# A Complex System Approach to Social Studies

**Research Statement**

20 December 2019

**Taha Yasseri**
*Senior Research Fellow in Computational Social Science*
Oxford Internet Institute
University of Oxford
1 St Giles', OX1 3JS Oxford, UK
and
*Turing Fellow,* Alan Turing Institute, London, UK
e-mail: taha.yasseri@oii.ox.ac.uk
web: www.oii.ox.ac.uk/people/yasseri

## Abstract

I have been trained as a physicist with a PhD in the Physics of Complex Systems from the University of Göttingen, Germany. Even though my fascination with social systems and the urge to study them in a social science department brought me to the Oxford Internet Institute, I have remained committed to the "scientific methods", in that I follow the complete cycle of observation (using descriptive statistical analysis), theory building and hypothesis generation (through literature review and theoretical analysis), and experimentation and theory testing (through lab and online experiments, data mining, agent-based modelling, and machine learning).

Even though observation and experimentation become immensely complicated as soon as the subject of the study changes from natural phenomena to social phenomena and humans instead of particles, the emergence of digital technologies and their ever-growing presence in our daily lives in the past two decades have made it possible to take such an approach to empirical social science research.

Over the past 10 years, after completion of my PhD, I have been applying this philosophy to a number of social phenomena in different domains leading to publication of peer-reviewed articles in internationally known journals, obtaining research grants from national and international funding agencies as well as industries, and presentation of my work at international conferences. Moreover, I have had regular appearance in public media as I strongly believe that a research project cannot be considered successful if it is not feeding back to the public that funds it in the first place.

## Research strands:

### 1. Information Dynamics and Social Contagion: A Case of Social Complexity

In the past years, I have studied the dynamics and spread of information and behaviour, social contagion, in different settings and cases. My work in this area started with prediction of movies box office revenue based on information seeking patterns of online movie fans [1]. Building on that, I studied political information seeking and electoral dynamics and investigated the possibilities of predicting elections based on different sources of data from the social web [2,3]. In a different domain, I studied how information exchange between Massive Open Online Course takers is limited to the modular network structure of interactions and hence questioned the notion of MOOCs as "global classrooms" that was popular in early 2010's [4,5]. Later, I applied the models of social contagion to online music listenership data to illustrate that social contagion is a complex phenomenon. My work has demonstrated that not only the network structure matters in determining the parameters of contagion of the social behaviour under study, but also the content and the type of behaviour which are often neglected in simplified modelling are important in determining how certain behaviours spread in social networks [6].

***Complex Systems Approach to Social Behaviour of Machines***
Artificial Intelligence (AI) has become the buzz word in academia, industry, and even public sector. While there is a surge in AI research that examines various aspects of machine intelligence, there has been little discussion on collective behaviour of machines. The very core idea behind complex systems theory is that the emergent behaviour of a complex system can be different and difficult to predict based on the linear sum of its parts. We already have built systems in which multiple automated agents interact with each other: for example, the editing bots in Wikipedia that simultaneously revise the articles in the encyclopaedia. Our past research on Wikipedia bots demonstrated that even though the bots share the same goal that is to improve the encyclopaedia and they have a very low level of intelligence, still conflict and disorganization are often inevitable [7], (mostly because they learn from humans). Extending on this work, I will be further examining how social behaviour of machines can be studied, designed to be ideally controlled to avoid conflicts and disagreements that are common in human social systems.

## 2. Online Collaboration and Crowdsourcing: Agent-based Modelling Meets Data

Wikipedia is the most successful example of an internet-based collaborative environment, based on the idea of collecting pre-existing knowledge from secondary sources. My previous studies have shown that contributions to Wiki-based projects are heterogeneously distributed in temporal patterns and across users. Whilst most of the users contribute very little, a few power-users produce a large share of the content [8]. However, all users contribute in bursts and have extremely irregular behaviour, modulated by daily and weekly patterns rooted in their individual socio-ecological diurnal habits and cultural differences [9]. I have studied the dynamics of conflict and edit wars among Wikipedia users [10] and how they differ in different language editions [11]. I have used agent-based modelling to study opinion dynamics among editors of Wikipedia to understand and explain the edit wars [12,13] and to study the role of extreme opinions among the groups and how to cope with them [14].

