Citation for published version:
Room, G, Spence, A, Evangelou, E, Zeppini, P, Pugliese, E & Napolitano, L 2017, *Dynamics of cumulative innovation in complex social systems (DCICSS)*. University of Bath.

Publication date:
2017

Document Version
Publisher's PDF, also known as Version of record

Link to publication

University of Bath

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Download date: 17. Nov. 2019

Institute for Mathematical Innovation | UNIVERSITY OF BATH

DCICSS/2017/1

DYNAMICS OF CUMULATIVE INNOVATION IN COMPLEX SOCIAL SYSTEMS (DCICSS)

PROJECT OVERVIEW

11 April 2017

KEYWORDS

- Dynamic network analysis
- From Auto-Catalytic Networks to Auto-Catalytic Societies
- Cumulative causation in social processes
- Empirically validated mathematical and statistical models
- From molecular networks to
    - Technology ecosystem of patents data
    - Financial networks and risk dynamics
    - Social digital networks and information flows
    - Mergers and acquisition and industrial districts

SUMMARY

Social and economic change often involves self-reinforcing processes and dynamic synergies. Think for example of local clusters of economic development, such as Silicon Valley or the proposed 'Northern Powerhouse', where the UK government seeks to promote a self-sustaining process of transformation. How we understand these dynamics raises fundamental questions for social science and public policy.

The existing literature relies mainly on qualitative methods plus standard regression models. This misses the dynamic interactions involved in such processes of transformation. In critical opposition to these, we bring together complexity theory and political economy, in order to analyse such self-reinforcing dynamics within different institutional settings.

We are a multi-disciplinary team of mathematicians and social scientists. We develop an approach in terms of co-evolving networked systems: modelled in mathematically precise terms and applied to empirical data sets. We investigate how such mathematical models of non-linear systems can illuminate processes of cumulative innovation and dynamic synergy in the social world. We develop statistical tools for investigating empirical data sets, so as to test competing theoretical accounts. We ask how policy makers can intervene in complex social and economic dynamics: and we expect to provide tools for evaluating policy interventions.

We take a series of empirical case studies, to develop and test this approach.

1. PROJECT OVERVIEW

We start from the research questions: how can we conceptualise and model self-reinforcing processes of social and economic change and how can we evaluate corresponding policy interventions? We address these questions by deploying mathematical models of non-linear systems and applying them to a range of well-chosen empirical data sets, using carefully crafted statistical tools.

Jain and Krishna (2003) (http://arxiv.org/PS_cache/nlin/pdf/0210/0210070v1.pdf) provide our starting point. They are interested in the dynamics of co-evolution that played a key role in the origins of life. They use network theory to explore autocatalytic sets (ACS): sets of simple molecular organisms, none able individually to self-replicate, but each providing a catalyst for its fellows: a process of symbiotic boot-strapping for collective self-replication.

Jain and Krishna (2003) imagine populations located at the different nodes of the network, with each population dependent on the rates of population change at other specified nodes. The populations of nodes which are part of an ACS then enjoy particularly rapid growth. In just the same way, insects and flowering plants have co-evolved, thriving disproportionately by virtue of their mutual adaptations. Jain and Krishna use computational models to simulate such a dynamic system. This reveals 'punctuated equilibria', with processes of growth and then partial collapse.

One interesting feature of the Jain and Krishna model is the two evolutionary scales. The fast (or short) dynamics correspond to the fitness propagation within the network, and the slow (or long) dynamics correspond to the update of the network. These features have interpretations in social networks. For example, the fast and slow dynamics can be interpreted respectively as local innovation (transfer of knowledge between similar technologies) and global innovation (updates to the global technology system).

This provides our distinctive focus and vantage point, for studying self-reinforcing processes of social and economic change. We do so across a range of diverse empirical problems; we develop ground-breaking **mathematical variations on the basic Jain and Krishna model,** appropriate to the various case studies; and we link our empirical data analysis to the **theoretical economic and social science literature on cumulative causation and innovation**.

The Jain and Krishna model has attracted a good amount of attention in the literature on complex network dynamics: and they have themselves discussed its potential application to patterns of technological innovation. Taking their model as our starting point, in relation to processes of economic and social change, seems therefore well-judged. Nevertheless, this is to our knowledge the first and most ambitious attempt to apply their model on a multi-disciplinary basis and with a strong empirical and policy emphasis.

Patents: This will be the initial empirical case study. We have in-house expertise and ready access to the datasets in question.

Using the case study of patents as our template, we will extend the research across further social and economic case studies, with a view to expanding the mathematical, statistical and empirical scope of the Jain and Krishna model. These additional case studies are likely to include some of the following:

Mergers and Acquisitions: to track the development and co-evolution of different capabilities in firms.

