

# Comparison of Network Topological Data versus Binary Representations of Drug-Protein Interactions for Drug Repurposing

Cheuk Lam Jocelyn Ho, Maria Herrero-Zazo

King's College London, Institute of Pharmaceutical Science, 150 Stamford Street, London SE1 9NH, UK.

## Introduction

This study aims to explore the most appropriate ways to use drug-protein interaction (DPI) information in machine learning (ML) methods, to predict new indications for approved drugs. It is hypothesised that using biological networks provides a better ML performance, followed by quantitative representation and binary classification.

### DRUG REPURPOSING

- Identifies new indications for approved drugs [1]
- Sildenafil:** From angina to erectile dysfunction [2]
- Minoxidil:** From hypertension to hair loss [2]
- Reduces risk of toxicology, time and cost [3]
- Higher successful rate and shorter approval time [2]
- New treatment options to unmet medical needs [4]

### DRUG-PROTEIN INTERACTION DATABASES

- DPI information stored in online databases.
- Develop methods to incorporate DPI databases for drug repurposing.



### DRUG-PROTEIN INTERACTION REPRESENTATION

Binary representation	Quantitative representation
• 1 = Have interaction	• Binding affinity values (the smaller the value, the stronger the DPI)
• 0 = No interaction	• Describes degree of binding, how tightly a drug binds to a protein
• Simplest representation	• Uses threshold cut-off values to differentiate strong and weak DPI [5]

### MACHINE LEARNING

- Makes predictions from DPI information.
- Extracts relationships and information from training data set, identifies patterns to test new data sets [6].
- Large number of variables affects performance of ML models, causes less accurate predictions.

### BIOLOGICAL NETWORKS

- Allow visualisation of DPIs.
- Topological data analysis (TDA) extracts features from large data sets to represent global relationships [7].
- Feature reduction:** Properties of each node substitute individual variables, reducing number of variables as input data in ML models.

## Methodology

DPI was collected from online databases and input into ML methods. Results were compared using TDP, quantitative representation, and binary representation. Data set included 308 drugs, 388 proteins, and 6750 DPIs.

