivity due to the selective pressure in favor of plasmid-carrying bacteria that occurs when a test is falsely negative and carbapenems are administered. False negative rates as high as 20% may be possible with fitness costs less than 5% given the small count sampling inherent in modern clinical microbiology practices.
Conclusion: Current US Federal Drug Administration guidelines recommend the use of selective carbapenem pressure after *in vitro* isolation of suspect *Enterobacteriaceae* to prevent plasmid loss prior to molecular testing for carbapenemases. Our results suggest that applying selective pressure throughout the culture isolation process could help prevent *in vitro* laboratory findings that are discordant with *in vivo* antimicrobial susceptibility. The potential role of false-negative tests on transmission merits further investigation. Accounting for multiscale-complexity may inform approaches in other scenarios where horizontal gene transfer is important.
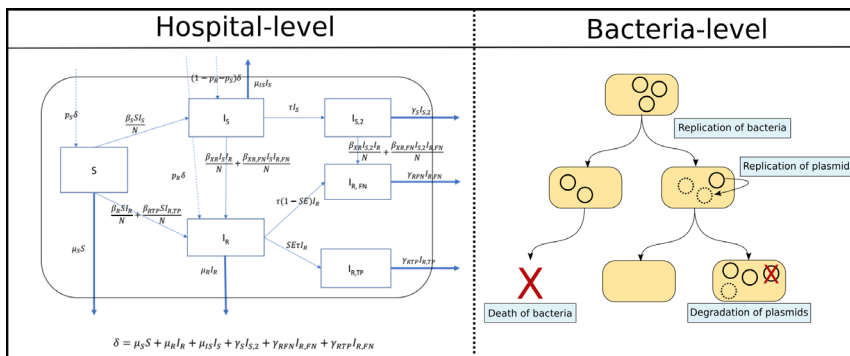
Figure 1. Schematics of the modeling approaches. At hospital level, dynamics of individuals that may be be susceptible, infected and sensitive to carbapenems (I_S, $I_{S,2}$), or infected and resistant to carbapenems (I_R, $I_{R,FN}$, $I_{R,TP}$) are modeled. At the bacterial level, an agent-based model describes stochastic events in a growing population of cells.

Classification of dragon-king phases in preferential attachment and failure models

Sandro Claudio Lera[a,b], Ulf Aslak Jensen[c]

[a]*ETH Zurich, Singapore-ETH Centre, 1 CREATE Way, #06-01 CREATE Tower, 138602 Singapore*
[b]*ETH Zurich, Department of Management, Technology, and Economics, Scheuchzerstrasse 7, 8092 Zurich, Switzerland*
[c]*Copenhagen Centre for Social Data Science*

Abstract

A new case of a preferential attachment model is considered. The probability that an already existing node in a network acquires a link to a new node is proportional to the product of its intrinsic fitness and its degree. We enrich this already known model by preferential deletion, which removes nodes at random with probability proportional to their fitness to some exponent ω. Under 'normal' conditions, the resulting node degree distribution is an asymptotic power-law (scale-free regime). We derive an exact condition for a phase transition after which one or a few nodes capture a finite fraction of all links in the infinite networks (dragon-king regime). By approximately 'parametrizing' the space of fitness distributions through the beta-density, we then show phase-diagrams that separate the two regimes.
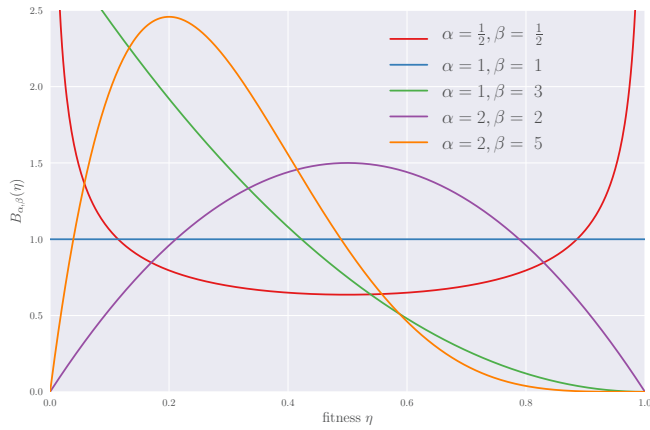
Figure 1: Beta distribution for different set of parameters α and β. We assume that this parametric class of distribution approximates the space of fitness distributions sufficiently well for all practical cases. Dragon-king nodes emerge when the bulk of the probability mass is concentrated on nodes of low fitness, and there are just a few fit nodes, that then end up dominating the system (e.g. $\alpha = 2, \beta = 5$).
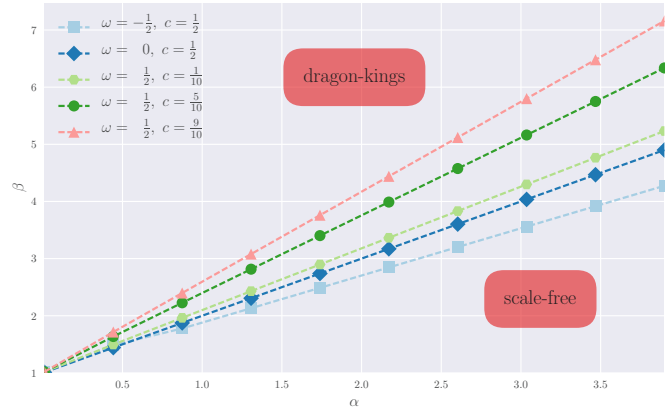
Figure 2: Phase-diagram of the scale-free vs. dragon-king phases in the (α, β)-parameter plane of the beta-distribution, for different preferential deletion parameters ω and deletion rates c. The scale-free regime is the one below the separation line(s), corresponding to fitness distributions that tend to be either balanced or overemphasize the fit nodes. In contrast, if the distribution is such that most probability mass is concentrated on nodes with low fitness, dragon-king nodes can emerge. This effect is enhanced the smaller ω and the larger c, i.e. the more nodes of low fitness are removed. In contrast, if ω is large, the scale-free regime is widened, as the targeted attacks to fit nodes tend to remove dragon-kings. The larger the rate of removal c, the stronger this effect.