Financial system: to model the co-evolution of risks within connected financial systems.

Welfare regimes: to model how institutions concerned with social security, vocational training, employment, etc co-evolve to produce a number of distinct regimes typical of particular countries.

Developing countries: linking science and engineering innovations with the institutional contexts and 'soft technologies' of the developing world.

Digital social networks: examining the dynamics of social media, and the insights offered by our modelling for data analytics and digital governance.

2. EMPIRICAL CASE STUDY: PATENTS

Technological innovation is a key part of much economic and social change. If we want to track processes of technological innovation, one established method for doing so is to use patent data. We will adopt this approach, but using the novel approaches sketched above.

We will therefore use patent databases to track the co-evolution of different technologies and of the knowledge which they embody. Patents are registered within one or more classes of the overall patent registration system. These classes form the **nodes** of our network: they represent **different areas of knowledge**. New patents when registered cite earlier patents on which they draw: commonly these are drawn from a variety of existing classes. They thereby establish links between classes. These links (aggregated across many new patents) define the directed **edges** of our network: they represent the **transfer of knowledge**.

We model and analyse the co-evolution of patent classes. We also study the reconfigurations of the patent classification system that occur at various scales and their relationship to those co-evolutionary dynamics. Finally, by making comparisons between patterns of development in different national contexts, and by appreciation of different institutional configurations, we expect to derive policy guidance relevant to the improvement of national innovation performance.

We work with PATSTAT, a database which records the *International Patent Classification* codes used by patent offices worldwide. Within the patent application process, a patent is assigned to one or more technology classes on the basis of its end-use and its scientific or technological area. The classification system is a hierarchical one, with six scales of granularity (from 8 main 'sections' up to about 70000 'subgroups'). This will allow us to work at different levels of detail of the technology ecosystem, to study how properties 'scale' across classifications, and to detect the aggregation scale where they emerge.

Over time, some technologies disappear and new ones show up. The decline and termination of patenting activity in a class is a good proxy for this event. But more importantly, patent databases undergo *re-classification events*. Whenever a patent application does not fit any existing class, a new class is introduced, and some of previously granted patents are re-classified. These events pose empirical challenges for the researcher: but they also point to more or less dramatic shifts in the knowledge ecosystem.

Two stages of innovation are represented in the patents network, incremental and radical. Incremental innovation corresponds to the growth of the classes, while radical innovation corresponds to the re-classification. It is by analysing the history of the patents network that we can identify and distinguish these two stages.

3. MATHEMATICAL DEVELOPMENT

In order to illuminate the case studies, we explore a range of variations in the mathematics of the Jain and Krishna model. We expect thereby to address some intriguing and innovative mathematical questions.

We will take the basic Jain and Krishna template and recalibrate it to fit better the specific case studies. This might mean, for example, that variable edge weights are required in the network, or that negative edge weights (corresponding to a node inhibiting growth of another node) need to be introduced. This latter situation introduces several open mathematical questions. Also, how node importance (centrality) is measured will vary from case study to case study.

One feature of the Jain and Krishna model that distinguishes it from much of the work on evolving networks is that it identifies "weakest" nodes and rewires the network accordingly. This is natural in their application to the evolution of self-replicating sets of chemical species, but is quite unusual in comparison with other probabilistic models for evolving networks. In order to distinguish between important and weak nodes Jain and Krishna need to calculate the corresponding components of an eigenvector of a certain matrix (similar to the way Google ranks important webpages). The efficient calculation of eigenvectors of large matrices is an important problem in scientific computation, and recent randomized algorithms for the determination of a small number of dominant eigenvalues of a large matrix provide a technique to significantly reduce computational costs. The application of these techniques for analysis of evolving networks seems to be novel and has wide applicability across several different application areas.

The Jain and Krishna model for an evolving network can be thought of in the following terms. A sequence of networks behaves as a Markov chain where the probability distribution of the network at time t is specified conditional on the network at the previous time in terms of a rewiring probability p. In practice, p is an unknown parameter of the model and for application to a real data-set, p would have to be estimated. To do this, one would construct a likelihood function for the parameter p and use this to best estimate p. This approach is central to statistical inference and forms the basis of many inferential methods such as maximum likelihood and Bayesian estimation. We have explored the application of maximum likelihood estimation to simulated data from the Jain and Krishna model with satisfactory results. With the calibrated models, evaluation of the likelihood becomes computationally challenging because of the increasing complexity introduced by variable weights. In this case and approximate likelihood evaluation will be considered.

