A Data-driven, Graphical Approach to Dismantling Illicit & Counterfeit Medicine Markets (ICMs)

Jennifer A. Csicsery-Ronay, Dr. Ioannis A. Kakadiaris, Computational Biomedicine Lab

UNIVERSITY of HOUSTON

Introduction

This work looks to advance the field of machine learning and network science from a distributed machine learning approach. Unearthing illicit supply activity requires the development of distributed data mining methods to detect anomalies and inference of coordination and intent. Developing a centralized data lake and subsequent simulations of the data to address the constraints will be the first work to produce a detailed catalog and proposal of how to set up a fully distributed machine learning solution for various industries.

The tragic and detrimental effects of ICMs are a global phenomenon affecting every country on our planet, with the global pandemic exacerbating the impact of the problem due to supply chain disruptions and a lack of quality control. Beyond the apparent public health impact, socially disadvantaged, and underserved communities are more likely to suffer from the negative impact of counterfeit and poor-quality drugs. A cohesive and just society must protect its most vulnerable and this work promises to contribute to that protection.

Methodology

- **Task 1:** Develop a cloud-based (AWS) data lake that will serve as a centralized, curated, and secured repository for data processed by various data mining methods.
- **Task 2:** Conduct syndromic surveillance to identify regions of ICM spread by implementing automation techniques for monitoring news and health websites for early indications of outbreaks and disease clusters.
- **Task 3:** Construct a multiplex (MPN) representation of the ICM networks using the previously stated aggregated data as well as a 3D graphical representation of the MPN that is easy to explore and exhibit textual/visual explanations for a link or chain of links between entities.
- **Task 4:** Construct a hybrid Graphical Convolutional Network (GCN) - Recurrent Neural Network (RNN) to learn and predict the time evolution of the empirical graph and create synthetic data with time dependence.

Current Progress

The data mining methods implemented yielded 10,000+ images and text that were utilized in developing clustering models to learn of strong and weak connections between markets. A foundational graphic database has been constructed and is being tested using a synthetic dataset mimicking the structure and content of the mined data.

Once this task is complete, work on an automated syndromic surveillance system for identifying early indications of outbreak and disease clusters will begin.

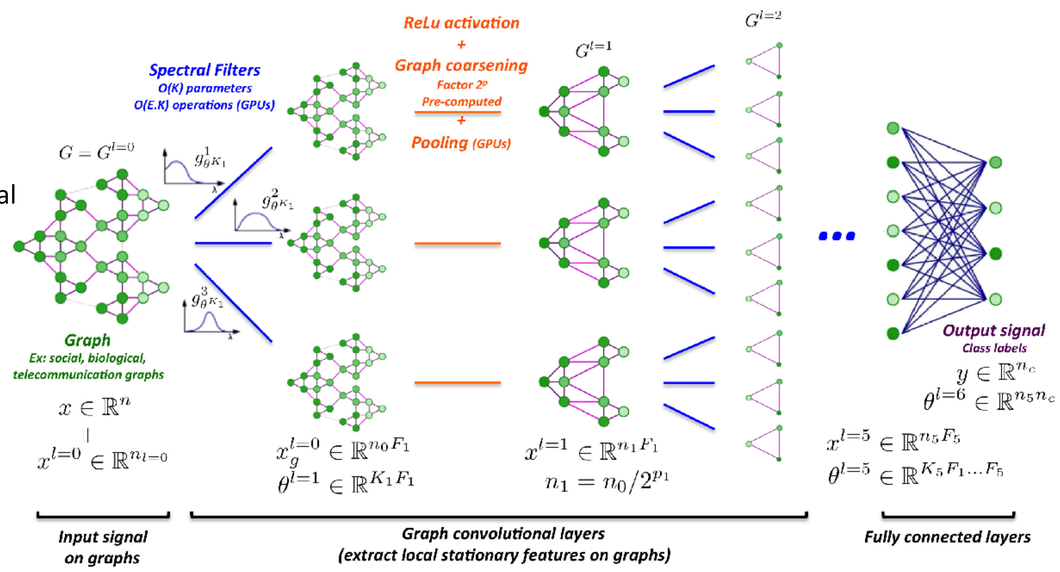
Acknowledgements

PURS Program, Honors College, University of Houston

Dr. Kakadiaris, Mentor, Director, Computational Biomedicine Lab, University of Houston

Timothy Burt, PhD. Candidate, FINDM Project Lead, Computational Biomedicine Lab, University of Houston

