2018 COMPLEX SYSTEMS SUMMER SCHOOL

AT THE SANTA FE INSTITUTE

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

2018 COMPLEX SYSTEMS SUMMER SCHOOL
GROUP PROJECT ABSTRACTS

GROUP PROJECT ABSTRACTS (CONTINUED)

Analysis of neutral networks based on global and local sampling of RNA sequences

Sarah J. Berkemer[1,2,+,*], **Kofi K. Khamit-Kush**[3,+], **Jacob P. Rutten**[4,+], **and Alice C. Schwarze**[5,+]

[1]Max Planck Institute for Mathematics in the Sciences, Leipzig, Germany
[2]Bioinformatics Group, Department of Computer Science, Universität Leipzig, Germany
[3]Department of Biology, Clark Atlanta University, Atlanta, GA
[4]Theoretical Biology group, Utrecht University, Netherlands
[5]Mathematical Institute, University of Oxford, Oxford OX2 6GG, United Kingdom
[*]bsarah@bioinf.uni-leipzig.de
[+]these authors contributed equally to this work

ABSTRACT

Fitness landscapes are used to explore the phenotype space based on fitness values. Neutral networks are used to explore fitness landscapes for RNA sequences (genotype) and corresponding secondary structures (phenotype). A neutral network is constructed based on genotypes showing the same phenotype. Sequence being the nodes of the network are connected by an edge if they differ by a single point mutation. Due to the exponential increase in sequence number, it is impossible to explore complete sequence spaces, even for relatively small sequences. Therefore, sampling of biological sequences based on a given secondary structure is used to explore parts of the space.

In this work, we use a new sampling approach where we first uniformly sample sequences from different distant parts of the sequence space using *RNARedPrint*. We then increase the size of the network by locally sampling sequences around each of the sequences sampled at first. In order to keep the network in a reasonable size, we additionally constrain the neutrality criterion such that not only the secondary structure but also the value of its folding energy is considered, using only the lowest 5% of folding energy values. This is done for three different type of structures, an artificial short hairpin structure, a Hammerhead ribozyme and a tRNA. We analyze the structure of the resulting networks and compare with already existing work. We are able to confirm earlier results and show that our sampling technique helps to explore distinct parts of the fitness landscape. We also show that local minima consist of more than one sequence and low folding energy sequences are spread throughout the network and separated by sequences with higher folding energy.

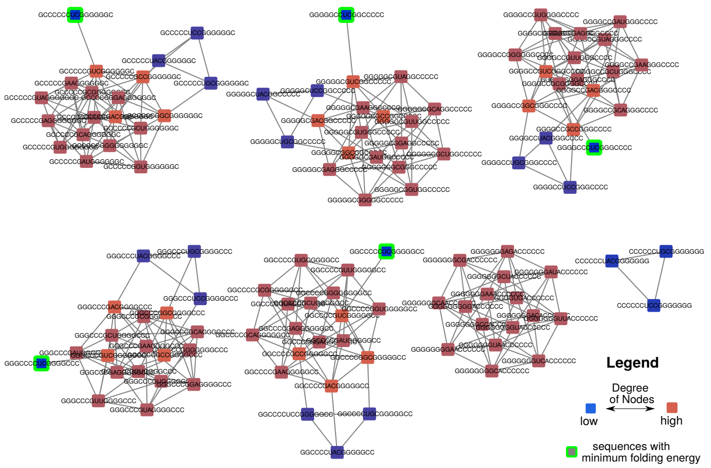


Figure 1. Neutral Network for the artificial small hairpin structure. Nodes represent sequences in the neutral network and are connected by an edge if sequences differ with Hamming distance 1. Node colors indicate the degree of a node which nicely shows strongly connected areas. Sequences folding with minimal folding energy are marked in green. Network visualized using Cytoscape 3.6.

Automated opinion detection analysis of online conversations

Yuki M. Asano
University of Oxford
yuki@robots.ox.ac.uk

Niccolo Pescetelli
MIT MediaLab
niccolop@mit.edu

Jonas Haslbeck
University of Amsterdam
jonashaslbeck@gmail.com

1 Introduction

People's ability to communicate complex concepts through language sets them apart from other animals. We are able to effectively read other people's views and opinions through their conversational language, mapping the spoken words to a lower dimensional space of concepts. Attempts to replicate this natural language understanding capability has lead to the development and application of artificial Natural Language Processing (NLP). Two applications are topic modeling and sentiment analysis, topic modeling maps text into topics whereas sentiment analysis maps text into an affective continuum (typically; bad vs good). However, people's opinions are multifaceted, topic-specific, and highly sensitive to context. The limitations of these two dominant methods have left a large part of communication untouched, for example the sentences *Climate change is a hoax.* and *Climate change is upon us.* express opposite opinions, yet have the same sentiment and topic. In this work, we develop a model that fills the missing third of the sentiment-topic-opinion Venn-diagram to enable a more complete automated understanding of human language.

2 Method

Data: We have chosen 'abortion legalisation' as the topic to analyse. This issue has been historically presented within two polarised camps within public debate: Pro-life (anti-abortion) and Pro-choice (pro-abortion). Both 'camps' carry specific arguments typically based on differing value set making it a good starting-point to train our model. We collect the training data for our model from the sub-reddit pages *pro-life* and *pro-choice*, yielding each about 50k sentences. By substituting synonyms we quadruple our dataset to 200k sentences and achieve better performance on the held-out test-set of classification accuracy of around 72%.

We test our model also on abortion related *Change my View* discussions of reddit (CMVs) and also on a real (fictional) debate between representatives from each camp for which we implemented an additional speech-to-text system.

Model: We use a Bi-directional Gated Recurrent Unit (GRU) with learned word-embeddings followed by a fully connected layer to classify the opinion of a two concurrent sentences. Our architecture parameters are:

3 Initial Results

We have analysed CMV threads and find diverse behaviors with some showing highly correlated opinion expression.

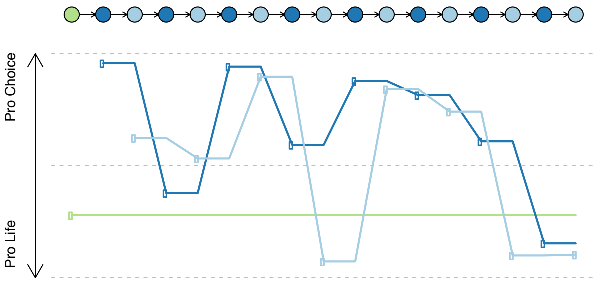


Figure 1: Example CMV conversation chain with two users in a dialogue.

A screen-capture of our speech-to-text system is available here.

4 Conclusion

We have shown the feasibility of classifying opinion in an automated way using machine learning. Our method can be extended to analyze political debates and polarization or serve as a tool for journalists for promoting a healthy online debate.

Characterizing Atrial Fibrillation Dynamics using Multiplex Visibility Graphs

Konstantinos N. Aronis, MD,[1, 2] Anastasiya Salova,[3] Ariadna Venegas-Li,[3, 4] and Andrea Santoro[5, 6]
[1)] *Division of Cardiology, Jonhs Hopkins Hospital, Johns Hopkins University, School of Medicine, Baltimore, MD, USA.*
[2)] *Institute of Computational Medicine, Department of Biomedical Engineering, Johns Hopkins University, Baltimore, MD, USA.*[a)]
[3)] *Department of Physics, University of California at Davis, Davis, CA, USA.*
[4)] *Complexity Sciences Center, University of California at Davis, Davis, CA, USA.*
[5)] *School of Mathematical Sciences, Queen Mary University of London, London E1 4NS, United Kingdom.*
[6)] *The Alan Turing Institute, The British Library, London NW1 2DB, United Kingdom.*[b)]

(Dated: 1 November 2018)

Atrial fibrillation is the most common arrhythmia world-wide. Despite technological advances, the current therapeutic options have only modest long-term success rates[1]. This partially reflects the incomplete understanding of the dynamics underlying atrial fibrillation. In this work we characterize healthy (normal sinus rhythm) and fibrillatory cardiac dynamics by using visibility graphs[2] on multivariate time series[3], derived from simulations and clinical data.

We performed simulations of normal rhythm and fibrillatory dynamics using the mono-domain formulation of the Rogers and McCulloch model. Clinical data were recorded from 4 patients with atrial fibrillation before and after successful catheter ablation procedure using a multi-electrode basket catheter. From multivariate times series of simulated and clinical data, we constructed multiplex visibility graphs. We then compared the structural properties of multiplex visibility graphs between time series from healthy and fibrillatory rhythms. For both simulations and clinical data, the giant component of the multiplex visibility graph was higher in time series of normal rhythm compared to time series of fibrillatory rhythms. The structural efficiency of the multiplex graphs was higher in normal rhythm compared to fibrillatory rhythms.

Multiplex visibility graphs provide novel insights in cardiac dynamics, paving the way for a more in-depth characterization of fibrillatory rhythms. In particular, normal sinus rhythm is associated with multiplex visibility graphs that have greater connectivity and structural efficiency compared to fibrillatory dynamics. Our long-term goal is to develop a network-based method that could be used to define procedural end-points and identify alternative therapeutic targets.