Email addresses: `slera@ethz.ch` (Sandro Claudio Lera),
`ulfaslak@gmail.com` (Ulf Aslak Jensen)

Forecasting Seasonal Influenza in the United States Using Nonlinear Time Series Analysis

Alicia Kraay,[1]* Rachel E. Gicquelais,[1]*, Ramona Roller,[2] Kyle Lemoi,[3] Elaine M. Bochniewicz[3]
Spencer J. Fox[4]

*Authors contributed equally}

Author Affiliations
[1]University of Michigan
[2]University of Amsterdam
[3]MITRE Corporation
[4]University of Texas at Austin

Forecasting of seasonal infectious diseases, such as influenza, can help in public health planning and outbreak response. We compared a traditional influenza forecasting method (an autoregressive moving average [ARMA] model) with a nearest neighbor forecasting approach (the Lorenz Method of Analogues), where nearest neighbors were identified from a reconstructed state space using delay coordinate embedding. Delay reconstruction and ARMA models used influenza-like-illness (ILI) surveillance data from 1997-2017 in the United States. We compared model forecasts of the 2015-2017 influenza seasons across 1-4 week prediction horizons. Each model's fit to ILI data (black line) is summarized in the figure below across the four prediction horizons. Overall, ARMA models (teal), more accurately predicted ILI than the method of analogues (salmon), especially when the prediction horizon was 1-2 weeks. The method of analogues with a single nearest neighbor predicted slightly better than the ARMA model for 3-4 weeks; however, a very large delay vector was required. Future directions will include exploring other methods to incorporate nearest neighbors into the method of analogues and combining both autoregressive and nearest neighbor approaches.
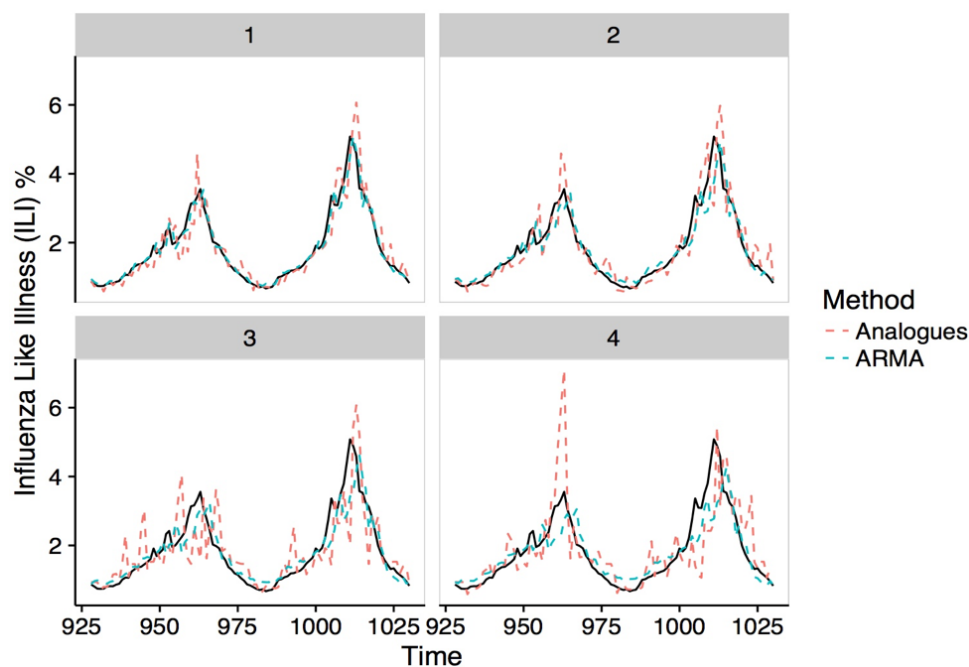
Hops across cultural boundaries: a regional analysis of beer recipe ingredients and their pairing principles.

V. Bleu Knight[1], Doheum Park[2], and Zhiya Zuo[3t]

[1] New Mexico State University, Department of Biology, Las Cruces, 88003, U.S.A.
[2] Korea Advanced Institute of Science and Technology, Daejeon, South Korea
[3] University of Iowa, Iowa City, IA, 52242
[t] Alphabetical order

Abstract

Brewing beer has come to be regarded as a form of art, yet the consumption of beer is a cultural-specific phenomenon that evolves regionally. In the United States, for example, commercially available beers have drastically increased in number and style over the course of the last 30 years. Here we quantify diversity in beer brewing within and across different regions of the globe using a data-driven approach. The Jaccard index is used to measure actual and functional diversity of regional beer recipe ingredients and the flavor compounds they contain. We identified highly prevalent ingredients, and found them to reflect many common vintage beer recipe ingredients. We also identified regionally authentic ingredients, and found that the cultural traditions of the region are represented within those ingredients. Overall, hops offered the most regional authenticity to the recipes. Aroma preferences for hops from Europe and North America are outlined, along with regional tendencies to share flavor molecules in beer recipe ingredients.