***AI Assisted Collective Intelligence***
Collective intelligence that emerges from the collaboration, collective efforts, and competition of many individuals was first illustrated in Galton’s famous Wisdom of the Crowd experiment. Since then there have been numerous examples of successful deployment of collective intelligence and crowed-sourcing to solve different problems that have been unsolvable to individuals or for machines. Prominent example of such projects are Wikipedia and citizen science platforms such as Galaxy Zoo and FoldIt. However, with the surge of AI and its numerous applications in decision making and forecasting, it is important to examine how AI can be used to improve the state of the art in crowed sourcing and collective intelligence. AI, and in particular Machine Learning can be used in assigning fitness scores to different members of the crowd in combining their contributions to the collective decision. In addition, AI can be used in training individual members of the crowd, and finally the crowd intelligence can be used to advance the learning by machines in an environment in which both humans and machines take part in creating collective intelligence. I will be examining these possibilities through a series of case studies that are concerned with each of these three elements.

## 3. Political Behaviour and Data Driven Political Since: Political Turbulence

As people go about their daily lives using social media, such as Twitter and Facebook, they are invited to support myriad political causes by sharing, liking, endorsing, viewing and following. Chain reactions caused by these tiny acts of participation form a growing part of collective action today, from neighbourhood campaigns to global political movements [15]. In previous work I have shown how most attempts at collective action online fail [16]. Those that succeed can do so dramatically, but are unpredictable, unstable, and often unsustainable [17].
I have used supervised machine learning to detect and identify Islamophobic behaviour on Twitter in order to study its social dynamics [18]. Additionally, I have analysed the content of petitions submitted

to the UK's government petitioning website using unsupervised machine learning to detect issues and topics that the British citizens are concerned about the most in the time of uncertainty [19].

***Data-driven Approach to Opinion Dynamics and Representative Democracy***

One of the main challenges in representative democracy is the divergence between decision makers and the people that they represent. The divergence can exist both in the agenda, and in the opinion on a specific issue. Simply put, what matters to policy makers might be different to what matters to the citizen. Even when they agree on what matters, the representatives might have different opinion to the opinion of people who elected them.

In modern days expression of opinion by citizens is much more common due to the freedom of speech and public media. In the age of the Internet, when people produce more than 2.5 quintillion of data each day, this divergence, that some argue has grown even in established democracies, seems paradoxical.

The distance between policy makers and citizens is practically non-existent online, yet we see a huge divide between main-stream parties and citizens' will. I will extend my previous work on methodologies to mine the issues and public opinions on them by analysing textual data generated on Internet-based platforms and to develop computational method to measure the divergence between public opinion and representative's opinion on the issues of mutual concern using clustering techniques.

## 4. Online and Mobile Dating: a Moving Target

Online technologies have entered our love lives no less than any other aspects of our lives. When it comes to gender differences, the initial promise of online dating is a more balanced mating practice in which male and female users have the same level of power and agency. However, there are arguments that gender asymmetries in mating are deeply rooted in humans' individual and social traits and it is less driven by the medium in which mating is practiced. My previous research on eHarmony users suggests that women put more emphasis on income, education, and religion, while men put more emphasis on age and physical health of their potential partner [20]. My previous study on the communication patterns in mobile dating has suggested differences in the way males and females converse on the dating platforms [21]. In a different study, I have reported that that the gender gap in initiating a conversation in an online dating app has been increasing over the years and men are 5 times more likely to initiate conversations in 2018 compared with 2008 [20]. The gender gap further widens when it comes to paid sex market, as I have shown here [22].

***Gender and Sexuality in the Internet Age***

Gender differences in human sexual behaviour have been reported extensively; compared to women, men masturbate more, use more pornography, are more permissive and more reactive to visual cues, and experience sexual desire more spontaneously. Today, considerable amount of our sexual behaviour happens online. The internet-based technologies not only affect and possibly change our approach to sexuality, they also produce massive amount of transactional data (aka big data) that can be studied in the framework of computational social science.

Some of the research questions regarding gender differences in sexual behaviour could be addressed in an unprecedented way by using the big data generated on online platforms. I will continue my research on this topic based on the collaborations I have set up over years with different industry partners and the datasets that I have generated by scraping online dating/sex platforms.