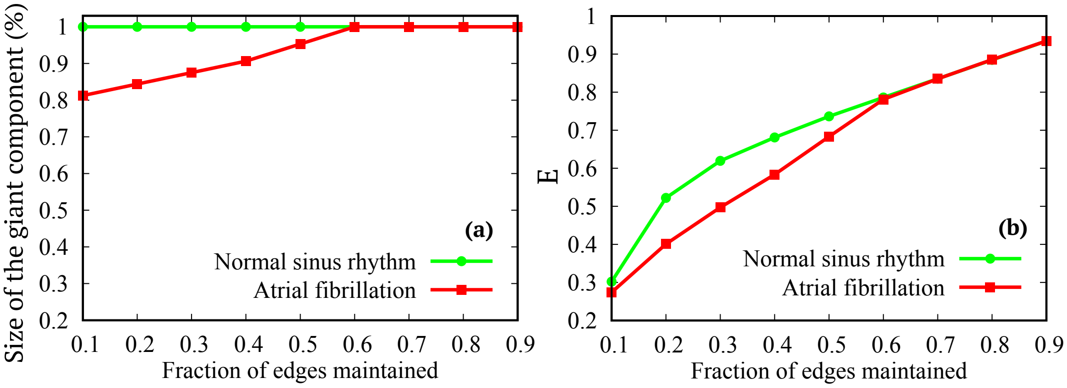


FIG. 1: Size of the giant component (a) and structural efficiency (b) of the network as a function of the fraction of the strongest edges maintained in the clinical data. The two rhythms are successfully distinguished when considering particular weight thresholding regimes (i.e. below 40% of edges maintained).

[1] H. Calkins, et al. 2017 HRS/EHRA/ECAS/APHRS/SOLAECE expert consensus statement on catheter and surgical ablation of atrial fibrillation. *Heart Rhythm*, 14(10):e275–e444, 2017.

[2] L. Lacasa, et al. From time series to complex networks: The visibility graph. *Proceedings of the National Academy of Sciences*, 105 (13):4972–4975, 2008.

[3] L. Lacasa, V. Nicosia, and V. Latora. Network structure of multivariate time series. *Scientific Reports*, 5(1), 2015.

[a)] Electronic mail: karonis1@jhmi.edu

[b)] Electronic mail: asantoro@turing.ac.k

Collective information processing in human rumor spreading networks: An Experimental Study of Rumor Content Fusion and Propagation

Aleksandra Aloric,[1] Javier Garcia-Bernardo,[2] Peter Krafft,[3] Allison Morgan,[4] Zohar Neu,[5] and Andrea Santoro[6]

[1]Center for the Study of Complex Systems, Institute of Physics Belgrade, University of Belgrade, Serbia
[2]University of Amsterdam, Amsterdam Institute for Social Science Research (AISSR), The Netherlands
[3]Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, USA
[4]Department of Computer Science, University of Colorado, Boulder, CO, USA
[5]Department of Engineering Mathematics, University of Bristol, Woodland Road, Bristol, BS8 1UB
[6]School of Mathematical Sciences, Queen Mary University of London, London, UK
(Dated: November 1, 2018)

This study aims to address a gap in the literature related to the mechanisms which humans use to share and filter out information from an unknown original source, and of an unknown quality, in the context of rumor spreading. The findings of past studies on the fidelity and accuracy of propagation of rumors in the real world are at odds with laboratory experimental studies, which find that rumors often become wildly distorted as they pass between individuals. We posit that this difference is largely due to the fact that classical laboratory experiments do not take into account complex interaction mechanisms in human societies. Namely, ways in which groups can combine noisy information to reduce overall uncertainty. We design an experimental setup intended to allow the observation and quantification of such mechanisms, which have been shown to be important in other collective information processing systems. The effect of feedback introduced into the rumor spreading process through redundant transmission pathways in a network structure is investigated. The scale of our experiments is not large enough to give statistically significant results with regards to the importance of network structure However, motif analysis of our results shows that non-trivial information fusion mechanisms are employed by human participants when presented with con accounts of a rumor. We hope that the work presented in this report will pave the way to future experiments and models of information fusion, which will allow us to quantify how human societies collectively process information in the form of rumors.

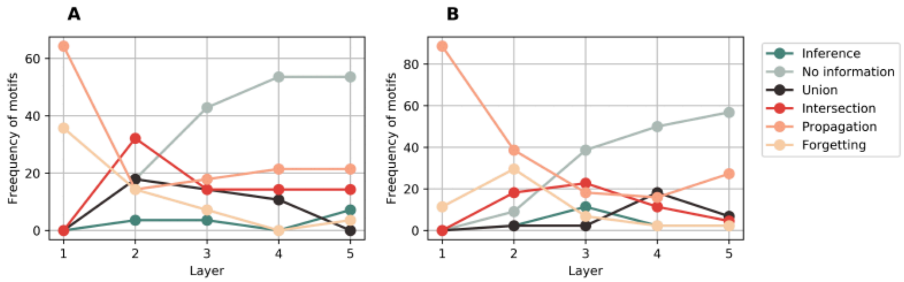


FIG. 1. Frequency of the motifs present in the proposition propagation network of story 1 (panel A) and 2 (panel B). Motifs denote different information processing mechanisms - Forgetting, information transmission and fusion.

Evolution of Water Narratives in Local US Newspapers: A Case Study of Utah and Georgia

Thushara Gunda[1,+], Matthew D. Sweitzer[2], Kevin T. Comer[3], Conor Finn[4,5], Sandra Murillo-Sandoval[6], and Jennifer Huff[7]

[1]Sandia National Laboratories, Albuquerque, NM 87185, United States
[2]School of Communication, Ohio State University, Columbus, OH 43210, United States
[3]MITRE Corporation, McLean, VA 22102, United States
[4]Centre for Complex Systems, Faculty of Engineering & IT, The University of Sydney, NSW 2006, Australia
[5]CSIRO Data 61, Marsfield, NSW 2122, Australia
[6]National Polytechnic Institute (IPN), Mexico City 07738, Mexico
[7]Burke Museum, Seattle, WA 98195, United States
[+] Corresponding author: tgunda@sandia.gov, Sandia National Laboratories, P. O. Box 5800, Albuquerque, New Mexico 87185-MS1137

Abstract

Narratives about water resources have evolved, transitioning from a sole focus on physical and biological dimensions to incorporate social dynamics. Recently, the importance of understanding the visibility of water resources through media coverage has gained attention. This study leverages recent advancements in natural language processing (NLP) methods to characterize and understand patterns in water narratives, specifically in 4 local newspapers in Utah and Georgia. Analysis of the corpus identified coherent topics on a variety of water resources issues, including weather and pollution (Figure 1). Closer inspection of the topics revealed temporal and spatial variations in coverage, with a topic on hurricanes exhibiting cyclical patterns whereas a topic on tribal issues showed coverage predominantly in the western newspapers. We also analyzed the dataset for sentiments, identifying similar categories of words on trust and fear emerging in the narratives across newspaper sources. An analysis of novelty, transience, and resonance using Kullback-Leibler Divergence techniques revealed that topics with high novelty generally contained high transience and marginally high resonance over time. Although additional analysis needs to be conducted, the methods explored in this analysis demonstrate the potential of NLP methods to characterize water narratives in media coverage.

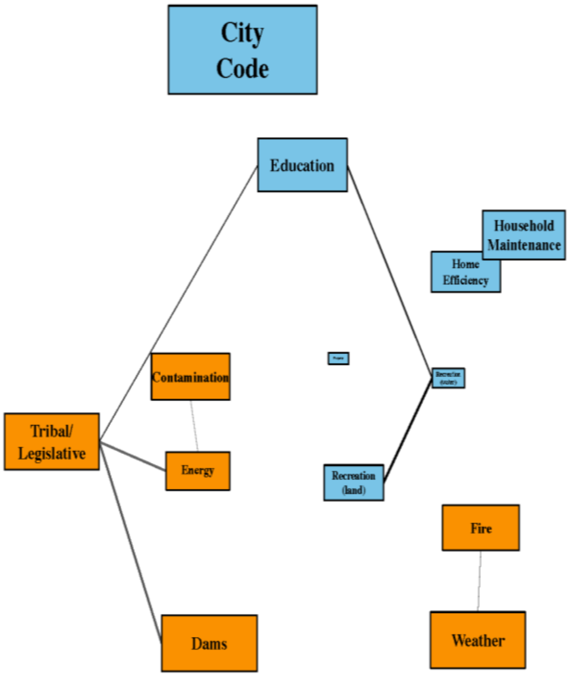


Figure 1: Network graph of final 13 topic correlations. Topics of interest within the final structural topic model are colored in orange. Line thickness indicates the strength of correlations.

This page intentionally left blank

Fluctuations in 1D point patterns

Rosalba Garcia-Millan[1,*], **George Cantwell**[2,†], **Konstantinos Aronis**[3,‡], **and Carlos Marcelo**[4,§]

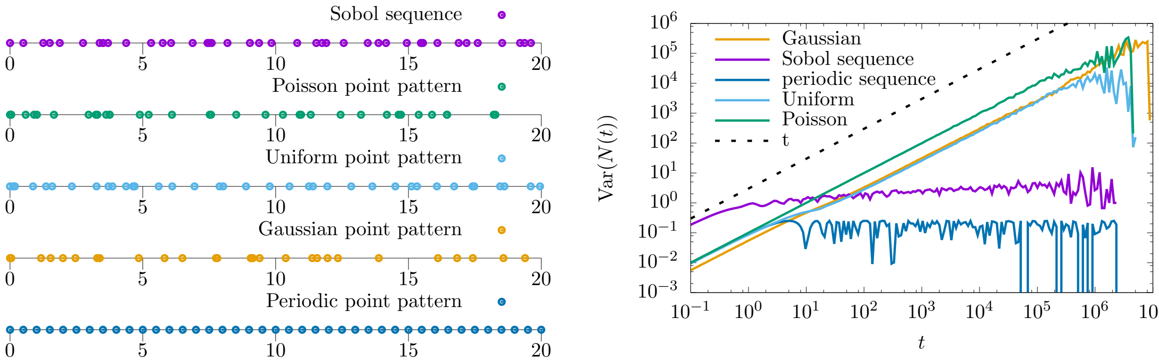
[1]Imperial College London
[2]University of Michigan
[3]Johns Hopkins School of Medicine
[4]University of Chicago
[*]garciamillan16@imperial.ac.uk
[†]gcant@umich.edu
[‡]karonis1@jhmi.edu
[§]cmarceloservan@uchicago.edu

ABSTRACT

Random point patterns with similar apparent characteristics (such as density of points) may display very different behaviours when it comes to the way these points are distributed in space. Strongly negatively correlated points (e.g. due to repulsive forces) show a high degree of order, whereas strongly positively correlated points (e.g. due to attractive forces) show clustering and hence, a high degree of disorder. These two extreme cases define a spectrum where point patterns are classified according to their fluctuations from small fluctuations (order, *hyperuniformity*) to large fluctuations (disorder, clustering). Point patterns with weak correlations show the same trend in their fluctuations as the typical uncorrelated case and are said to display Poisson-like behaviour.