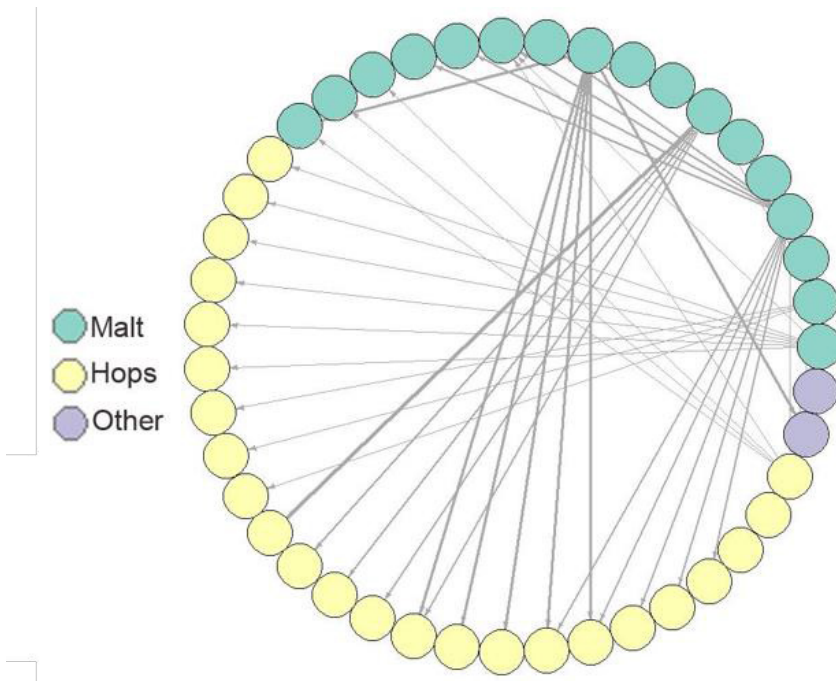**Figure 4:** Authentic Ingredient Network. The top 10 ingredients with the highest authenticity scores for each region were analyzed in a network format, revealing that the majority of authentic ingredients are hops, and that hop ingredients comprise fewer hubs than their malt ingredient counterparts. Line weight corresponds to region. Africa is the heaviest, followed by Oceania, Asia, North America, South America, and Europe.

Inferring Functional Taxonomy from Network Structure

Alje van Dam[1], Kayla R. Sale-Hale[2,3]*, Surendra Hazarie[4], and Carla N. Rivera[5]

[1]Copernicus Institute, Utrecht University, Utrecht, The Netherlands
[2]University of Arizona, Ecology & Evolutionary Biology, Tucson, 85721, United States
[3]University of Michigan, Ecology & Evolutionary Biology, Ann Arbor, 48109, United States
[4]University of Rochester, Physics, Rochester, United States
[5]Center of Applied Ecology and Sustainability (CAPES), Santiago, PC 6513677, Chile
*Correspondence to: kaylasale@icloud.com

ABSTRACT

The projection of bipartite graphs has proven to be a useful tool to infer similarity measures between nodes of either one of the node types, based on how many links they share with nodes of another type. In economics this has lead to the concept of the product space (Hidalgo *et al.* 2007), which is a network that connects products based on their co-locations in different countries. The relatedness between products in this networks is thought to be representative of the 'capabilities' that are required to produce such products. Here we apply similar concepts to plant-pollinator networks and evaluate the results in light of natural history. By projecting plant-pollinator networks onto plant-plant and pollinator-pollinator networks, we aim to derive a measure of relatedness between plants based on the ecological traits that a pollinator must have to feed on it, and likewise a measure of relatedness between pollinators in terms of the traits they share, that allow them to feed on the same plants. By running community detection algorithms on the projected networks we create a functional taxonomy based on these (unobserved) ecological traits, and compare this taxonomy to taxonomies based on observable traits. As opposed to evolutionary taxonomies, our inferred taxonomies can represent a 'partner space,' describing how plants see their pollinator community in terms of their capabilities and vice versa. Trait-based community detection formalizes the concept of 'pollination syndromes,' the idea that pollinators in a pollination network with similar traits prefer similar plants with similar traits (e.g. flies pollinators prefer white-flowered plants) (Willmer 2011). We find that detected communities (Figure 1) are well-aligned with pollination guilds formed from field observations of flower visitors but that most basic (empirically-recorded) plant traits are not predictive of the communities observed from the projected plant-pollinator networks, suggesting that other traits play a significant role in plant choice for pollinators. This indicates the non-trivial nature of the communities identified by looking at plant-pollinator networks, reinforces criticisms of pollination syndromes, and may inspire future ecological research such as the identification of characteristic ecological traits that form these communities.
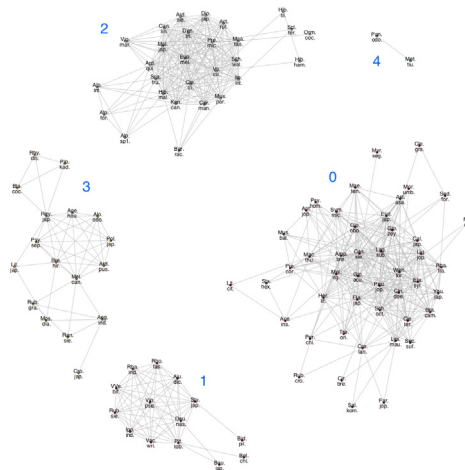
Figure 1. Plant similarity network.

It from Bit: Cellular Automata, Driven Ising Model, and Emergent Computation

Chris Miles[1], Elliot Nelson[2], Kyle Reing[3], Shing H. Zhan[4], and Mark Kirstein[5]

[1]Department of Physics, University of Michigan, Ann Arbor, MI, U.S.A.
[2]Perimeter Institute, Waterloo, ON, Canada
[3]Department of Computer Science, University of Southern California, Los Angeles, CA, U.S.A.
[4]Department of Zoology, University of British Columbia, Vancouver, BC, Canada
[5]Department of Economics, Technische Universität Dresden, Dresden, Germany

Abstract

We explored emergent computation within two systems — cellular automata (CA) and a driven Ising model. We attempted to generate emergent computation with a CA using a genetic algorithm, with the aim of understanding general conditions on information processing dynamics which increase computational capacity. Specifically, we aim to search the space of CAs to identify minimal computational units, which may in turn be used to execute specific computational tasks or optimize further search for collective computational behavior. In order to select for CAs with more computational capacity, we use an objective function defined in terms of dual total correlation (DTC), an information-theoretic quantity capturing the information in higher-order correlations between cells in a given spatial (and temporal) region of the time-dependent CA grid. The genetic algorithm evolves a population of CA update rules by testing average performance over a range of initial conditions, creating a new population favoring more successful rules, and iterating over a number of generations.