## References:

[1] Mestyán, M., **Yasseri, T**., & Kertész, J. (2013). Early prediction of movie box office success based on Wikipedia activity big data. *PloS one*, *8*(8), e71226.
[2] **Yasseri, T**., & Bright, J. (2014). Can electoral popularity be predicted using socially generated big data?. *it-Information Technology*, *56*(5), 246-253.
[3] **Yasseri, T**., & Bright, J. (2016). Wikipedia traffic data and electoral prediction: towards theoretically informed models. *EPJ Data Science*, *5*(1), 22.

[4] Gillani, N., **Yasseri, T**., Eynon, R., & Hjorth, I. (2014). Structural limitations of learning in a crowd: communication vulnerability and information diffusion in MOOCs. *Scientific reports*, *4*, 6447.
[5] Eynon, R., Hjorth, I., Gillani, N., & **Yasseri, T**. (2014). 'Vote me up if you like my ideas!'experiences of learning in a MOOC. *Experiences of Learning in a MOOC (May 21, 2014)*.
[6] Ternovski, J., & **Yasseri, T**. (2017). Social Complex Contagion in Music Listenership: A Natural Experiment with 1.3 Million Participants. *arXiv preprint arXiv:1711.05701*.
[7] Tsvetkova, M., García-Gavilanes, R., Floridi, L., & **Yasseri, T**. (2017). Even good bots fight: The case of Wikipedia. *PloS one*, *12*(2), e0171774.
[8] **Yasseri, T**., & Kertész, J. (2013). Value production in a collaborative environment. *Journal of Statistical Physics*, *151*(3-4), 414-439.
[9] **Yasseri, T**., Sumi, R., & Kertész, J. (2012). Circadian patterns of wikipedia editorial activity: A demographic analysis. *PloS one*, *7*(1), e30091.
[10] **Yasseri, T**., Sumi, R., Rung, A., Kornai, A., & Kertész, J. (2012). Dynamics of conflicts in Wikipedia. *PloS one*, *7*(6), e38869.
[11] **Yasseri, T**., Spoerri, A., Graham, M., & Kertész, J. (2014). The Most Controversial Topics in Wikipedia. *Global Wikipedia: International and cross-cultural issues in online collaboration*, *25*.
[12] Török, J., Iniguez, G., **Yasseri, T**., San Miguel, M., Kaski, K., & Kertész, J. (2013). Opinions, conflicts, and consensus: modeling social dynamics in a collaborative environment. *Physical review letters*, *110*(8), 088701.
[13] Iñiguez, G., Török, J., **Yasseri, T**., Kaski, K., & Kertész, J. (2014). Modeling social dynamics in a collaborative environment. *EPJ Data Science*, *3*(1), 7.
[14] Rudas, C., Surányi, O., **Yasseri, T**., & Török, J. (2017). Understanding and coping with extremism in an online collaborative environment: A data-driven modeling. *PloS one*, *12*(3), e0173561.
[15] Margetts, H., John, P., Hale, S., & **Yasseri, T**. (2015). *Political turbulence: How social media shape collective action*. Princeton University Press.
[16] Hale, S. A., Margetts, H., & **Yasseri, T**. (2013, May). Petition growth and success rates on the UK No. 10 Downing Street website. In *Proceedings of the 5th annual ACM web science conference* (pp. 132-138). ACM.
[17] **Yasseri, T**., Hale, S. A., & Margetts, H. Z. (2017). Rapid rise and decay in petition signing. *EPJ Data Science*, *6*(1), 20.
[18] Vidgen, B., & **Yasseri, T**. (2018). Detecting weak and strong Islamophobic hate speech on social media. *arXiv preprint arXiv:1812.10400*.
[19] Vidgen, B., & **Yasseri, T**. (2019). What, When and Where of petitions submitted to the UK Government during a time of chaos. *arXiv preprint arXiv:1907.01536*.
[20] Dinh, R., Gildersleve, P., & **Yasseri, T**. (2018). Computational courtship: understanding the evolution of online dating through large-scale data analysis. *arXiv preprint arXiv:1809.10032*.
[21] Zhang, J., & **Yasseri, T**. (2016). What happens after you both swipe right: A statistical description of mobile dating communications. *arXiv preprint arXiv:1607.03320*.
[22] Mergenthaler, A., & **Yasseri, T**. (2021). Selling sex: what determines popularity and rates?, Under Review.