In this report, we show numerical results on different one-dimensional point patterns both computer generated and experimentally obtained (heartbeat recordings). Furthermore, we regard the case of anomalously suppressed fluctuations (*hyperuniformity*) as a case of special interest and we try to derive the underlying mechanisms that give rise to this behaviour. We study the Dirichlet distribution as a model that can be tuned to display any behaviour from a periodic signal to a clustered signal.

This report is a summary of early work that we have done during the Complex Systems Summer School 2018 organised by the Santa Fe Institute, New Mexico. Throughout the report we discuss a number of directions that we can take to further research about the questions that we pose here.



(a) (b)

Figure 1. (a) Sample of the considered 1D point patterns (b) Curves $\mathrm{Var}(N(t))$ for the different point patterns shown in Fig. 1(a), obtained from drawing 10^6 waiting times in each case and using Eq. (2). We observe that the point patterns "periodic" and "Sobol" show anomalous suppression of fluctuations (hyperuniformity) and "Gaussian", "uniform" and "Poisson" show the typical Poisson-like behaviour. The dashed line indicates the typical Poisson-like behaviour in 1D point patterns, namely the scaling form $\mathrm{Var}(N(t)) \propto t$.

The impact of technology intervention on the sustainable development goals[1]

Magdalena M. Klemun[a], Sanna Ojanperä[b], Amy Schweikert[c]

[a]*Massachusetts Institute of Technology, Cambridge, Massachusetts, USA*
[b]*University of Oxford, Oxford, and The Alan Turing Institute, London, UK*
[c]*The Colorado School of Mines, Golden, Colorado, USA*

Achieving the 17 United Nations Sustainable Development Goals (SDGs) set out in the 2030 Agenda for Sustainable Development adopted in 2015 requires progress along multiple dimensions of human development. In many cases, the progress will involve using new or improved technology. However, some technological interventions can lead to contradictory changes among other indicators. As a very simplified example, building a new factory may increase employment (positive progress on Goal 8.2, 8.3, 8.5), creating higher incomes and potentially reducing hunger in a community indirectly (positive progress on Goal 2.1), but might simultaneously increase greenhouse gas emissions from manufacturing in the country (reverse progress on Goal 13.2).

This simple example epitomizes larger and more complex questions, including a need to understand these types of interactions between different goals more comprehensively. Specific questions include: how do simplified, and more complex, contradictions such as the example described above emerge? Are there combinations of technologies that make them less likely, and if so, why? Does the sequence (of how technologies are introduced) make a difference? This project builds on recent work addressing the interconnections among SDG indicators using a complexity and network theory approach. We take a technology-focused view and investigate the effects of introducing a new or improved technology portfolio into an existing network of resources, technologies, and industries.

The 17 SDGs are delineated into 232 indicators detailing sub-components necessary to the achievement of each goal. Since many technologies require a similar set of resources and industrial support regardless of where they are being manufactured or used, this project focuses on building a location-independent network of interdependencies between technological intervention and the resulting interconnections at the level of the SDG indicators. A specific technological intervention (addition of photovoltaic electricity) aimed at positive progress in a specific SDG goal (Goal 7) is then studied as a network problem and evaluated in terms of resulting changes cascading through the larger SDG indicator network. One goal of this work is to set up a methodology that can be adapted to address country-specific policy questions and study a variety of technologies.

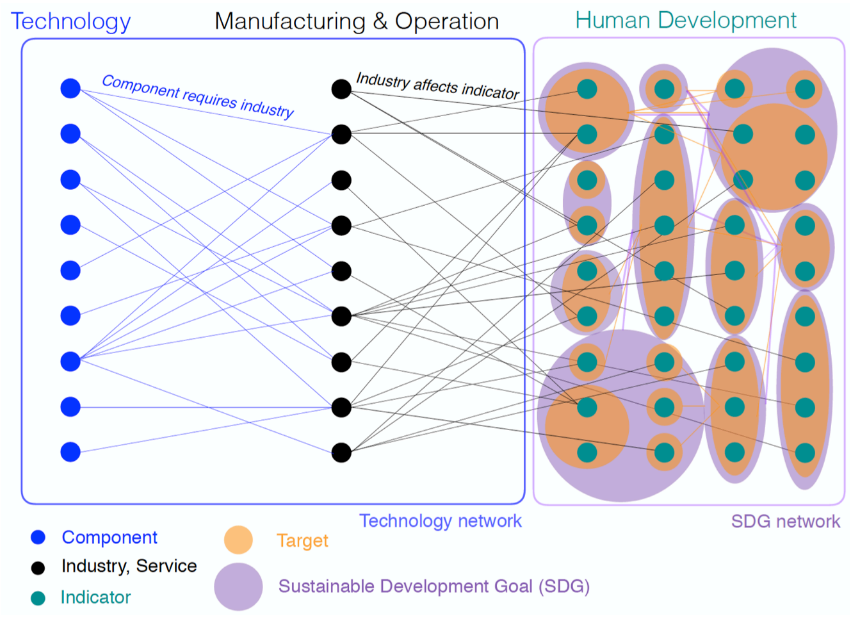


Figure 1: Technology networks (large blue box, left) consist of technology components (blue circles: e.g., the module or the inverter of a photovoltaic system) and the industries and services needed to manufacture, deploy, operate, and maintain these components (black circles). SDG networks (large purple box, right) consist of indicators (green circles) that measure progress towards the targets (orange shapes) underlying each SDG (purple shapes). Technology networks are connected to SDG networks through development indicators (green circles) that are affected by technology-supplying industries and services. One SDG is connected to another (purple lines) if one or more targets related to that SDG are related to targets related to the second SDG (orange lines).

---

[1] This project was conceptualized at The Sante Fe Institute's Complex Systems Summer School (2018)

Industrial Economic and Spatial Clustering in Singapore

Shantal Lichang Cheong[1,*], **Ricky Laishram**[2], **Sanna Ojanperä**[3], **Alex Shannon**[4], **and Xindi Wang**[5]

[1]Strategic Planning Group, Urban Redevelopment Authority, Singapore
[2]Electrical Engineering and Computer Science, Syracuse University, NY, USA
[3]Oxford Internet Institute, University of Oxford and the Alan Turing Institute, United Kingdom
[4]Center for Urban Science and Progress, New York University, NY, USA
[5]Network Science Institute, Northeastern University, Boston MA, USA
[*]shantal_cheong@ura.gov.sg

ABSTRACT

Economic activities between and within industrial sectors serve as a primary driver of urban growth. Clustering effects between sectors can occur organically or inorganically (e.g. through government intervention via subsidies or grants or policy regulations), and are defined by geography, value creation, and business environment[?]. Industrial clustering facilitates a business ecosystem that can benefit from economies of scale and the flow of information, goods, and services. Clustering effects may be observed differently depending on the nature and type of industrial activities. A better understanding of industrial clustering effects for industries will enable governments to formulate increasingly informed policies and strategies to foster the growth of economic ecosystems. It will also bolster strategic decision-making for firms and companies in evaluating the location, size, and type of economic activity for their businesses.

In this study, we analyzed economic and geographical data from Singapore in an attempt to understand if economic and spatial features in the context of a small city-state exhibit clustering effects. A a small island of 719 km^2, Singapore is ranked among the top ten countries in the world for per capita Gross Domestic Product (GDP). Using measures on the Total Requirements Coefficient between all economic sectors in the economy, our study analyzed the clustering network to identify sectors that exhibit strong economic linkages among each cluster. Geographical clustering effects are studied in detail using data on the geographical location of sectoral clusters as nodes, weighted by the cluster size in terms of nett floor area (NFA) at each cluster node.

Using the clustering network detected in the first stage of the study, we examined the relationship between economic linkages and spatial clustering across and within industrial sectors. To address this, cluster matching analysis was conducted to study the strength of the relationship between both types of clustering network. This study found that the economic and geographical clustering activity exhibit global similarity of approximately 0.6, which suggests that on a broad level, the economic linkages and spatial clustering of industrial sectors tend to be correlated. Lastly, pairwise asymmetric comparison between industries in the broad sector was proposed to evaluate if industries grouped together in the spatial cluster were also grouped together in the economic clusters.

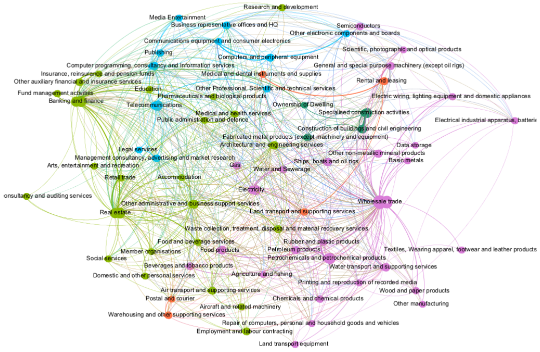


Figure 1. Cluster result on 71 sub-sectors of Singapore.