In addition, we have explored stochastic thermodynamics within a driven Ising model. This initial investigation has given us insight into thermodynamic relations present that will be beneficial for exploring this system's computational capacity in future work. In this preliminary study, the work distribution produced by many realization under the influence of a time-symmetric drive is consistent with known fluctuation relations in the field of stochastic thermodynamics. In future work, we hope to explore dissipative adaptation as a potential mechanism for promoting emergent computation.
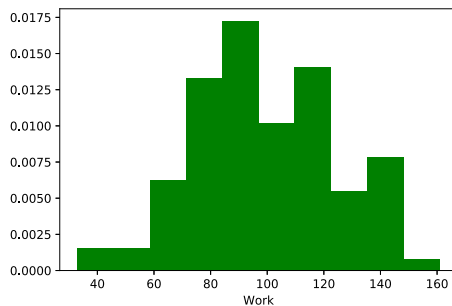
Figure 1: A histogram of the thermodynamic work of many realizations produced by a time-symmetric magnetic field drive on the Ising model.

Modeling and simulating the emergence of Internet communities: impact of the spread of memes and agent memory

Aida Huerta-Barrientos[1], Rémi Lamarque[2], Stephen Leese[3]

1 National University of Mexico, e-mail: aida.huerta@comunidad.unam.mx

2 Aix Marseille Univ, CNRS, LPL, Aix-en-Provence, France, e-mail: remi.lamarque@lpl-aix.fr

3 Davenport, Iowa, e-mail: whatcroft@gmail.com

Abstract

The spreading of memes on the web played an important role in the emergence of Internet communities. The principal purpose of the study is to implement a simulation model to analyze the process of emergence of Internet communities. The model shows the importance of factors such as the interactions between online agents as well as their propensity to adopt and remember new memes. In addition, it explores the threshold between isolated cultural short-lived trends and the viral spreading of many cultural features. The simulation model is an agent-based model built using NetLogo™ software, designed such that agents represent Internet users and memes are represented by features appearing randomly on each agent. The model illustrates the spread across the whole network as shown through interactions of agents indicating further, that the structure of the network, especially the number of indegree and outdegree links between agents, has a crucial influence on how many memes are shared among agents in the long run. In other words, a greater connectivity leads to the quick sharing and sustainability of several cultural features, which is the basis for the emergence of a community.

Figure 1. The simulation results of memes spread through a of 100 nodes and a memes rate equal to 2, after 1000 ticks
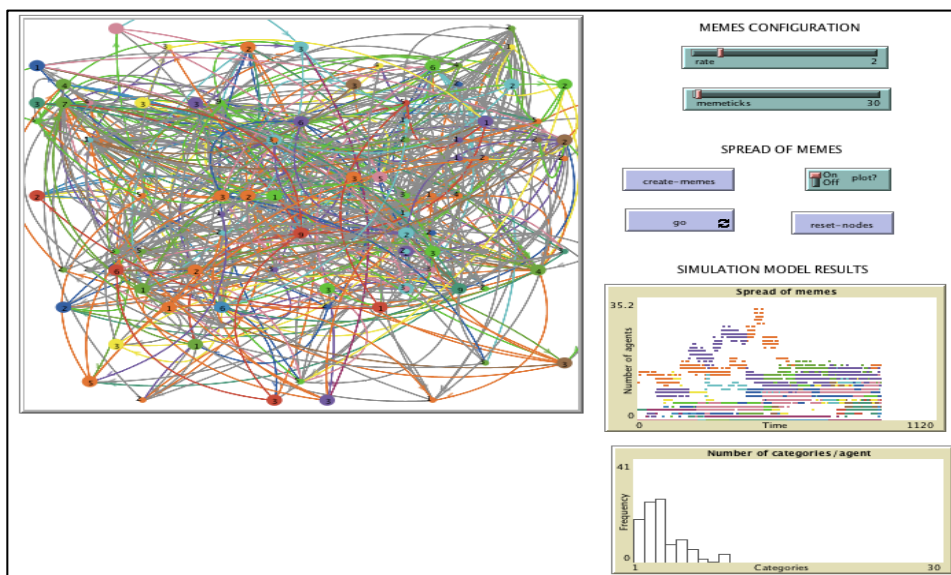
Key words: memes, agent-based modeling, Internet communities, complex adaptive systems.

,

Modeling the effects of preferences and biases to the emergence of totalitarianism

Hugo Barbosa,[1] Freya Casier,[2] José Alejandro Coronado,[3] Marjan Fadavi,[3] Madison
Hart,[4] Valérie Reijers,[5] Carla Natalia Rivera,[6] Adrián Soto,[7, 8] and Carlos Viniegra[9]

[1]*Department of Physics and Astronomy, University of Rochester, Rochester, NY, USA*
[2]*The Danish National Centre for Social Research, Copenhagen, Denmark*
[3]*New School for Social Research, New York, NY, USA*
[4]*University of Texas at Austin, TX, USA*
[5]*Department of Aquatic Ecology & Environmental Biology,*
Institute for Water and Wetland Research, Radboud University, Nijmegen, the Netherlands
[6]*Center of Applied Ecology and Sustainability (CAPES), Santiago, Chile*
[7]*Department of Physics and Astronomy, Stony Brook University, NY, USA*
[8]*Institute for Advanced Computational Science, Stony Brook University, NY, USA*
[9]*Cutter Consortium, Mexico*