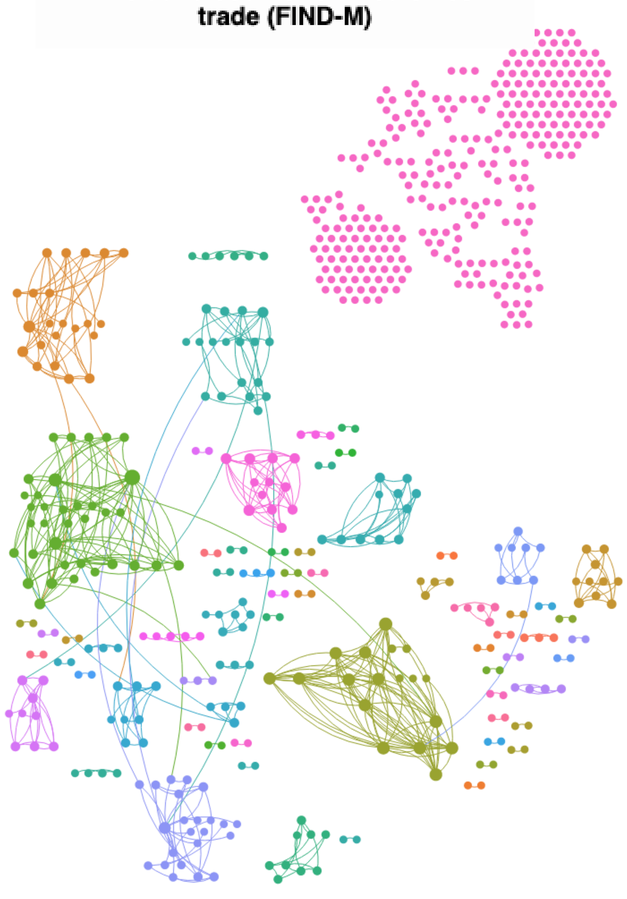
College of Natural Sciences & Mathematics, University of Houston



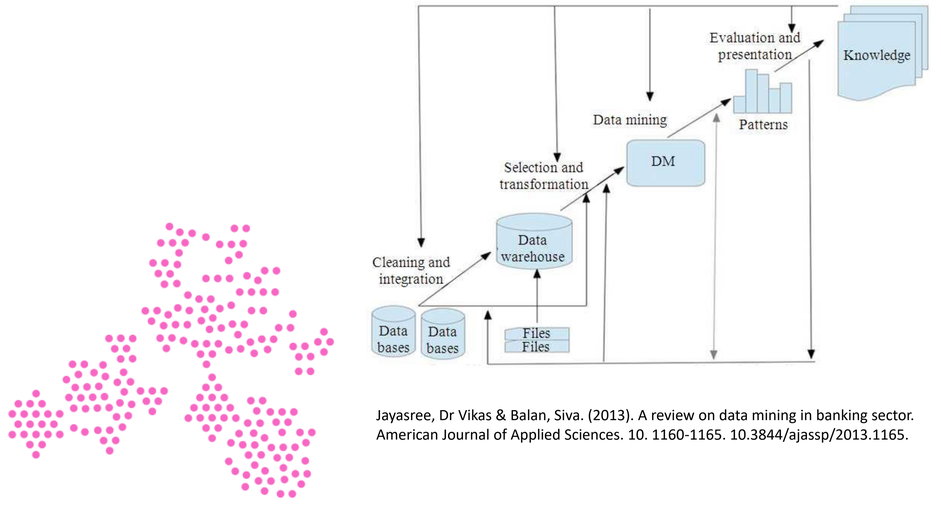
Defferrard, M., Bresson, X., & Vandergheynst, P. (2016). Convolutional Neural Networks on Graphs with Fast Localized Spectral Filtering. NIPS.

FINDM

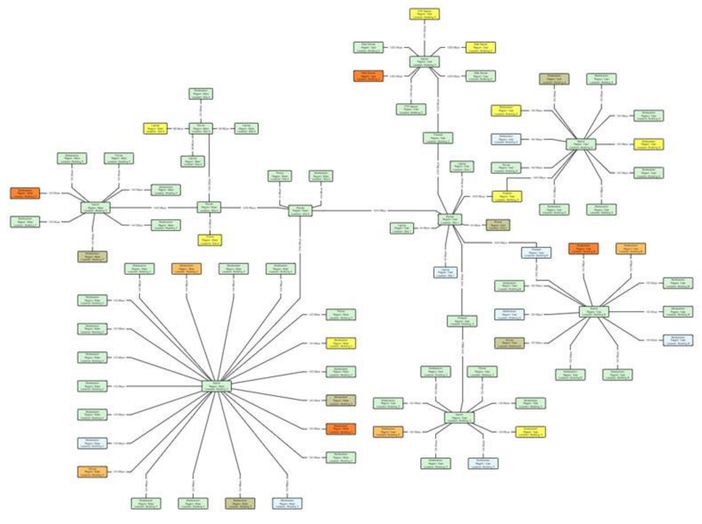
Financial Network Disruptions in illicit and counterfeit Medicines trade (FIND-M)



FIND-M visualizer, Timothy Burt, Ph.D. Candidate, FINDM Project Lead



Jayasree, Dr Vikas & Balan, Siva. (2013). A review on data mining in banking sector. American Journal of Applied Sciences. 10. 1160-1165. 10.3844/ajassp/2013.1165.



Graph database, Tom Sawyer
https://www.tomsawyer.com/graph-database-browser

UNIVERSITY of HOUSTON
COLLEGE of NATURAL SCIENCES & MATHEMATICS