Investigating Phase Transitions in a Cardiac System through Informational Properties of Renewal Process Models

Konstantinos N. Aronis, MD,[1,2] Ariadna Venegas-Li,[3,4] Anastasiya Salova,[4] and Andrea Santoro[5]
1) *Division of Cardiology, Jonhs Hopkins Hospital, Johns Hopkins University, School of Medicine, Baltimore, MD, USA.*
2) *Institute of Computational Medicine, Department of Biomedical Engineering, Johns Hopkins University, Whiting School of Engineering, Baltimore, MD, USA.*a)
3) *Complexity Sciences Center, University of California at Davis, Davis, CA, USA.*
4) *Department of Physics, University of California at Davis, Davis, CA, USA.*
5) *School of Mathematical Sciences, Queen Mary, University of London, London, UK.*

(Dated: 1 December 2018)

Development of ventricular tachycardia and degeneration of it into ventricular fibrillation can be considered a phase transition in cardiac dynamics. Characterization of the underlying physical mechanisms may lead to better prediction of the onset of this phase transition and enable efficient treatment of patients at risk. Aiming to contribute to this characterization we study time series of activity in simulated heart tissue. We use Bayesian structural inference to find the best fit models in a set of hidden Markov models representing renewal processes. We compute the entropy rate and statistical complexity of these models and compare between results for healthy heart activity, heart activity at the onset of the phase transition and heart activity after the phase transition. We discuss the advantages and shortcomings of this approach, as well as alternative improvements.

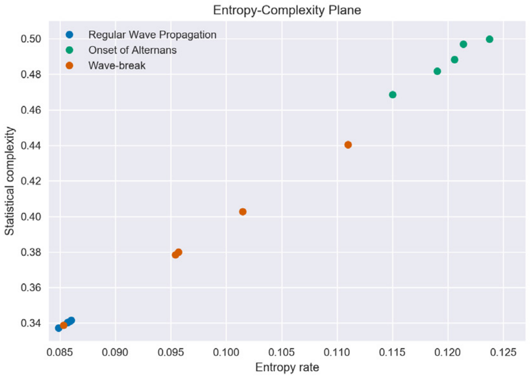


FIG. 1. Points in the plot represent estimated values of entropy rate and statistical complexity for time series of voltage activity recorded in five different sites of a cardiac dynamic simulation. Each of the sites is simulated in three different regimes: regular wave propagation (representing healthy heart tissue), near the onset of alternans (representing a phase transition into ventricular fibrilation), and after the wave-break (representing heart tissue undergoing ventricular fibrilation).

---

a) Electronic mail: karonis1@jhmi.edu

Nighttime Light, Superlinear Growth, and Economic Inequalities at the Country Level

Ore Koren
Department of Political Science, Indiana University
Bloomington, USA
Email: okoren@iu.edu

Laura Mann
Jet Propulsion Laboratory, California Institute of Technology
Pasadena, USA
lauraerinmann@gmail.com

Abstract—Research has highlighted relationships between size and scaled growth across a large variety of biological and social organisms, ranging from bacteria, through animals and plants, to cities an companies. Yet, heretofore, identifying a similar relationship at the country level has proven challenging. One reason is that, unlike the former, countries have predefined borders, which limit their ability to grow "organically." This paper addresses this issue by identifying and validating an effective measure of organic growth at the country level: nighttime light emissions, which serve as a proxy of energy allocations where more productive activity takes place. This indicator is compared to population size to illustrate that while nighttime light emissions are associated with superlinear growth, population size at the country level is associated with sublinear growth. These relationships and their implications for economic inequalities are then explored using high-resolution geospatial datasets spanning the last three decades.

Keywords-nighttime light, scaled growth, economic development, Inequality

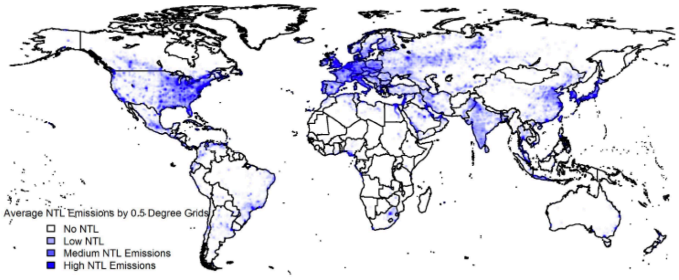


Figure 3. Average NTL Emissions by 0.5 Degree Grids, 1992–2013

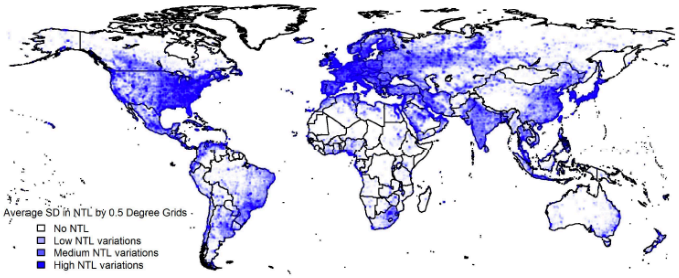


Figure 4. SD in NTL Emissions by 0.5 Degree Grids, 1992–2013

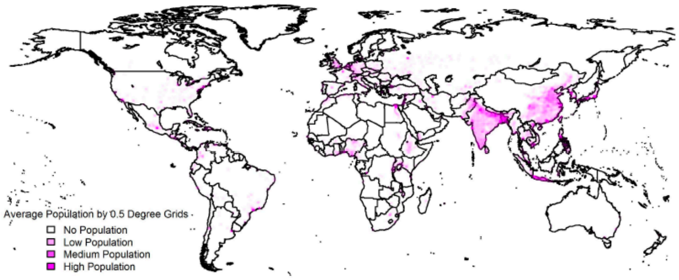


Figure 5. Average Population by 0.5 Degree Grids, 1995–2010

Mean First Saturation Time (MFST) and its application to Refugee movement

R. María del Río-Chanona[1], Benjamin F. Maier[2], Guillaume St-Onge[3], George T. Cantwell[4]

1. Mathematical Institute, University of Oxford
2. Robert Koch Institute, Nordufer 20, D-13353 Berlin
3. Département de Physique, de Génie Physique, et d'Optique, Université Laval, Québec (Québec), Canada
4. Department of Physics, University of Michigan, Ann Arbor, Michigan, USA

In this work, we study the mean first saturation time (MFST), a generalization to the mean first passage time, on networks and show an application to the 2015 Burundi refugee crisis. The MFST between a sink node j, with capacity s, and source node i, with n random walkers, is the average number of time steps that it takes for at least s of the random walkers to reach a sink node j. We expand the literature by exploring the behaviour of the MFST for node-biased random walks in ErdősRényi random graph and geographical networks. Furthermore, we apply MFST framework to study the distribution of refugees in camps for the 2015 Burundi refugee crisis. For this last application, we use the geographical network of the Burundi conflict zone in 2015. In this network, nodes are cities or refugee camps, and edges denote the distance between them. We model refugees as random walkers who are biased towards the refugee camps which can hold s_j people. To determine the source nodes (i) and the initial number of random walkers (n), we use data on where the conflicts happened and the number of refugees that arrive at any camp under a two-month period after the start of the conflict. With such information, we divide the early stage of the Burundi 2015 conflict into two waves of refugees. Using the first wave of refugees we calibrate the biased parameter β of the random walk to best match the distribution of refugees on the camps. Then, we test the prediction of the distribution of refugees in camps for the second wave using the same biased parameters. Our results show that the biased random walk can capture, to some extent, the distribution of refugees in different camps (see figure 1). Finally, we test the probability of saturation for various camps. Our model suggests the saturation of one or two camps (Nakivale and Nyarugusu) when in reality only Nyarugusu camp saturated.

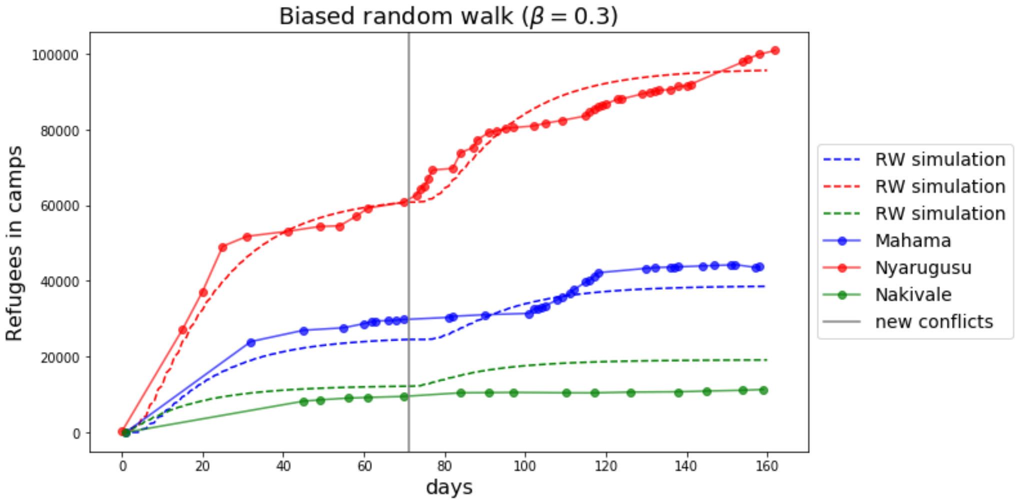


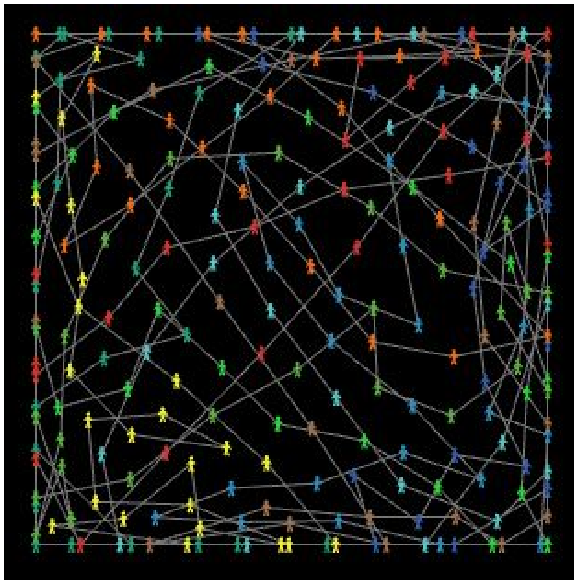
Figure 1: Number of refugees in 3 camps across time. The dotted lines show the results of our model (biased random walk), while the straight lines show the real distribution. The grey line indicates the time at which the second wave of refugees starts.