Modern societies has undergone profound transformations in the last decade. While on the one hand, we have observed an increase in awareness about individual rights, liberties and inclusion, in contrast, we have also seen the growth and strengthening of movements pushing their agenda in the opposite direction. Indeed, while we observe a popularization of the scientific knowledge, we also witness an increase in the dissemination of ideas and theories in direct contradiction with basic common sense. In short, modern societies have witnessed an increase in polarization in its various dimensions, be they religious, scientific or ideological. In this work we develop an agent based model to explore the influence of different biasing mechanisms to social and economic interactions and how these elements can produce robust and resilient socioeconomic structures. Our results suggest that simple biasing mechanisms can produce non-trivial social structures with resemblances to political structures observed in human societies. We also show that different biasing mechanisms can produce stable or unstable macro-level social structures. Finally, our results also suggest that ideological polarization might be an unavoidable consequence of the overexposure to social interactions provoked by the widespread dissemination of on-line social platforms.
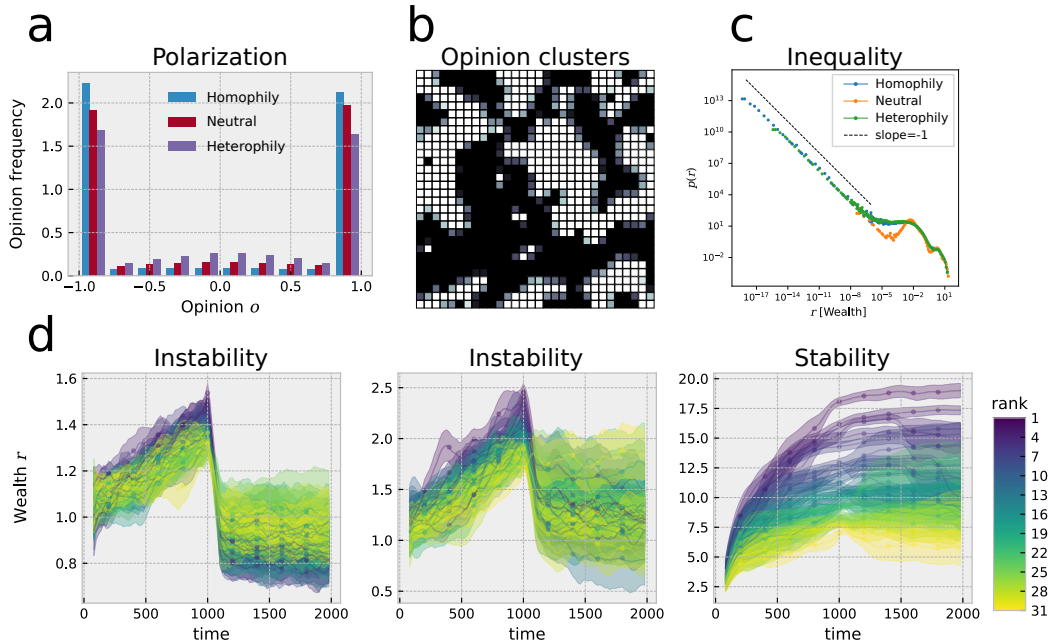
FIG. 1. **(a)** Opinion polarization is persistent regardless the biasing mechanisms in all simulation scenarios. **(b)** The emergence of opinion clusters are fueled by a combination of intra-group convergence and inter-group divergence. **(c)** Heavy-tailed wealth distribution emerges when local interactions are biased in favor of the wealthier. **(d)** Inequality creates a hierarchical power structure that stable and robust, protecting the top of the pyramid from abrupt changes at the bottom.

Organic Urban Boundaries and Clusters – Analyzing Singapore as an Example

Yimin Zhou[1]**, Doheum Park**[2]**, and Sean Wu**[3]

[1]Ministry of National Development of Singapore, Centre for Liveable Cities, Singapore, 069118, Singapore
[2]Korea Advanced Institute of Science and Technology, Graduate School of Culture Technology, Daejeon, 34141, Republic of Korea
[3]University of California, Berkeley, Division of Epidemiology & Computational Biology, Berkeley, 94720, The United States

ABSTRACT

Cities are dynamic complex systems evolving through their functions and human activities. While the regional boundaries are artificially created, dwellers move around them by walking, cycling or taking vehicles with various purposes and such movements in the cities generate "organic" boundaries. In this paper, we introduce how to extract such organic boundaries of cities based on the human mobility data and apply the method to analyze Singapore. We first create Voronoi cells, i.e. unit regions, of the city by using the public transportation and population datasets. Then we spatially interpolate the cells and create the mobility network of the interpolated cells by using the mobility dataset. Applying community detection unveils the organic regions in Singapore and we see that in those self-organizing regions the number of population and facilities have scaling relationships. We believe that our preliminary result is promising and shows its potential in various applications for urban planning, statistics and administration.
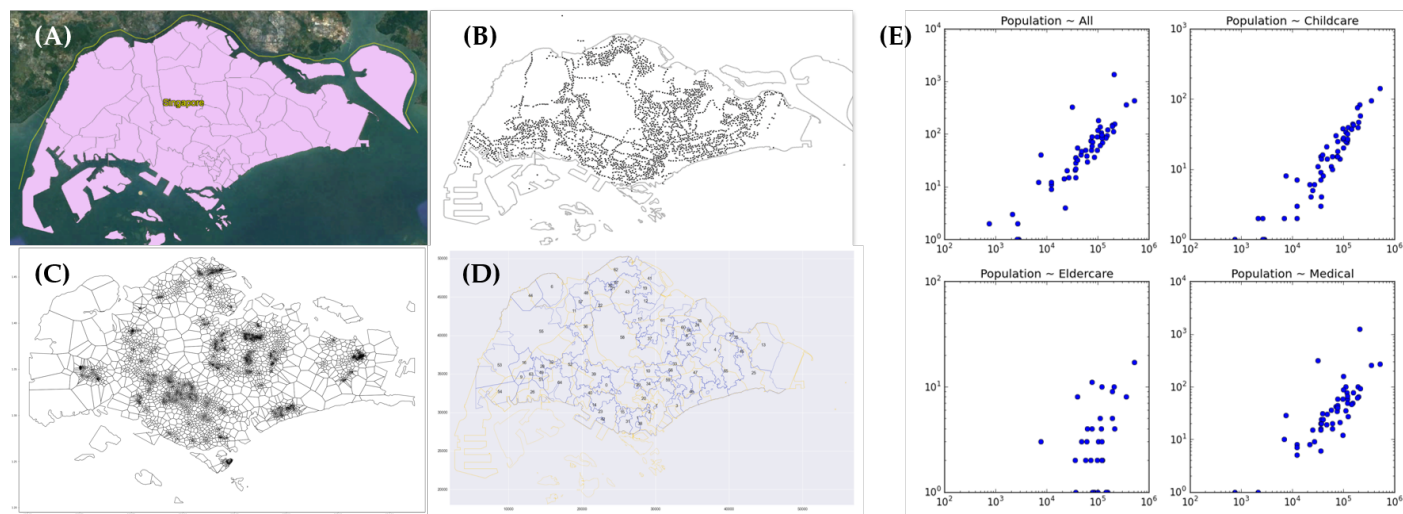
Figure 1. (A) Boundaries of Singapore's 55 planning areas. (B) More than 5,000 transport nodes across Singapore. (C) A map of Voronoi cells for Singapore. (D) Map of 67 clusters. The clusters are in blue while the official planning area boundaries are demarcated in yellow. (E) Examples of scaling relationship. We plot the population size of the clusters against the total number of three types of facilities, and against the number of each type of facilities.