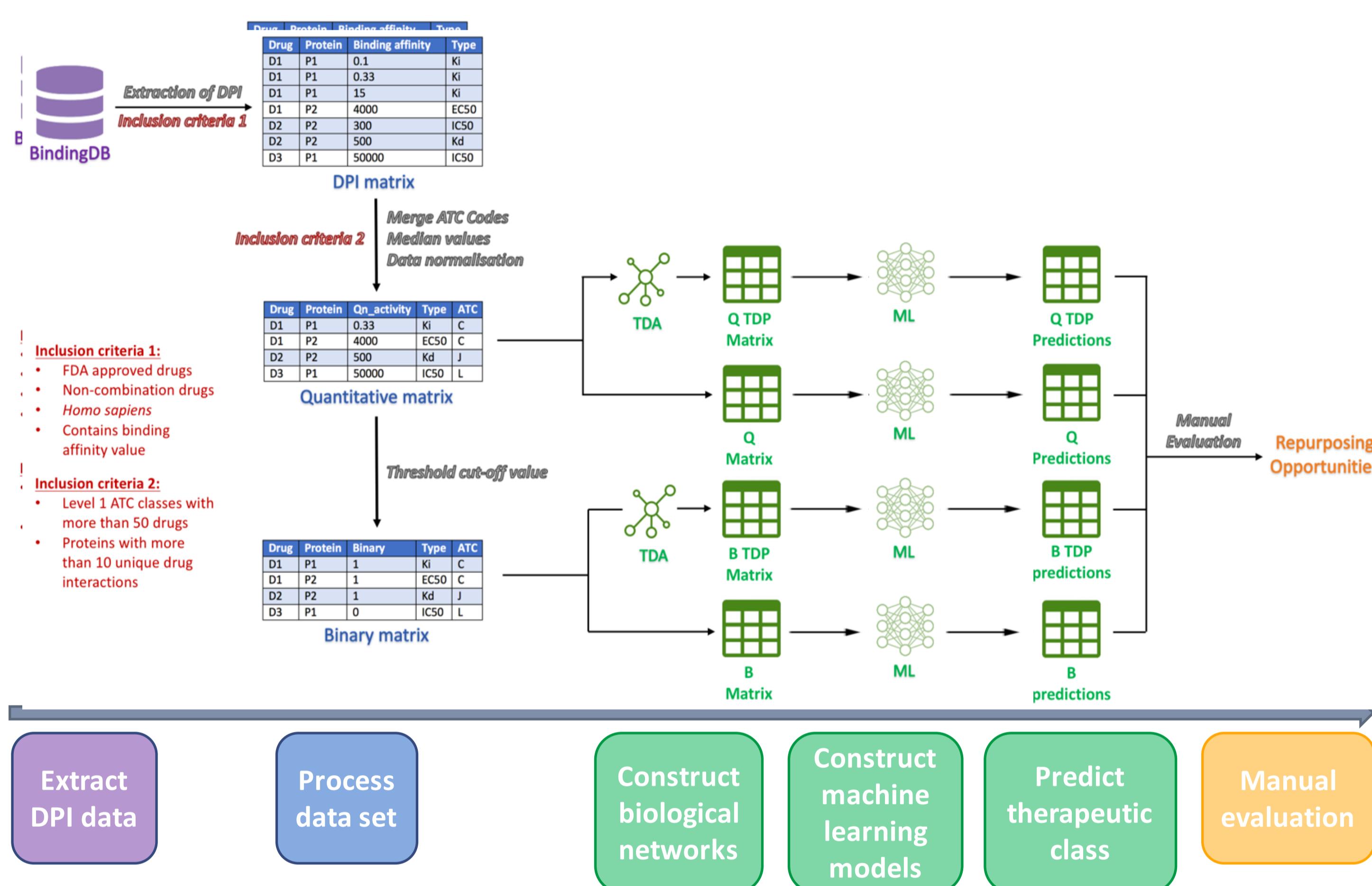


Fig. 1 Flowchart showing the overview of methodology used in this study.