Modeling the Dynamics of Wealth Inequality

Haripriya Chakraborty[1], Vandana R. Venkateswaran[2], Tom Pike[3], Duy Huynh[4],
Eleonora Mavroeidi[5], Javier Garcia-Bernardo[6], Cesar Montalvo[7]

[1]*The Graduate Center, City University of New York, USA*
[2]*Max Planck Institution for Evolutionary Biology, Germany*
[3]*George Mason University, USA*
[4]*MITRE, USA*
[5]*OECD, FRANCE*
[6]*University of Amsterdam, Netherlands*
[7]*Arizona State University, USA*

Many experts regard increasing inequality as one of the most pressing issues facing our world. The persistent rise in wealth inequality over the last 40 years is a result of increasing income inequality, a large transfer of wealth from the public to private hands, and systemic barriers to economic mobility. Global inequality (share of wealth to the top 1%) has increased during the period 1980 and 2015, from 16% to 20%, respectively, while the bottom 50% of the population accumulated a share of global income of approximately 9%. Inequality has drastic negative consequences for those living in poverty as well: around 56% of the global population lives on between $2 and $10 a day. However, Oxfam recorded the biggest increase in billionaires in history during 2017. Billionaires increased their wealth by $762 billion during that year. This could have ended global extreme poverty seven times. Moreover, 82% of the new wealth created has gone to top 1%, while 0% has gone to the world's poorest 50%. In light of these staggering numbers, it is useful to have an accessible framework which can be used to analyze the dynamics of wealth inequality and to simulate the effects of various policies that may reduce or exacerbate wealth inequality. In this work, we propose a theoretical framework to study the evolution of wealth inequality in a given population. We consider, among other things, demographic variations, policy effects, and other types of inequality. In addition, we present an agent based model to simulate some of the dynamics of wealth inequality.



Caption: The python based model using 10,000 agents and 30 simulated years. The results consistently returned either agents becoming super rich or agents going into massive debt.

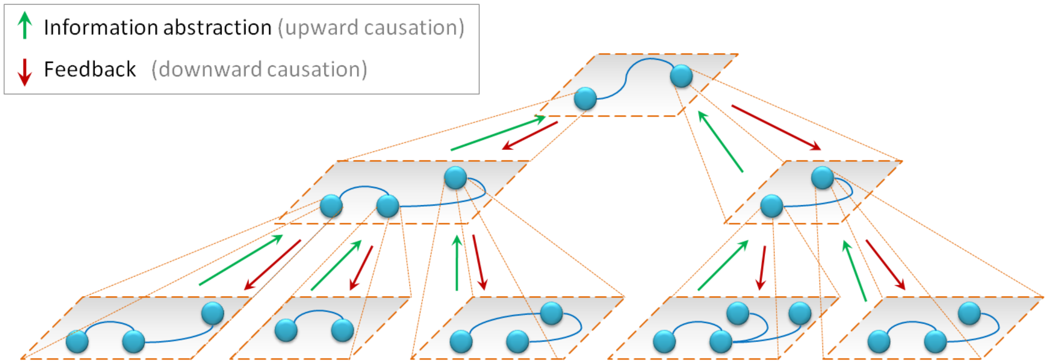
Multi-scale Adaptive Systems

Structural Features and Qualitative Properties

Ada Diaconescu[*], **Louisa Jane Di Felice**[**], **Patricia Mellodge**[***], **Josefine Bohr Brask**[****]

[*] Télécom ParisTech, France; [**]Autonomous University of Barcelona, Spain;
[***]University of Hartford, U.S.; [****]University of Exeter, UK

Abstract. Most large-scale systems, both natural and artificial, appear to feature *hierarchical* architectures, with multiple levels operating at multiple *scales* -- in terms of space, time and information -- and interrelated via various *feedback* loops. Considering the wide-spread occurrence of such multi-scale adaptive systems, or 'hierarchies', we aim to determine the key *structural properties* that lead to the viability and success of this system design. We also aim to link variations of such structural properties to *qualitative system properties* such as reactivity, adaptability, stability and resilience. We intend to develop this project via an iterative methodology, based on: i) analysing existing system examples and theories; ii) distilling a *conceptual framework* based on reoccurring structural features; and, iii) evaluating the framework based on various model simulators. The present report sets the initial foundation for the conceptual framework, based on previous work, analysed examples and theories. It also proposes an agent-based simulation model for future evaluations. Future work will analyse further examples from a wide range of natural and artificial domains; extend and refine the conceptual framework; implement the proposed simulator, and employ it to further validate the conceptual framework. Considering the increasing deployment of complex networked technologies that may fundamentally alter the underlying structure of our socio-economic systems, it seems rather urgent to more thoroughly understand and manage such systems.



Structural overview of a multi-scale adaptive system (or hierarchy)

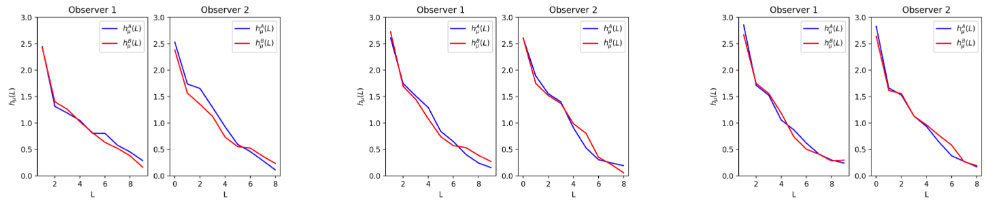
A Quantitative Approach to Contact Improvisation Dance

Sarah Holmes[1,*], Gianroco Lazzari[2], Patricia Mellodge[3], Alexandra Mikhailova[4], Vandana Revathi Venkateswaran[5], and Ariadna Venegas-Li[6]

[1] Kennesaw State University, Department of Dance, Kennesaw, GA
[2] Global Health Institute, School of Life Sciences, Ecole Polytechnique Féedérale de Lausanne (EPFL), Lausanne, Switzerland
[3] College of Engineering, Technology, and Architecture. University of Hartford, West Hartford, Connecticut
[4] Center for Neuroscience, University of California, Davis, Davis, California
[5] Department of Evolutionary Theory, Max Planck Institute for Evolutionary Biology, Plön, Germany
[6] Physics Department and Complexity Sciences Center, University of California at Davis, Davis, CA.
*sholme39@kennesaw.edu

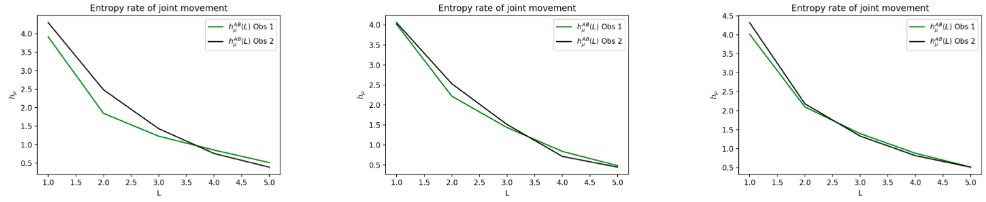
ABSTRACT

Partner improvisation is characterized by prediction, adaptation and emergent outcomes that are unique to the pair. These definitions have not been stringently applied to improvisational dance, therefore we propose a way to measure the influence of partners on the emergent movements of each other. As a pilot study, we use a simple alphabet of body contact (torso, arms, legs) to code 3 videos of improvisation pairs. We record the point of contact for each partner, and then use this information to compute measures of mutual information and entropy rate. We find that mutual information between partners is higher as the level of the improvisation grows, which is consistent with our understanding of prediction in improvisation. Our study finds that the level of experience in contact improvisation of the annotator makes a difference when scoring the contact between partners. Entropy rate calculations have a big bias due to shortage of data, and do not distinguish between the pairs of dancers.



(a) Lower expertise CI video. **(b)** Intermediate expertise CI video **(c)** Advanced expertise CI video

Figure 1. Entropy rate estimates of each dancer individually, as a function of length of sequence considered (L). Different colored lines represent results obtained from sequences of contact coded by different scorers.



(a) Lower expertise CI video. **(b)** Intermediate expertise CI video. **(c)** Advanced expertise CI video.

Figure 2. Entropy rate estimates from time series of joint movement between two dancers as function of length of sequence considered (L). Different colored lines represent results obtained from sequences of contact coded by different scorers.

Quantifying transmission and identifying factors associated with the malaria outbreak in Venezuela

Talia M. Quandelacy[1], Gianrocco Lazzari[2], Inga Holmdahl[3], and Cesar Montalvo[4]

1. Centers for Disease Control and Prevention, San Juan, Puerto Rico
2. Global Health Institute, School of Life Sciences, Ecole Polytechnique Federale de Lausanne, Lausanne, Switzerland
3. Center for Communicable Disease Dynamics, Harvard TH School of Public Health, Boston, Massachusetts
4. Simon A. Levin Center, Arizona State University, Tempe, Arizona

Venezuela is currently in a humanitarian crisis, stemmed by high inflation rates and decreasing economic security. The political and economic instability has caused a food shortage and the re-emergence of infectious diseases, including malaria, a disease previously eliminated in Venezuela in 1977. However, malaria had resurged since 2015 and continues to spread, due to the migration of workers to and from gold mines along the border of Guyana. Large proportions of the population are also leaving Venezuela, and the effect of this out-migration on the spread of malaria in other countries is unknown.