Quantifying and comparing "memory" in biological, ecological, physical, and socio-economic systems

Alexandra Jurgens[1], Alicia Kraay[2], Jake L. Weissman[3], Jingnuo Dong[4], Marco Pangallo[5], Sean Wu[6], Shanee Stopnitzky[7], Shing Hei Zhan[8], Yael Gurevich[9], and Yao Liu[10]

[1]Department of Physics, University of California Davis, CA, USA

[2]Department of Epidemiology, University of Michigan, Ann Arbor, MI, USA

[3]Department of Biology, University of Maryland College Park, MD, USA

[4]Department of Plant and Soil Sciences, Oklahoma State University, Stillwater, OK, USA

[5]Institute for New Economic Thinking at the Oxford Martin School and Mathematical Institute, University of Oxford, Oxford, UK

[6]Divisions of Epidemiology and Biostatistics, University of California Berkeley, CA, USA

[7]Department of Ecology and Evolutionary Biology, University of California Santa Cruz, CA, USA

[8]Department of Zoology, University of British Columbia, Vancouver, BC, Canada

[9]Department of Molecular Biology and Ecology of Plants, Tel-Aviv University, Tel-Aviv, Israel

[10]School of Informatics, Computing, and Cyber Systems, Northern Arizona University, Flagstaff, AZ, USA

ABSTRACT

Most real-world systems demonstrate in their behavior some degree of dependence on past behavior and conditions, or some capacity to store information about the past in their dynamics. This "memory" of a system, be it cognitive or ecological, cultural or physical, has vital implications for those who wish to infer and predict the states and behavior of that system. Each discipline defines memory in its own way, as legacy effects, information transmitted from past to future, information stored in a process, cross-generational response to stimulus, etc. We use an information theoretic framework to unite these definitions and attempt to draw useful comparisons between systems spanning the biological, physical, and social spheres. To do this, we worked with epsilon machines (εM), a type of hidden Markov model in which the states represent causal pasts. We used a technique known as causal state splitting reconstruction (CSSR) to infer these machines from both synthetic data and real-world observations. To develop an inferential basis for statistical complexity using the εM approach, we performed benchmarking experiments, and showed that statistical complexity is positively correlated with memory strength and the pattern is consistent across most realistic perturbation levels and alphabet sizes. We then applied this approach to time series data of precipitation, soil moisture, game theory, animal movement, net ecosystem exchange of CO2 (NEE), and coral reef metabolism. Within and across these systems, the relationship between estimated entropy rate (randomness/unpredictability of the time series) and statistical complexity follows the expected bell-shape, with the highest statistical complexity occurring at intermediate unpredictability (Figure 1). With the cases of soil moisture, game theory, and NEE, we also highlight the usage of statistical complexity as a criterion for determine suitability of process-based system models.
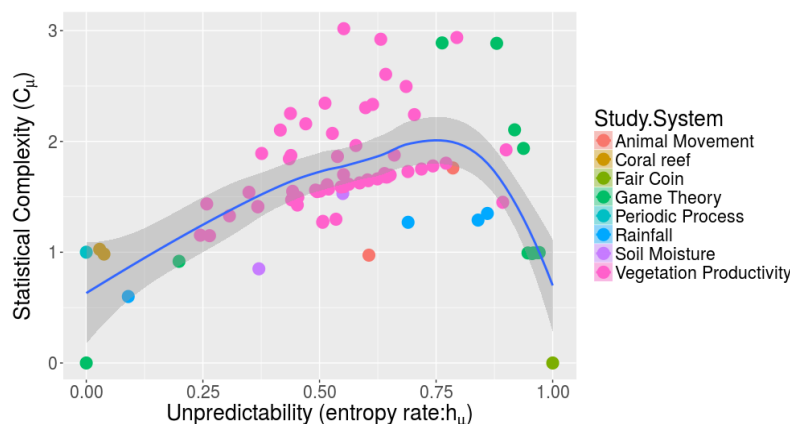
Figure 1. For each dataset we studied we calculated the statistical complexity (C_μ) and entropy rate (h_μ) of the time series, allowing us to build a complexity-entropy diagram that spanned a diversity of system types. Only binary time-series shown. Note the characteristic hump-shaped curve.

Specialists vs. Generalists – Who will conquer the (League of Legends) world?

Martina Balestra[1], Abdel R. Abdelgabar[2], V. Bleu Knight[3], Mario A. Muñoz[4], Shing Hei Zhan[5]

[1] New York University, Department of Management & Innovation, New York, 11201, U.S.A.

[2] Erasmus University Rotterdam, Erasmus School of Economics, Rotterdam, 3062 PA, The Netherlands

[3] New Mexico State University, Department of Biology, Las Cruces, 88003, U.S.A.

