



HORIZON 2020

The EU Framework Programme for Research and Innovation

REN^oIR

Reverse EngiNEering of sOcial Information pROcessing

MAIN PILLAR

Excellent Science

TYPE OF ACTION

Marie Skłodowska-Curie
Research
and Innovation Staff
Exchange

DURATION

1. Jan. 2016 - 31. Dec. 2019
Researcher Months: 292
EC funding: 1.3 M€

 www.renoirproject.eu

 [RENOIRprojectEU](https://www.facebook.com/RENOIRprojectEU)

 [@RENOIRprojectEU](https://twitter.com/RENOIRprojectEU)



This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 691152.

Abstract

In today's world, access to information is a decisive factor advancing industry, society and even culture. It is therefore of great importance to understand why and how some information (e.g. some memes) spreads virally with great ease, while other is met with disinterest and omission. Uncovering the reasons may allow promoting important information, like warnings about cyber-attacks, while stifle harmful rumors, such as vaccines causing autism. The aim of the project is to treat the vast complexity of such information dynamics in social systems by involving researchers in social sciences, journalism, computing, data mining and complexity science.

The Project's objectives are:

- discovery and reverse-engineering the mechanisms of information spreading in social media, such as dynamics of news releases, blogs, Twitter messages, e-mails etc.,

- training and exchange of knowledge between partners in different domains,

- bidirectional knowledge transfer between academia and media industry by exposing researchers to real-life problems and giving business access to innovative methods and tools for information analysis.

The project is based on three pillars: data acquisition, data mining/machine learning and complex systems modeling. The specific problems addressed include understanding rules of and predicting information spreading in different media and about different topics, finding information sources and uncovering hidden information channels.

The secondments will accelerate individual careers of involved researchers, especially early stage ones. The project will lay foundations for long-term collaboration by strengthening existing links between partners and creating new ones.

RENOIR Consortium

WUT, WARSAW UNIVERSITY OF TECHNOLOGY, POLAND

Project Coordinator Janusz Hołyst, *Physics of Complex Systems* (jholyst@if.pw.edu.pl)

JSI, JOZEF STEFAN INSTITUTE, SLOVENIA

Marko Grobelnik, *Computer Science*

PWR, WROCŁAW UNIVERSITY OF SCIENCE AND TECHNOLOGY, POLAND

Przemysław Kazienko, *Computer Science*

STA, SLOVENSKA TISKOVNA AGENCIJA, SLOVENIA

Aljoša Rehar, *Media Company*

PROJECT MANAGER

Anna Grzywacz (agrzywacz@if.pw.edu.pl)

STF, STANFORD UNIVERSITY, UNITED STATES

Jure Leskovec, *Computer Science*

RPI, RENSSELAER POLYTECHNIC INSTITUTE, UNITED STATES

Boleslaw Szymanski, *Computational Network Science*

NTU, NANYANG TECHNOLOGICAL UNIVERSITY, SINGAPORE

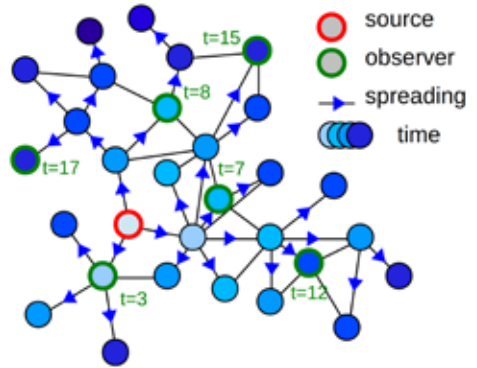
Peter Sloot, *Physics of Complex Systems*

PROJECT OFFICER

Katarzyna Lyson (katarzyna.lyson@cc.europa.eu)

Warsaw University of Technology

The spreading of information in a social network can sometimes be hard to trace and we may only have limited knowledge about where and when the information appeared. With such limited number of “observers” in the network, is it possible to uncover the original source of the information ?



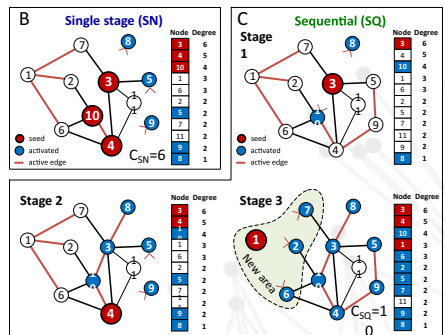
The information spreads from source (red) in time (darker blue means later arrival). Can we find source only knowing the information arrival times observers (green) report ?

We are working on finding such sources. Methods for simple, epidemic-like spreading have been already created. Our work focuses on improving existing algorithms to extend them to more realistic, complex spreading models as well as taking into account variety of real network properties.

By exploiting data from collaboration, we also develop such more realistic models, goal being ability to apply these methods for real systems.

Wrocław University of Science and Technology

Our team focuses on research and development in the area of data science and network intelligence using versatile methods of efficient computing in a distributed computer network environment, intelligent human-computer interaction (HCI), and smart decision-making based on inference from data. Smart decisions are mainly achieved through intelligent processing and analysis through the detection of associations and dependencies in complex, distributed and dynamic data sets and streams.



Information spreading in complex networks is often modeled as diffusing information

Within the RENOIR project, the team is responsible for developing methods for relational learning and inference in dynamic and multiplex networks. The research also focuses on methods to find influencers in multiplex networks.

The Jožef Stefan Institute

Cross-lingual Global Media Monitoring: Event Registry

The system comprises the cross-lingual global media monitoring platform <http://EventRegistry.org> for collecting media information from 300,000 news and social media sources, performing linguistic and semantic processing in multiple languages, forming events and event sequences, streaming information about events, rich visualizations, and search with complex queries for analyzing global social dynamics.

Event Registry plays a key role in the RENOIR project as a tool for analysing cross-lingual news reporting bias with respect to topics, languages, location, political orientation, source and sentiment. A methodology for the analysis of information spreading has already been developed and incorporated into the system. Further work will focus on aggregating similar event sequences into a single probabilistic temporal model to be used for reasoning, prediction and interpretation.

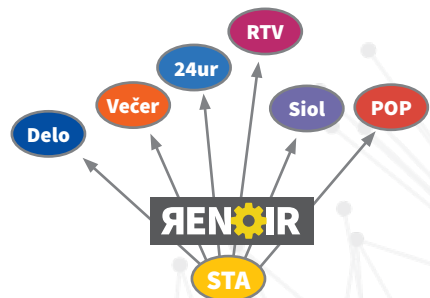


Event Registry collects information from 300,000 news and social media sources to generate events in near real time.

Slovenian Press Agency

STA joined the project as the industry partner tasked with providing academic partners with real-world problems while potentially harnessing the results for operational implementation. The project has already proved invaluable on both counts.

Researchers have developed a prototype tool for the tracking of content uptake by our subscribers, and two more tools facilitating content management are in the pipeline. At the same time, working closely with researchers has expanded, and deepened, our knowledge of the technology underlying the news business, providing us with valuable insights that will inform our business decisions going forward.



The article tracking prototype uses {2,3,4}-gram cosine similarity with TF-IDF for pairwise comparison of articles in order to track STA wire citations by Slovenian media.