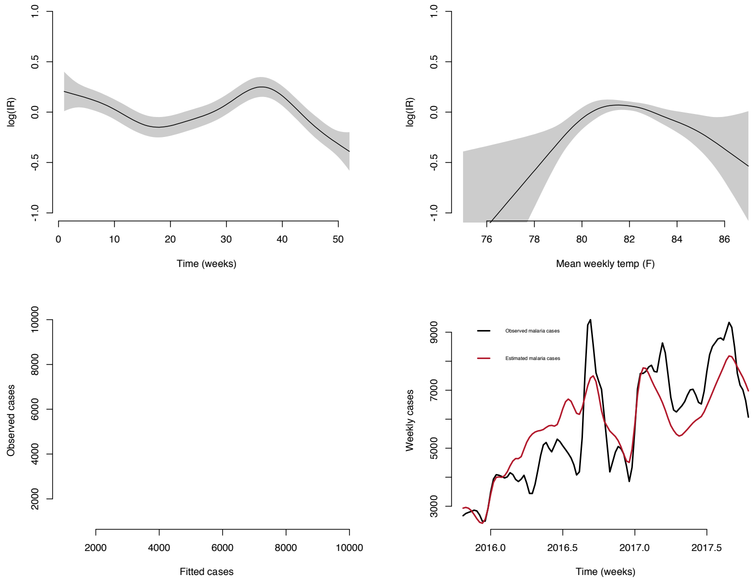


Figure 1. Economic and seasonal factors associated with weekly malaria incidence rates (95% CI are shaded), observed and estimated cases, and estimated cases over time.

Understanding the characteristics of the outbreak and the extent of its spread has been challenging because the government of Venezuela have stopped publicly reporting on the outbreak since the 42nd week of 2017. Here we provide summary of factors associated with the malaria outbreak, and assess the effect of imported cases from Venezuela on malaria transmission in the Colombia. We estimated of incident malaria cases occurring in Venezuela from the 43rd week of 2017 to the 35th week of 2018 (i.e. the time period since weekly reported surveillance stopped), using a Poisson regression model. We examined temporal, economic (inflation rate), and climatic variables (temperature, rainfall, and precipitation) as linear and spline terms using a generalized additive model. Increases in the exchange rate were associated with a 19% (IR: 1.19, 95 CI: 1.15, 1.24) increase in the weekly incidence rate of malaria after adjusting for time and mean weekly temperature (Figure 1). Our model used a spline term for weeks, average weekly temperature, and weekly exchange rates had the highest R^2 (R^2=0.68). An ARIMA model was also applied to weekly malaria cases to forecast weekly malaria cases in 2018. We also used case and out-migration data to estimate the contribution of cases imported from Venezuela towards local transmission in Colombian provinces. Our analyses provide a quantitative examination of the current malaria outbreak, both within Venezuela, and its impact on neighboring countries.

Scaling of Sensory Information Intake in Organisms

Elan Stopnitzky[1,*] **and Louisa Jane Di Felice**[2]

[1]University of Hawaii at Manoa, Physics and Astronomy, Honolulu, United States
[2]Universitat Autonóma de Barcelona, Institute de Ciència i Tecnologia Ambientals, Barcelona, Spain
[*]elanstop@hawaii.edu

ABSTRACT

The metabolic resources required by organisms are known to scale with their mass raised to the $3/4$ power. Like energy and nutrients, information about the environment is also a necessary resource for life. However, little attention has been placed on how the capacity to absorb information scales with size. We have taken steps toward answering this question by a) synthesizing data on sensory capacity from multiple organisms to try and answer the question empirically and b) exploring theoretical considerations that might allow us to deduce the scaling laws from first principles. We hypothesize that as organisms increase in size and complexity, their sensory capacity per unit mass decreases, owing to the fact that less total information is needed if it can be processed in a more sophisticated manner. Preliminary results, both theoretical and empirical, are supportive of this claim. In particular, we derive a fundamental bound on the sensory capacity of a spherical cell arising from the heat it dissipates, which turns out to be a linear function of the cell radius r. As the mass of a cell scales as r^3, our bound implies that the maximum achievable channel capacity per unit mass decreases with the size of the organism. We also investigate the hypothesis that organisms at all scales devote a fixed fraction of their metabolic rate to gathering sensory information, and find tentative support for this claim.

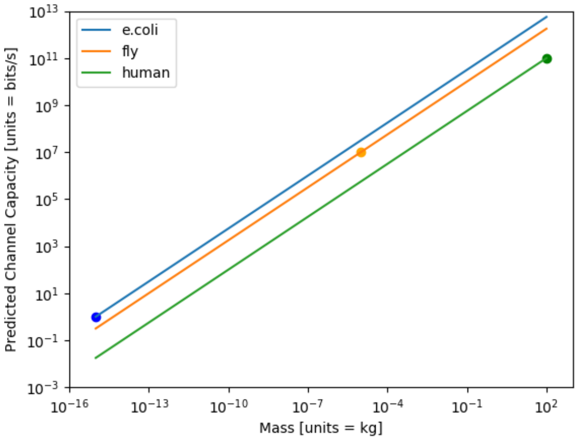


Figure 1. Predicting the scaling of information intake under the assumption that organisms devote a fixed fraction of their metabolic rate to information gathering. Note that flies fall below the extrapolated line for E. coli and that humans fall below the extrapolated line for flies.

Spatio-temporal patterns of smallpox epidemics in the United States from 1915 to 1950

Nikunj Goel[1], Gianrocco Lazzari[2], Inga Holmdahl[3], Peter Geissert[4], Talia M. Quandelacy[5]

1. Yale University, New Haven, Connecticut
2. Global Health Institute, School of Life Sciences, Ecole Polytechnique Federale de Lausanne, Lausanne, Switzerland
3. Center for Communicable Disease Dynamics, Harvard TH Chan School of Public Health, Boston, Massachusetts
4. Portland State University, Portland, Oregon
5. Centers for Disease Control and Prevention, San Juan, Puerto Rico

Prior to its full eradication in 1980 Smallpox was a highly infectious disease, with a mortality rate as high as 50% in some countries. In the US, historical records report a first introduction in Florida, in the 1890s, after which the epidemics spread around the country, until its official elimination in 1952.

Despite its historic ubiquity across the globe and concerns about bioterrorism, the dynamics of smallpox spread have not been extensively studied, with the exception of the United Kingdom.

Here we use a large dataset of historical smallpox cases, complied under Project Tycho, to assess the presence of spatio-temporal patterns in US smallpox outbreaks. We applied continuous wavelet transform and empirical mode decomposition analyses to time-series of historical smallpox cases aggregated at the county level from 1915 to 1950.

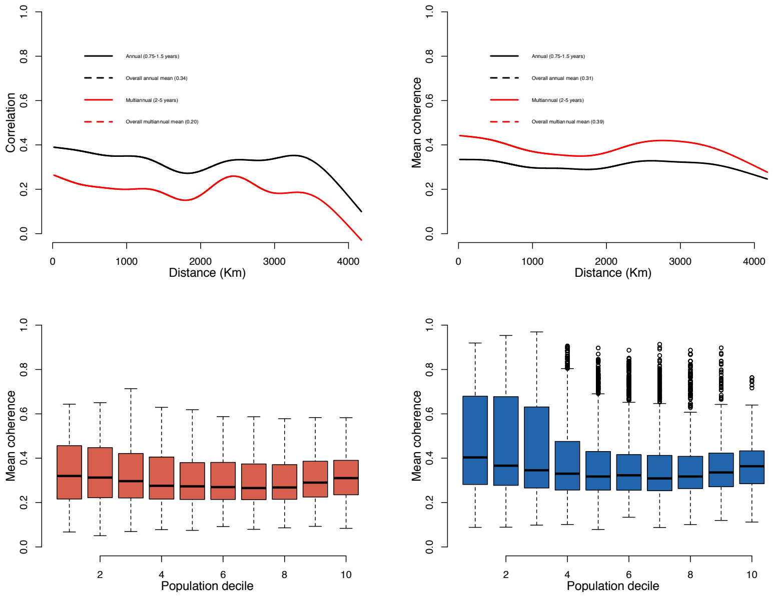


Figure 1 Synchrony in timing of smallpox cases by distance (Km) and by population deciles. Correlation is the pearson correlation coefficient between the wavelet coefficients of the county-pairs. Dashed lines refer to the overall mean correlation across all pairs. Coherence refers to the strength in frequency and timing of normalized bi-weekly cases between two counties. Population categories are the deciles of the population product of two counties. Red population colors refer to the annual and blue refers to the multi-annual periods.

We observed an overall level of synchronization of major smallpox epidemics across different US counties, and in particular during 1918, 1919, and 1921. Furthermore, we observed spatial dependence of coherence between epidemics in counties up to 2000 Km from one another. On average, the coherence of annual epidemics did not vary as population size changed. County-pairs with the smallest and largest populations had higher average coherence of major multiannual epidemics, though these differences were not significant. Further examination of the mechanisms driving observed epidemic synchrony is ongoing, and will include an examination of climatic, demographic, and immunologic factors. Our analysis will hopefully lead to a better understanding of smallpox transmission at smaller spatial scales, which will aid in future drug development and bioterrorism intervention and prevention strategies.

A Study on Public Transport Mobility Flows in Singapore

Sarah J. Berkemer[1, 2], **Shantal Lichang Cheong**[3], **Jared Edgerton**[4], **Marina Kogan**[5], **Ricky Laishram**[6], **Alan R. Pacheco**[7], **Alex Shannon**[8], **Matthew D. Sweitzer**[9], **and Xindi Wang**[10]

[1]Bioinformatics Group, Department of Computer Science, and Interdisciplinary Center for Bioinformatics, University of Leipzig, Härtelstraße 16-18, D-04107 Leipzig, Germany
[2]Max Planck Institute for Mathematics in the Sciences, Inselstraße 22, D-04103 Leipzig, Germany
[3]Strategic Planning Group, Urban Redevelopment Authority, Singapore
[4]Department of Political Science, The Ohio State University, Columbus, OH 43210, USA
[5]Computer Science Department, University of New Mexico, Albuquerque, NM 87131, USA
[6]Electrical Engineering and Computer Science, Syracuse University, Syracuse, NY 13210, USA
[7]Bioinformatics and Biological Design Center, Boston University, Boston, MA 02215, USA
[8]Center for Urban Science and Progress, New York University, New York, NY 10003, USA
[9]School of Communication, The Ohio State University, Columbus, OH 43210, USA
[10]Network Science Institute, Northeastern University, Boston, MA 02215, USA
+All authors contributed equally to this work

ABSTRACT

In the study of cities as complex systems within which many sub-systems operate and interact, the sub-system of people movement within a city is a key component in the activities and interaction of people and the urban environment. Understanding human mobility behavior within a city allows us to gain insights on human choice and community behavior. It has the potential to inform the work of urban planners in the planning of land use and transportation infrastructure to support mobility flows within a city.