[4] University of Melbourne, School of Mathematics & Statistics, Parkville, VIC 3010, Australia

[5] University of British Columbia, Department of Zoology, Vancouver, BC, V6T 1Z4, Canada

ABSTRACT

Contemporary online social participation systems are designed to harness the knowledge, experience, and cooperation of distributed and anonymous participants in a community to complete a task. As a result, they are largely open to anyone who wants to participate - bringing together people with diverse experiences, skill levels, and motivations (among other factors) to collaborate. Participants often have the capacity to act with a high degree of autonomy: they are frequently able to choose the role or task that they want to engage with, and an approach to the task in line with their interests and experience. In this study, we seek to understand how the degree of a participant's specialization in particular types of tasks influences their success in group tasks. Specifically, we examined the behavior of League of Legends players over successive games, and used Shannon entropy to measure their diversity of behavior (Figure 1 shows player behavioral diversity). We then correlated this measure with individual-level outcomes. We found that players who engage in fewer different types of activities over time (i.e., more specialized) gain more in-game wealth than players whose activities are more diverse. Implications are discussed.
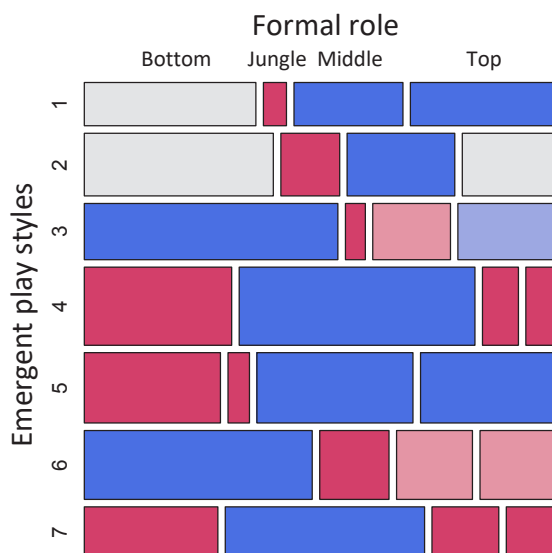
Figure 1. Overlap between emergent play styles and formal roles taken on by players at the outset of a game. Pearson residuals were used as a measure of fit between emergent play styles and formal player roles (red indicates good fit and blue poor fit). The size of each bar corresponds to the proportion of players assigned to a formal role who play in an emergent style.

The stability of Lotka-Volterra metacommunities on random dispersal networks[*]

Burcu Tepekule[1][†] & Gregory L. Britten[2][‡]
[1]Institute of Integrative Biology, ETH, Zurich, Switzerland
[2]Department of Earth System Science, University of California, Irvine, USA

We study the comparative linear stability of Lotka-Volterra metacommunities in cases where dispersal coefficients between individual communities within the larger metacommunity are either uniformly or normally distributed. We also vary the mean interaction strength between species within a community. In the case of uniform dispersal coefficients, stability (as quantified by the leading eigenvalue of the linearized metacommunity steady state) increases with both number of species S and number of nodes N in the metacommunity when mean interaction strength is low. When mean interaction strength is high we observe decreasing stability with S and N. The rate of change in the leading eigenvalue is higher with respect to N than with respect to S in both high and low mean interaction strength cases. In the case of normally distributed dispersal coefficients, we observe similar behavior with increasing (decreasing) stability for weak (strong) mean interaction strength as a function of S and N. However, the relative rate of change in the leading eigenvalue with respect to S is relatively weak for normally distributed dispersal, relative to uniformly distributed dispersal, which may imply that stability of normally dispersing metacommunities is less sensitive to species loss or gain than uniformly dispersing metacommunities. This result would have broad implications for our understanding of metacommunity spatial dynamics and the role of connectivity between environments.
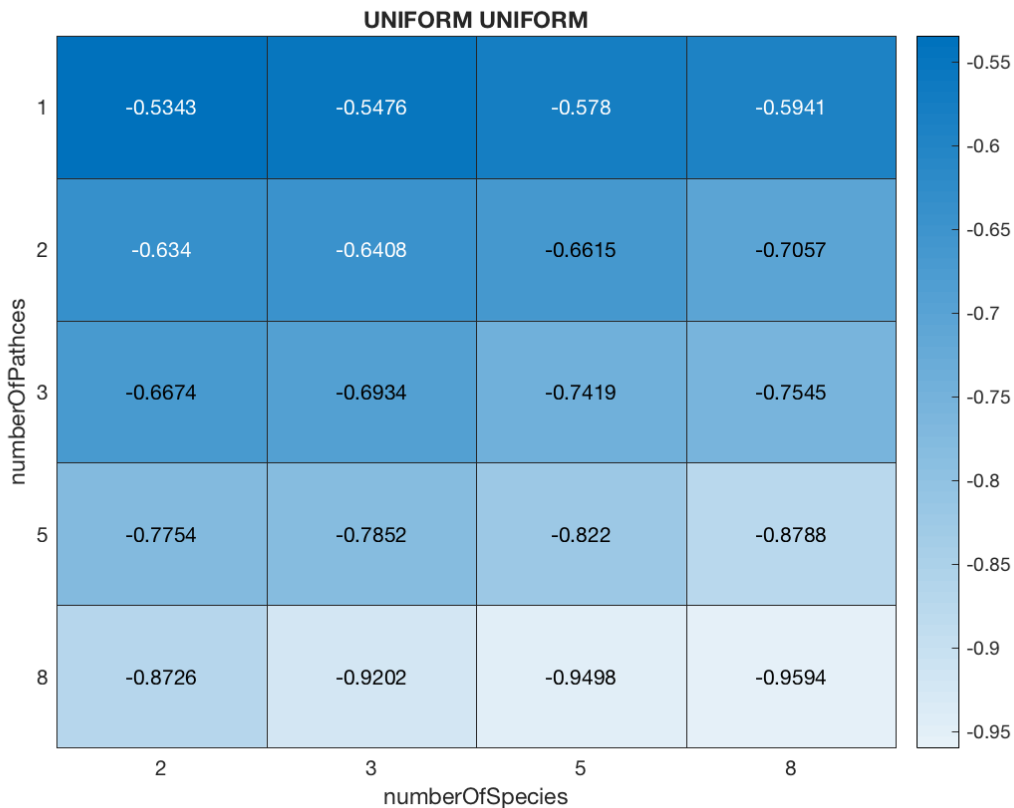
Figure 1: Distribution of the mean leading eigenvalue for a Lotka-Volterra metacommunity as a function of number of species S and number of nodes N for the case of uniform interaction coefficients, uniform distributed dispersal coefficients, and weak mean interaction coefficient

---

[†]burcu.tepekule@env.ethz.ch
[‡]gregleebritten@gmail.com

Topological Analysis of Idealized and Empirical Fitness Landscapes

Dmitriy Kunisky[1], Ximo Pechuan[2]

[1]Department of Mathematics, Courant Institute of Mathematical Sciences, New York, NY 10012.
[2]Department of Systems & Computational Biology, Albert Einstein College of Medicine, Bronx, NY 10461.