4. STATISTICAL INDICATORS

We search for statistical tools with which to probe datasets – tools which have not previously been exploited for this sort of dynamic analysis.

An important choice concerns the level of data aggregation. In the case of patents, we know that patent classes are divided into subclasses, sub-subclasses and so on down to the finer level of so-called 'groups'. Working in the finest level of granularity may not be the best option because there is too much detail, which is difficult and unnecessary to model. On the other hand, data at the highest level contain less information. It is important to test our network models at different levels of granularity to obtain the best goodness-of-fit of our model.

Even when a model is available for studying the network of patent technology classes, we still need to check that it is an accurate representation of the underlying data-generation process. We will consider structural descriptors of the network, starting with conventional metrics such as connectivity and centrality. Beyond that however, we will look at more innovative concepts of network measures that focus on cumulative causation mechanisms: in particular, the distributions across the network of the number and size of cycles and the distribution of the number and size of the Autocatalytic Sets (ACSs) in the network.

We will need to assess goodness-of-fit: examining how well the data are represented by the model. This is imperative for justifying and strengthening our conclusions. Because of the complexity of the model, the distribution of many useful structural descriptors cannot be deduced in closed form. Instead, we will compute these using simulations from the fitted model and compare them against the same statistics computed from the data, using a hypothesis test procedure.

5. POLICY GUIDELINES

Darwinian evolution is a blind process. When however we apply evolutionary models to the social world, there is a place for purposeful activity by social agents. Within the various case studies we expect to study, these agents include firms, local communities and public authorities: intervening at many different scales, in an effort to promote the development and direction of dynamic synergies and autocatalytic sets.

We expect our modelling and our empirical case studies to illuminate such interventions. We give particular attention to public policy makers, seeking to nurture the self-reinforcing dynamics of social and economic development: we also however consider the role of private power, the corporations which also intervene in such transformative processes. Sometimes they may come to some agreement – sometimes they are involved in a long struggle, involving a mixture of bullying, bribing, bargaining and disrupting.

Within each case study, we aim to interpret the parameters of the model by reference to the range of possible policy interventions. Thus within our first case study we can examine the rate of growth of patents as a measure of incremental innovation: and consider how this varies as between different regional and national innovation systems. Equally however, we can use the reclassification events of patent systems as a measure of more radical innovation.

6. POTENTIAL BENEFITS TO SOCIETY

It is important, both for scholars and for the general public, to recognise the dynamic inter-connections among different elements of our social and economic world and the ways in which they co-evolve. It is also therefore of benefit to have models and methods of analysis that are rooted in that vision, rather than approaches that tend to analyse each element in isolation.

This is also true of the dynamic and complex world with which policy makers are faced: the policy makers of the public realm but also the corporate policy makers of the business world. It does not for example make sense to examine particular policy interventions in isolation and to hope to establish 'what works' without reference to their co-evolving contexts. If policy makers are sensitive to such co-evolutionary dynamics, they can use them to steer and accelerate the effects of their interventions.

Our initial case study is concerned with technological change. The research should shed light on the process of technological change: both in terms of incremental innovation rates, and in terms of radical shifts. It will help uncover the structure of technology as an ecosystem, and be able to detect 'hot' areas in terms of their linkages to other areas. This systemic approach, based on a structured empirical technology map will be turned into usable tools for whoever deals with technology. In particular, we envisage benefits for policy makers who aim at improving the economic competitiveness in a region, or directing resources optimally towards sustainable sectors. There might even be insights for private enterprises, seeking advice on their investments, and for venture capitalists.

Research Team at the University of Bath:

Professor Graham Room
Professor of European Social Policy
http://www.bath.ac.uk/sps/staff/graham-room/

Professor Alastair Spence
Professor in Numerical Analysis
http://www.bath.ac.uk/math-sci/contacts/academics/alastair-spence/

Dr Evangelos Evangelou
Lecturer in Statistics
www.bath.ac.uk/~ee224

Dr Paolo Zeppini
Lecturer in Environmental Economics
http://www.bath.ac.uk/economics/staff/paolo-zeppini/

Research Associates:

Dr Emanuele Pugliese
Research Fellow at the Institute for Complex Systems, National Research Council (Italy)

Dr Lorenzo Napolitano
Post-doctoral Fellow, Institute for Complex Systems, National Research Council (Italy)

http://pilhd.phys.uniroma1.it/PILgroup_Economic_Complexity/Home.html

Acknowledgements

We acknowledge gratefully the assistance kindly provided by PATSTAT at the European Patent Office, in supplying the datasets we are using for our patent case study.
http://www.epo.org/searching-for-patents/business/patstat.html#tab1