In this study, a subset of data comprising of one day of public transport (PT) journeys from commuter card records from Singapore was analyzed to understand the PT mobility network structure (see Figure 1, left panel) and detect community clustering within the city-state at the broad level. Preliminary findings from assortativity analysis suggest key differences in travel patterns throughout the day, particularly during the lunch hour. Evidence of higher distance assortativity by PT mode (i.e. bus and train) was found for mobility flows during the lunch hour, while low distance assortativity by PT mode was found during other times of the day. Using a walktrap community detection algorithm, preliminary analysis suggests that the number of communities detected are highest during the morning and evening peak hours, with a slight increase in communities detected during the lunch hour.

An analysis of mobility patterns in relation to neighborhood characteristics was also conducted to better understand if different groups of the population from neighborhoods exhibiting varying spatial characteristics travel on the PT network differently. A community clustering study by Planning Areas (see Figure 1, right panel) found that the communities detected tended to be clustered along the train network, suggesting that mobility flows on the train network have a stronger influence on community clustering compared to mobility flows on the bus network. To understand if people are more likely to travel to a destination Planning Area that exhibits similar spatial and demographic characteristics as their origin area, a cosine distance similarity analysis was proposed to measure the trend in PT mobility flows between areas that are highly similar (including within-area travels) to highly different.

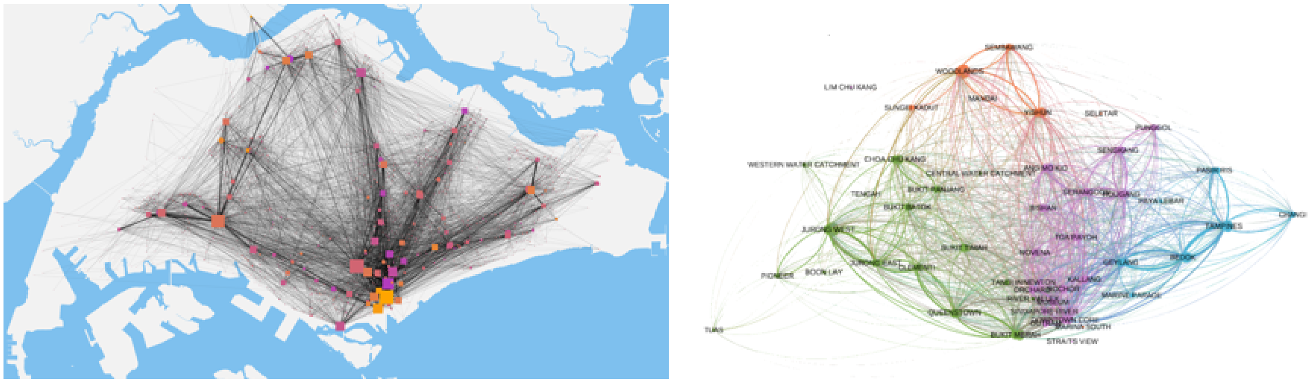


Figure 1. Islandwide network structure (full day). Node size and edge weight correspond to the number of travelers to a destination or along a route, respectively. Left panel shows the full network. Right panel shows the results of community detection methods.

Talking loud or talking often: How the interplay between network structure and agent influence affects the time to reach consensus in collective decision-making

Cedric Perret,[1] Guillaume St-Onge,[2] Zohar Neu,[3] and Andrea Santoro[4, 5]

[1] *School of Computing, Edinburgh Napier University, Edinburgh, EH10 5DT, United Kingdom*
[2] *Département de Physique, de Génie Physique, et d'Optique,*
Université Laval, Québec (Québec), Canada G1V 0A6
[3] *Department of Engineering Mathematics, University of Bristol, Woodland Road, Bristol, BS8 1UB*
[4] *School of Mathematical Sciences, Queen Mary University of London, Mile End Road, E1 4NS, London, UK*
[5] *The Alan Turing Institute, The British Library, London NW1 2DB, UK*

(Dated: November 2, 2018)

Consensus-seeking processes are core mechanisms for social organization in non-human animal groups, human societies, artificial and multi-agent systems. Such processes consist of a series of negotiations between individuals, which ultimately aggregate different individual opinions into one common agreed decision. Related works have shown that a small number of influential and highly centralized agents i.e. leaders, considerably speed up the consensus-seeking process. Yet the important features of hierarchy and leaders have not been clearly identified and their investigation in experiments provide contradictory evidence. To disentangle these effects, we separate two features of hierarchy (i) the level of centralization ν and (ii) the number of leaders N_L, i.e. influential individuals. We therefore developed a continuous opinion formation model to explore the effect of each feature and their interplay on the time to reach a global consensus. The results show that the most important feature is the topology of the network structure, with higher centralization always decreasing the time to reach consensus. In addition, it also demonstrates that the effect of influential individuals are deeply entangled with the underlying network structure. Indeed, it can be beneficial to have a few leaders if the structure is centralized, but surprisingly it can also be detrimental if the population is homogeneous. Our results provide a mechanistic explanation of the "babble" hypothesis which stresses the importance of talkativeness and extraversion of leaders. Further directions involve a full characterization of the main mechanisms leading to polarization or fragmentation as final states in more complex network structures, such as modular structures with nested hierarchies.

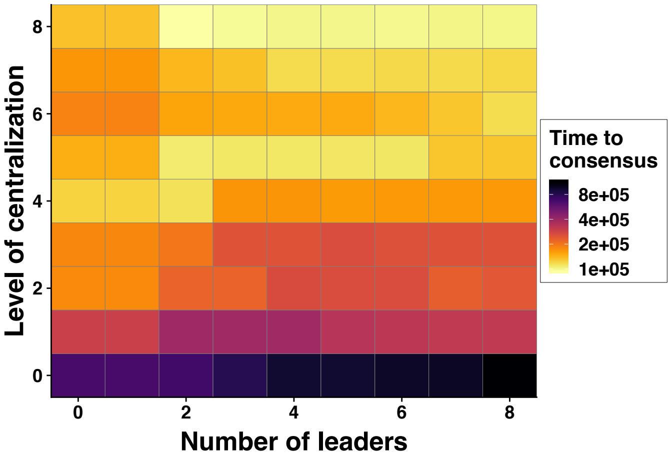


FIG. 1. Time to reach consensus τ as a function of the number of leaders N_L in the network and the level of centralization ν, which dictates the heterogeneity of the network. Simulations are averaged over 200 realizations, with networks having $N = 200$ node, an average degree $\langle k \rangle = 4$ and influence parameters $\alpha_L = 0.9$, $\alpha_F = 0.1$ for the leaders and the followers respectively.

Thresholding normally distributed data creates complex networks

George Cantwell[1], **Yanchen Liu**[2], **Benjamin Maier**[3], **Carlos A. Serván**[4], **Alice C. Schwarze**[5], **Jordan Snyder**[6], **and Guillaume St-Onge**[7]

[1]Department of Physics, University of Michigan, Ann Arbor, Michigan, USA
[2]Center for Complex Network Research, Northeastern University, Boston, Massachusetts, USA
[3]Robert Koch Institute, Nordufer 20, D-13353 Berlin, Germany
[4]Department of Ecology and Evolution, University of Chicago, Chicago, Illinois, USA
[5]Mathematical Institute, University of Oxford, Oxford OX2 6GG, United Kingdom
[6]Department of Mathematics, University of California, Davis, California, USA
[7]Département de Physique, de Génie Physique, et d'Optique, Université Laval, Québec, Canada

ABSTRACT

Network data sets are often constructed by some kind of thresholding procedure. The resulting networks frequently possess properties such as heavy-tailed degree distributions, non-zero clustering, large connected components and short average shortest path lengths. These properties are considered typical of complex networks and appear in many contexts, prompting consideration of their universality. Here we introduce a very simple model for continuous valued relational data and study the effect of thresholding it. We find that some, but not all, of the properties associated with complex networks can be seen after thresholding, even when the underlying data is not "complex".

We examine the properties of networks created by thresholding relational data. To do this we introduce a basic null model of the underlying data, which is then thresholded. The model is derived from three assumptions: (1) all nodes are statistically identical; (2) any correlations are local; and (3) the underlying data is normally distributed. If there are n nodes, the underlying data will be an $n \times n$ symmetric matrix, $\boldsymbol{X}$. The listed assumptions lead to a simple one parameter normal model for the data:

$$\boldsymbol{X} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}), \qquad \text{where,} \qquad \Sigma_{(i,j),(i,j)} = 1; \qquad \Sigma_{(i,j),(i,k)} = \rho; \qquad \Sigma_{(i,j),(k,l)} = 0. \tag{1}$$

Networks are created by thresholding $\boldsymbol{X}$. For each pair of nodes i, j, if $X_{ij} \geq t$ we say there is an edge. Despite the fact that the underlying data is fully connected and normally distributed (a situation not usually considered "complex") we find the thresholded networks display some of the properties typically ascribed to complex networks. As an example, below we show some degree distributions for real-world networks along with the normal model predictions.