Abstract

Fitness landscapes are an essential component of our current understanding of the evolutionary process, summarizing at the highest level of abstraction the dependence of an organism's reproductive fitness on its genetic composition. In this way, fitness landscapes capture one aspect of the mapping between a discrete combinatorial property of an organism, the genotype, and a biological observable it gives rise to, a phenotypic characteristic. The evolutionary dynamics of a population can then be described as a trajectory or an ensemble of trajectories on such a landscape. While fitness landscapes originated in idealized mathematical models of evolutionary biology, there has recently been a proliferation of computational and experimental methods to generate empirical fitness landscapes for many biological molecules of practical interest. One of the most pressing needs in this line of work is to obtain a set of comprehensive, reliable summary statistics to describe and classify the geometry of fitness landscapes in order to answer relevant biological questions.

In this work, we provide such summary statistics using the tools of topological data analysis, in particular based on the technique of persistent homology. We adapt the methods of persistent homology to the lattice structure of fitness landscapes and define two cubical complex filtrations reflecting two different notions of landscape complexity, one capturing local optima and trap-like "bumpy" landscape geometry, and the other capturing epistasis and dependencies between different mutations. We demonstrate that the topological invariants we compute resolve the full range of landscape complexity produced by the mathematical NK and Rough Mount Fuji models of "tunably rugged" landscapes. We then apply the same methods to a large dataset of empirical transcription factor binding affinity landscapes, extending the results of a recent work on simpler summary statistics of these landscapes. We argue that the barcode and Betti number plots of persistence diagrams popular in the persistent homology literature give a finer-grained description of landscape complexity than numerical summary statistics. To further demonstrate the usefulness of these descriptions, we use the notions of bottleneck and Wasserstein distance between persistence diagrams to define a metric structure on fitness landscapes, and describe how clustering in this metric space relates to conventional categorizations of transcription factors.
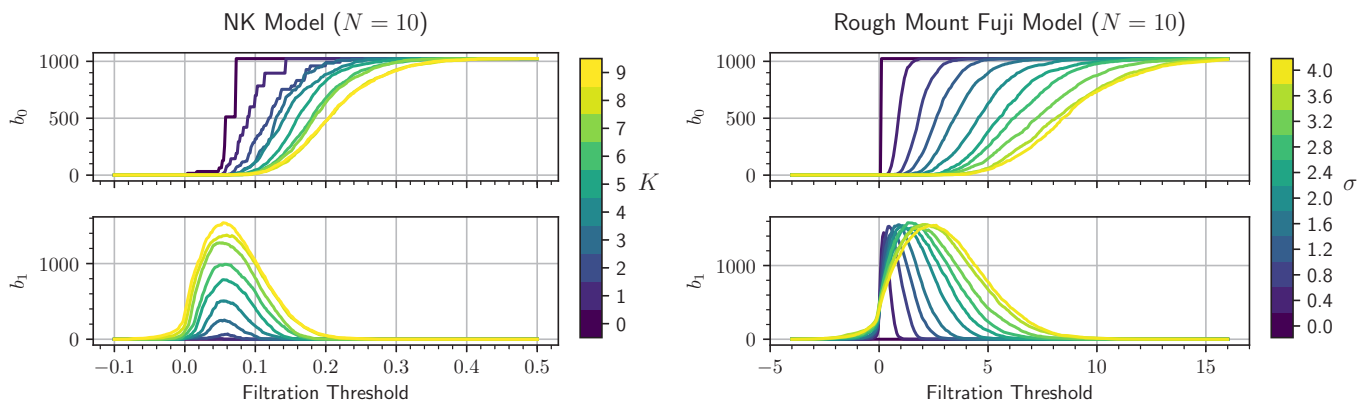
Figure. Betti number summaries of the edge consistency filtration persistent homology of the NK and Rough Mount Fuji mathematical models of fitness landscapes. The computed topological invariants effectively distinguish between different levels of ruggedness.

Trade patterns in international fishery markets

Maartje Oostdijk, Laura Elsler, Andrew F. Johnson, Basak Taraktas, Elaine Bochniewicz, Junfu Zhao & Alberto Micheletti

Abstract

Global trade is a driver of the state of present ecosystems, but the mechanisms through which trade affects ecosystems remains poorly understood. Seafood is one of the most globally integrated food commodities with 40% traded internationally. The increasing globalization of trade in marine products has been pointed to as a potential reason for the overexploitation of wild stocks. Traders can establish new trade relations with other countries which means there is less incentive for the conservation of local resources. We analyzed patterns in trade relations and stock status in exporting countries using global trade and stock assessment databases. We analyzed the evolution of the trade networks of several stocks over time and found increased connectivity and an expanding trade network, though the overall volume of traded stocks stayed relatively stable. Whole networks changed significantly over the years and shared very little similarity (~20%) between years. The 100 and 25 most connected nodes within these networks, however, stayed more stable over time (~50-60 percent similar). In the future we will explore the changing networks by using tools for anomaly detection as well as descriptive network statistics to investigate the patterns that enable the sequential/serial exploitation of marine stocks. We hope that these efforts will culminate in a network model or exponential random graph that describes the underlying mechanisms that cause sequential overexploitation of marine populations.

SIMILARITY BETWEEN THE NETWORKS - COMPARING 1 YEAR TIME DIFFERENCE