Reg. No: 2021/41/B/HS6/02798; Principal Investigator: dr inż. Radosław Witold Michalski

Modelling social interactions using data streams

Abstract for the general public

Modern societies are deeply immersed in technology. For some researchers, this phenomenon removes the borders and allows spreading ideas faster. Others raise the issues about how much this affects our relationships with others leading to new risks. Surely, this will be a long-lasting process of observing the impact of technology on our lives, but one can say for sure that ICT systems opened the ability to reach and analyse data that was not existing before. Some of it is generated passively, such as communication metadata or our location. However, Web 2.0 enabled us to become the creators of online content, and social media is the way how we communicate and develop social relationships today - both of these types of data also provided new capabilities for social science. Computer scientists, jointly with sociologists and physicists have been able to provide us with new insights about many aspects of our society, e.g. what choices do we make, how we evaluate information, how we communicate or influence ourselves. This would not have been possible at this scale if not the availability of massive data provided by us.

This does not necessarily mean that Big Data is the only way of deriving conclusions about our society. Still, experimental studies performed within small groups or surveying people do matter, and this would not change in the nearest future. It can be rather said that both approaches support each other: sometimes data can be used to provide insights about emerging social phenomena to be further analysed by sociologists, or it serves for large-scale evaluation of the ones that have been observed at a smaller scale by a series of controlled experiments. Undoubtedly, the overwhelming amount of data gives us the opportunity to look at the social phenomena at a completely different scale. However, this comes with a cost. Dealing with Big Data introduces new challenges, such as the requirement for new models, computational complexity or in many ways lack of ground truth that was often observable for small scale studies. Some of these problems are already solved, but many more still require significant efforts of researchers.

In many cases, the network-oriented approach is typically used for modelling social interactions. Here, we gather the events on interactions and build networks of different types and next, by using computational methods, we draw conclusions about different social phenomena. Albeit this approach is still widely used, for some problems, it is not applicable. For instance, some network measures or influence maximisation methods are computationally too complex to be applied for larger graphs, even with the support of GPUs. Secondly, large social networking sites, such as Facebook, Twitter or Weibo restrict researchers from accessing full datasets. In these cases API access to real-time events is provided, sometimes even only to sample of these. Lastly, the era of IoT is shifting the trend from storing all historical data to natively providing access only to recent records due to storage limitations. This imposes the question whether instead of constructing the networks, the streaming approach would be more appropriate since it is closer to the way how the data is being delivered.

In order to provide an answer to that question, the following project aims to investigate the benefits and challenges of using data streams to model social interactions. The joint effort of computer scientists, sociologists and physicists will be carried out in order to find out whether the data streams are capable of representing social interactions in a way that is computable, interpretable and justified by social science theories.

In this project, we will be looking at a number of phenomena observed in society, such as forming groups, information diffusion or spread of influence. However, contrary to typical approaches that base on social networks, we will be using data streams as the primary input and as an outcome of the project, we propose a solution for modelling social interactions based on streamed data. In order to verify the performance and accuracy of our models and methods, we plan to evaluate our approach using multiple datasets that contain social interactions, such as social media, communication data, virtual world environments and datasets provided by our partners that also contain valuable information about ground truth.

Summing up, in this project, we anticipate to propose a comprehensive set of representations, measures and methods for modelling social interactions when these are being provided as data streams. This knowledge can be applied for multiple areas that study the way how we build society and communicate within.