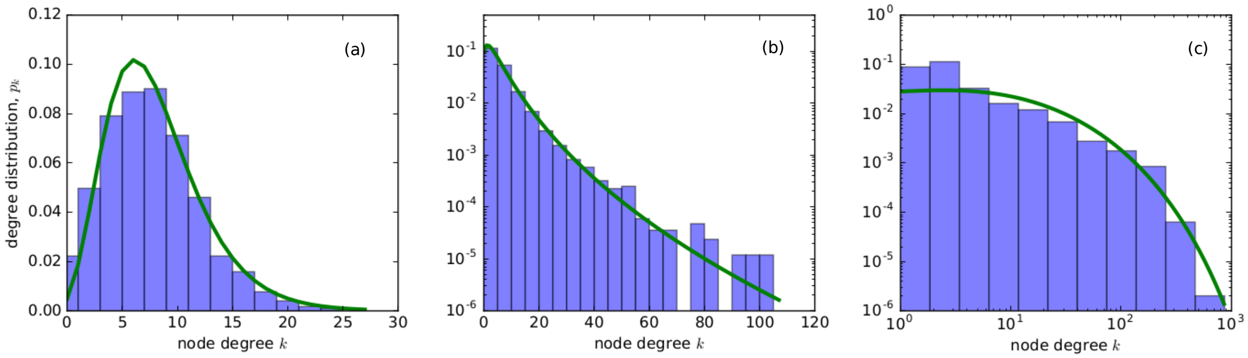


Figure 1. Degree histograms for three real-world networks along with fitted distributions from the threshold model. We show (a) A high school friendship network, (b) a co-authorship network between scientists, and (c) a protein–protein interaction network. These networks were chosen for their different degree distributions – note the different scales on the axes: linear, log-linear, and log-log.

Translating from nature to technology:
Framework to compare innovation pathways in biomimicry

Magdalena M. Klemun[a], Sarah Berkemer[b], Alexandra Mikhailova[c]
[a]Massachusetts Institute of Technology, Cambridge, Massachusetts, USA
[b]University of Leipzig, Leipzig, Germany
[c]University of California, Davis, California, USA

Biomimicry is a growing area of research that seeks to abstract functions and architectures observed in nature to solve problems in science and engineering. Previous studies have documented biomimicry efforts in individual subfields, and discussed the philosophical foundations of biomimicry as a sustainability-oriented innovation strategy. Less work has been done, however, on describing the individual steps of abstraction taken in the course of biomimicry innovation efforts. Studying this question can add to an understanding of how the process of replicating or mimicking nature works in practice, and whether successful outcomes result from similar inputs.

Here we begin to address this gap by compiling a database of the most widely discussed biomimicry applications. We use this database to develop a conceptual framework to describe the process through which biological phenomena have been translated into technologies (Fig. 1). We use the term 'translate' to indicate that there are different ways to start from an observation of nature and end with a technology, just like there are multiple ways to translate words from one language to another.

We find that a variety of biomimicry pathways have led to inventions and commercial products, ranging from those where the biological phenomenon is partially or fully explained and mechanisms or entire architectures can be replicated through engineering design (mechanistic pathway), to those where the biological phenomenon is observed and mimicked, but may not be fully understood (phenomenological pathway). We also classify biomimicry applications along other dimensions, including whether the translation process encompasses changes in the spatial scales or the media in which mechanisms operate. The goal of this work is to take stock of past biomimicry efforts in a way that identifies common and differentiating characteristics of biomimicry efforts.

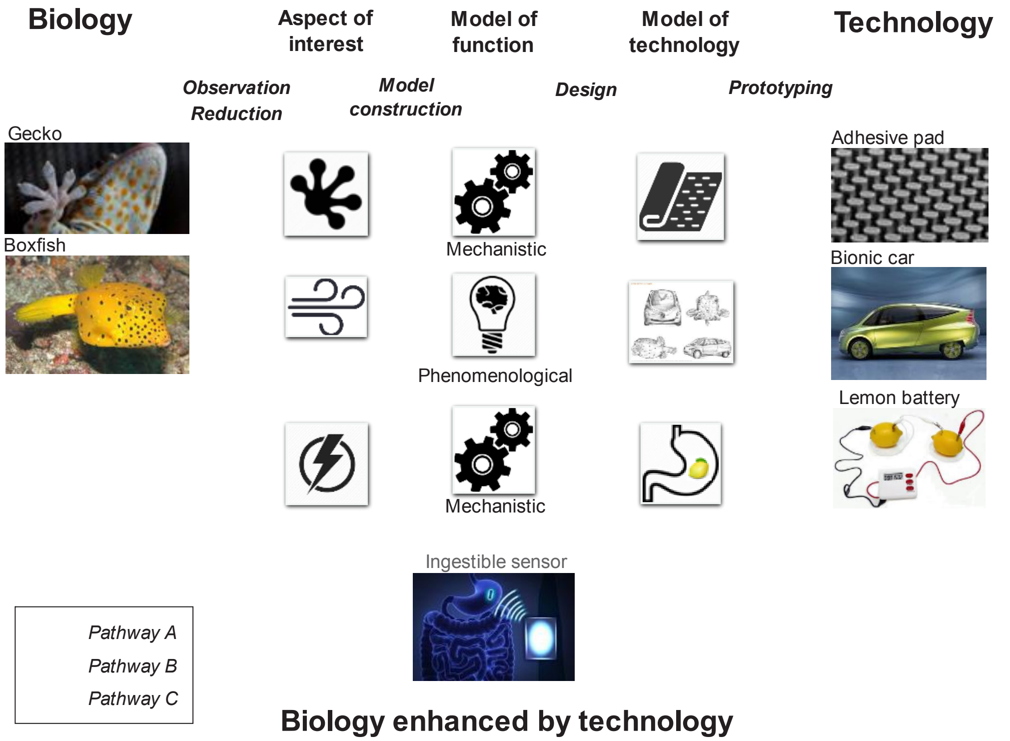


Figure 1: Pathways to translate functions observed in nature (left) into technological applications (right), or vice versa. Observation and reduction, model construction, design, and prototyping are common to all three pathways. Design may involve changing the context of a biological function (spatial and temporal scale, media such as water, air) to match the engineering problem at hand. Pathway A (blue) involves an understanding of the mechanisms governing a biological function (mechanistic model), while pathway B (red) draws on interpretations of why functions occur to inspire designs (phenomenological model). Pathway C (grey) involves the use of technologies in interaction with biology, for instance in medicine, and therefore starts off in the opposite direction of A and B. Our examples draw on innovations in materials science (pathway A: adhesive materials replicating van der Waals forces governing Gecko foot hair), automotive design (pathway B: Bionic car modeled after boxfish), and biomedicine (pathway C: long-lived, low-

Understanding Music with Higher Order Networks

Josefine Bohr Brask[1], **Ricky Laishram**[2], **Carlos Marcelo**[3], **and Xindi Wang**[4]

[1]University of Exeter, United Kingdom (j.bohr-brask2@exeter.ac.uk)
[2]Syracuse University, Syracuse, NY, USA (rlaishra@syr.edu)
[3]University of Chicago, IL, USA (cmarceloservan@uchicago.edu)
[4]Northeastern University, Boston, MA, USA (wang.xind@husky.neu.edu)

ABSTRACT

Music is a natural complex system: notes interact in a sequential way to create melody and interact in a “vertical" way with different instruments to create harmony; only with 12 notes, composers are able to create different styles and music genres. Simple networks have been used to model and analyze music. However, this representation assumes there are no dependencies between notes. To include these dependencies, we model musical pieces as a *Higher Order Network (HON)*. In a higher order network, the dependencies between the nodes are encoded by using sequence of dependent notes as nodes.

We created two sets of higher order networks: HON for music pieces (HON-piece) and HON for music genres (HON-genre). In HON-piece, each musical pieces are represented by a HON; and in HON-genre, we aggregate all the trajectories from musical pieces of the same genre and represent it by a HON. These representations are complementary to each other and serve different purposes. In Figure 1 we show an example of HON-piece.

To demonstrate the utilities of HONs, we proposed various features extracted from the HON-piece and observed that these features align with the common perception of music genres. Additionally, we used the features extracted from HON-piece and random walk on HON-genre to classify music genres.

By using the features from HON-piece, we achieved a classification accuracy of up to 77.7% in Classical vs Folk, and 81.9% in Rock vs Jazz. The genre classification based on random walk on HON-genre is less accurate: 64.7% and 60.9% for Classical vs Folk and Rock vs Jazz respectively. This result indicates that the features that we proposed are indeed able to capture the characteristics that differentiates music of different genre.

To further understand the relationship between HON and music pieces, we conducted a case study on *Blackbird* by *The Beatles*. We identified different parts of the network that corresponded to the beginning of verses, the rest of the verses, and the B-section. This shows the potential of using HONs to analyze the structure of the music piece.

We also investigated the spectral properties of the transition matrix of HONs. We found that the eigenvalue distributions are qualitatively different between different music pieces and have the potential to capture characteristics of music pieces.

There are various avenues for future works in this area, including: 1) incorporating temporal information which is a vital property for music; 2) incorporating multiple instruments for a music piece since our current approach ignores the "vertical" interactions. One potential way to incorporate that is using multilayer networks to represent different instruments; 3) exploring different coding methods such as pitch difference coding.

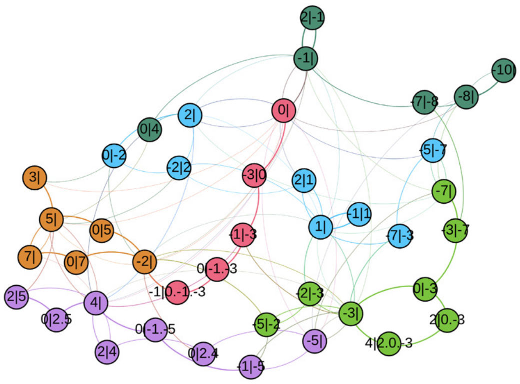


Figure 1. *Bach Prelude from Sonata N6* as Higher Order Network

SANTA FE INSTITUTE

Searching for Order in the Complexity of Evolving Worlds

www.santafe.edu