## Results

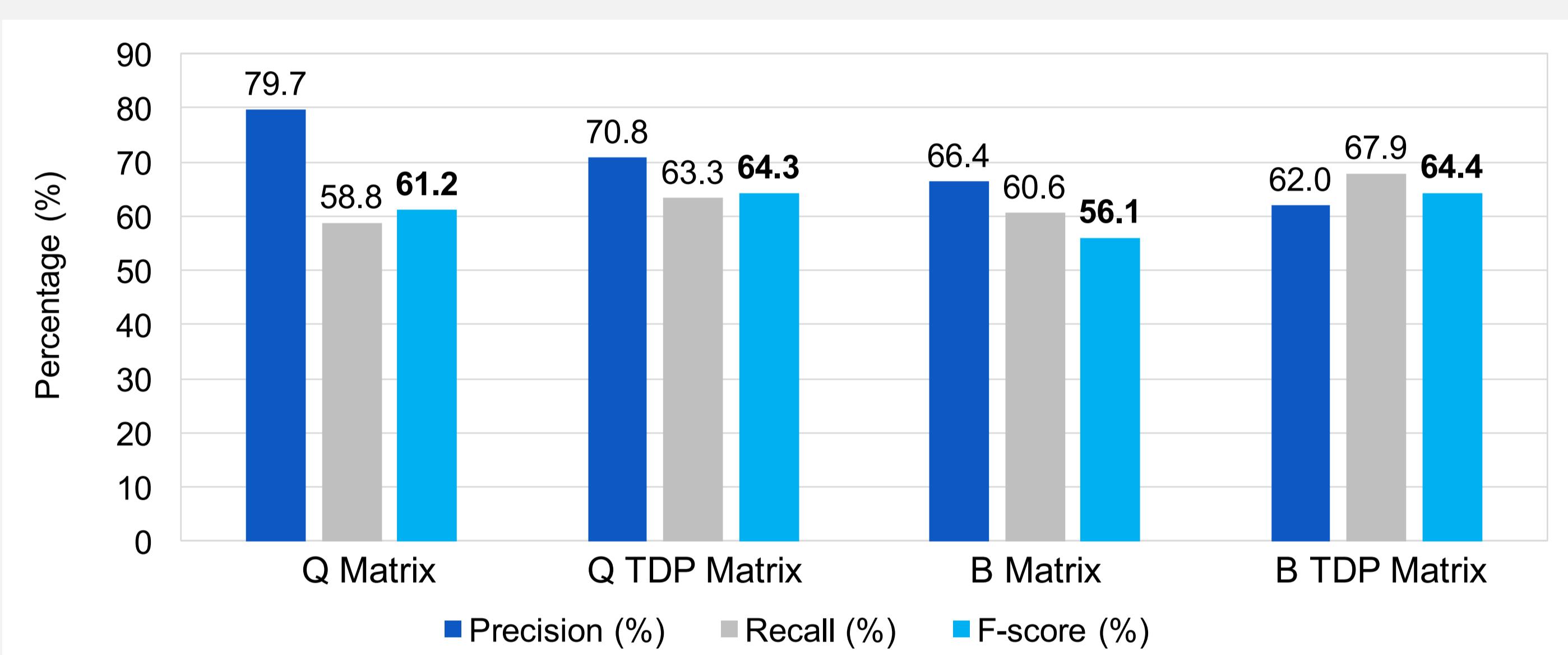


Fig. 2 Bar chart showing precision (%), recall (%), F-score (%) for all four matrices. F-score was used for comparison of machine learning performance. Values were rounded to 1 decimal place.

### Manual evaluation for repurposing opportunities

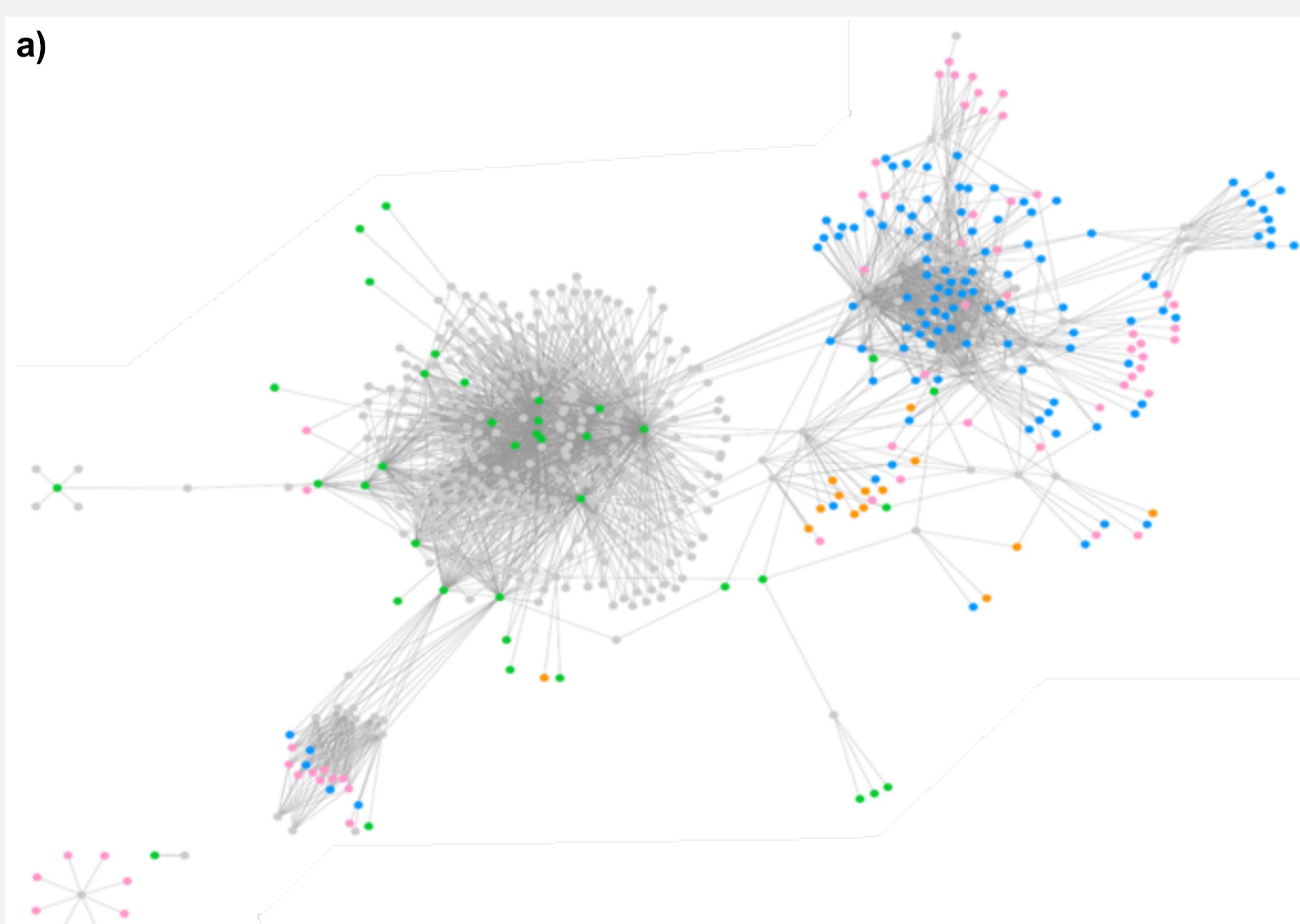
Phentolamine:

- Repeatedly predicted from cardiovascular agent to nervous system agent.
- Causes nervous system disorders, commonly as dizziness [8]
- Intravenous phentolamine can potentially alleviate complex region pain syndrome [9]

Zonisamide:

- Repeatedly predicted from nervous system agent to cardiovascular agent.
- Cardiovascular side effects found during clinical trials, such as palpitation and tachycardia [8]

a)



b)

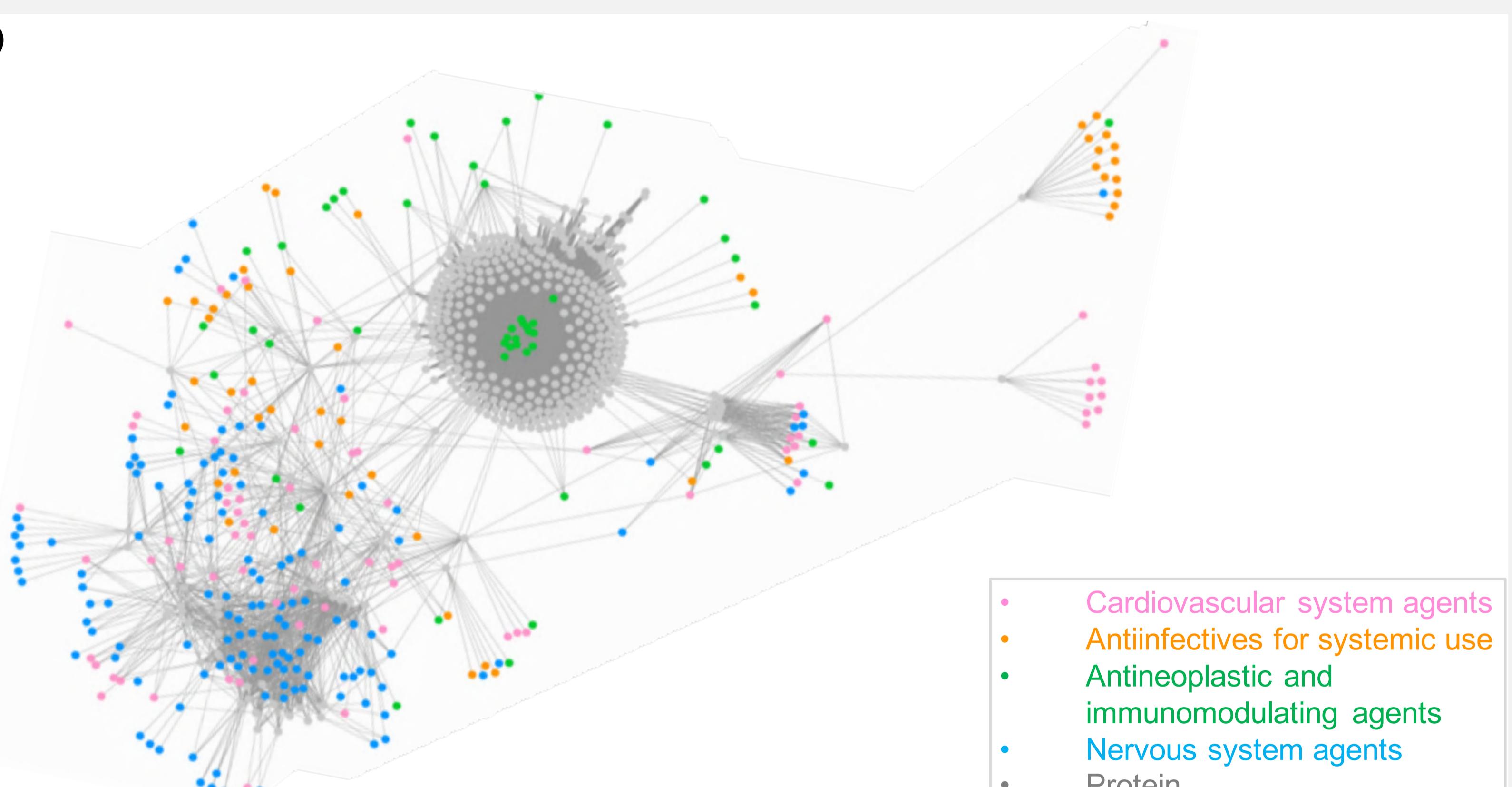


Fig. 3 Biological networks constructed from a) binary matrix, b) quantitative matrix.

## Conclusion

DPI representations	<ol style="list-style-type: none"> <li>Using biological networks was useful for feature reduction in ML and provided more accurate predictions, followed by quantitative representations and binary classification.</li> <li>Similar drugs were grouped together using DPIs, TDA reduced number of variables from over 300 proteins to 10 properties.</li> </ol>
Drug repurposing opportunities	<ol style="list-style-type: none"> <li>Research revealed one possible repurposing opportunity of phentolamine in managing sympathetic pain.</li> <li>Our approach was useful in predicting off-target side effects of drugs. These may be developed as useful therapeutic activities for other patients, and may help establish safety profiles of drugs.</li> </ol>
Future work	<ol style="list-style-type: none"> <li>To include more data on drugs and DPIs.</li> <li>To explore other more sophisticated ML methods.</li> </ol>

## References

- [1] Iorio et al. *Drug Discov. Today*. 2013; [2] Hodos et al. *Wiley Interdiscip. Rev. Syst. Biol. Med.* 2016;8:186–210. [3] Cavalla D. *Drug Discov. Today*. 2013;18:523–32. [4] Napolitano et al. *J. Cheminform.* 2013;5:30. [5] Cao et al. *PLoS One*. 2013;8. [6] Nidhi et al. *J. Chem. Inf. Model.* 2006;46:1124–33. [7] Camara et al. *Cell Syst.* Elsevier Inc.; 2016;3:83–94. [8] Kuhn et al. *Nucleic Acids Res.* 2016;44:D1075–9. [9] Niaki et al. *Acta Med. Iran.* 2011;49:523–6.

## Acknowledgements

CH thank Isabella Granli-Khan for providing assistance in merging online DPI databases with their respective drug therapeutic classes (ATC codes).