Artificial Intelligence for Medical Data with Python

10 SAMPLE SLIDES

8th session – Medical Systems based on Graph Algorithms

UNIVERSITY OF THE AEGEAN

SCHOOL OF ENGINEERING

DEPARTMENT OF INFORMATION AND COMMUNICATION SYSTEMS ENGINEERING

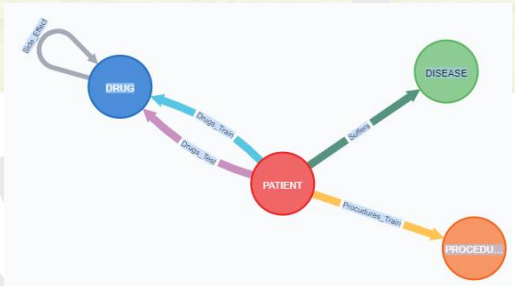
Presenter: Panagiotis Symeonidis

Associate Professor

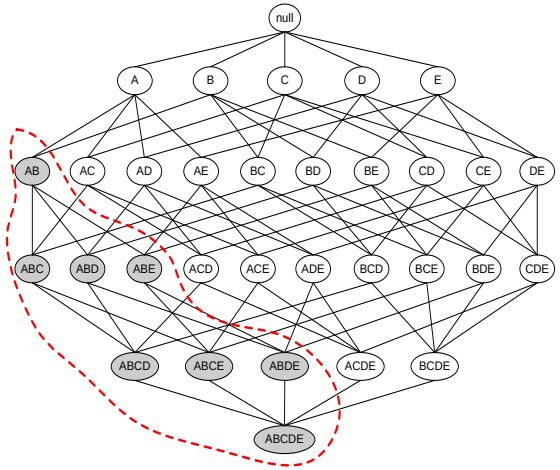
http://panagiotissymeonidis.com

psymeon@aegean.gr

Problem Definition



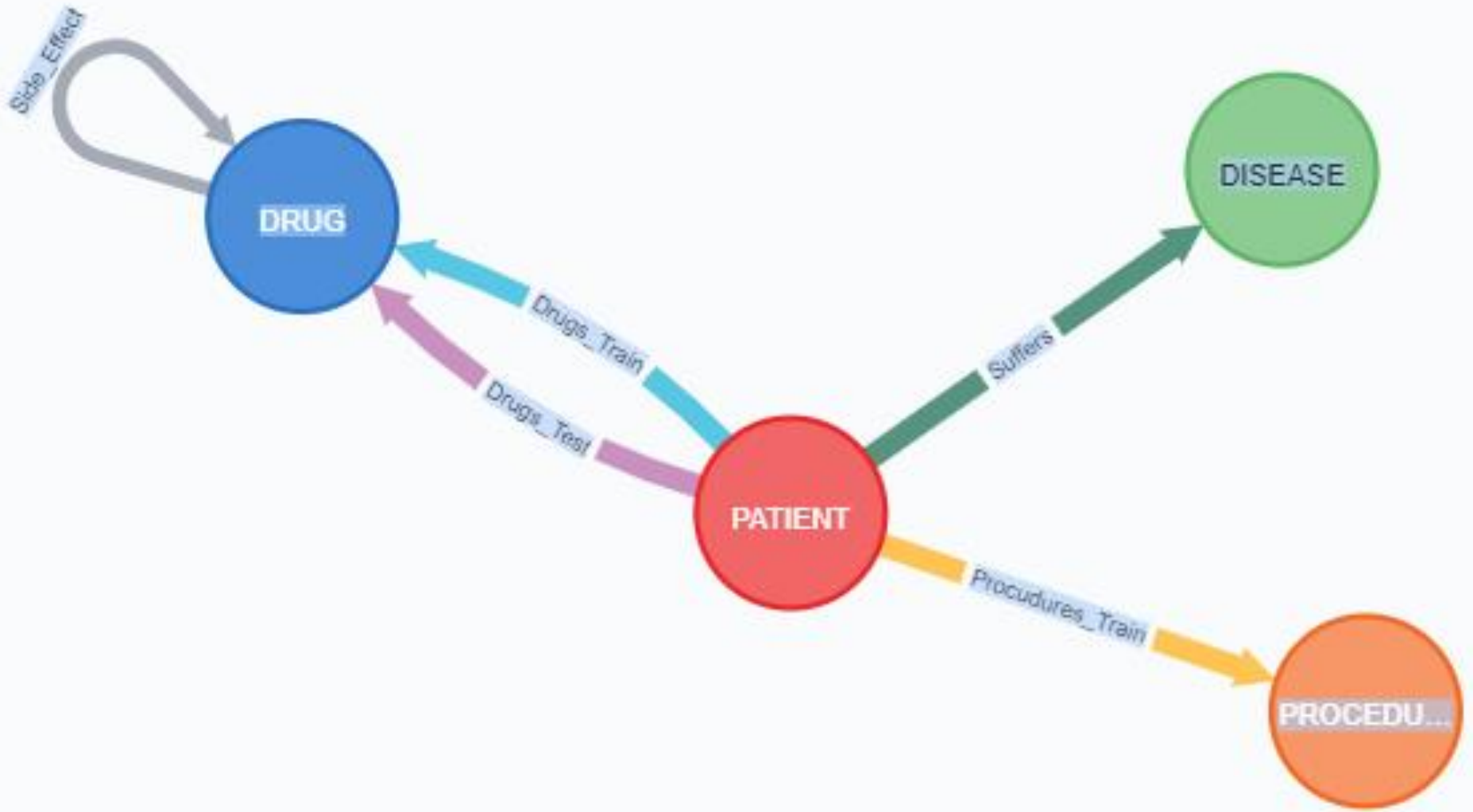
- Can we integrate heterogeneous medical data and represent the patient's clinical status with a Knowledge Graph?



- How can we extract frequent patterns and infer relevance among the participating entities of a knowledge graph in Health?

- How can we support Medical Doctors with recommendation of drugs to be prescribed to patients and explain our recommendations?

Knowledge Graph for Medical Data (RDF triples)



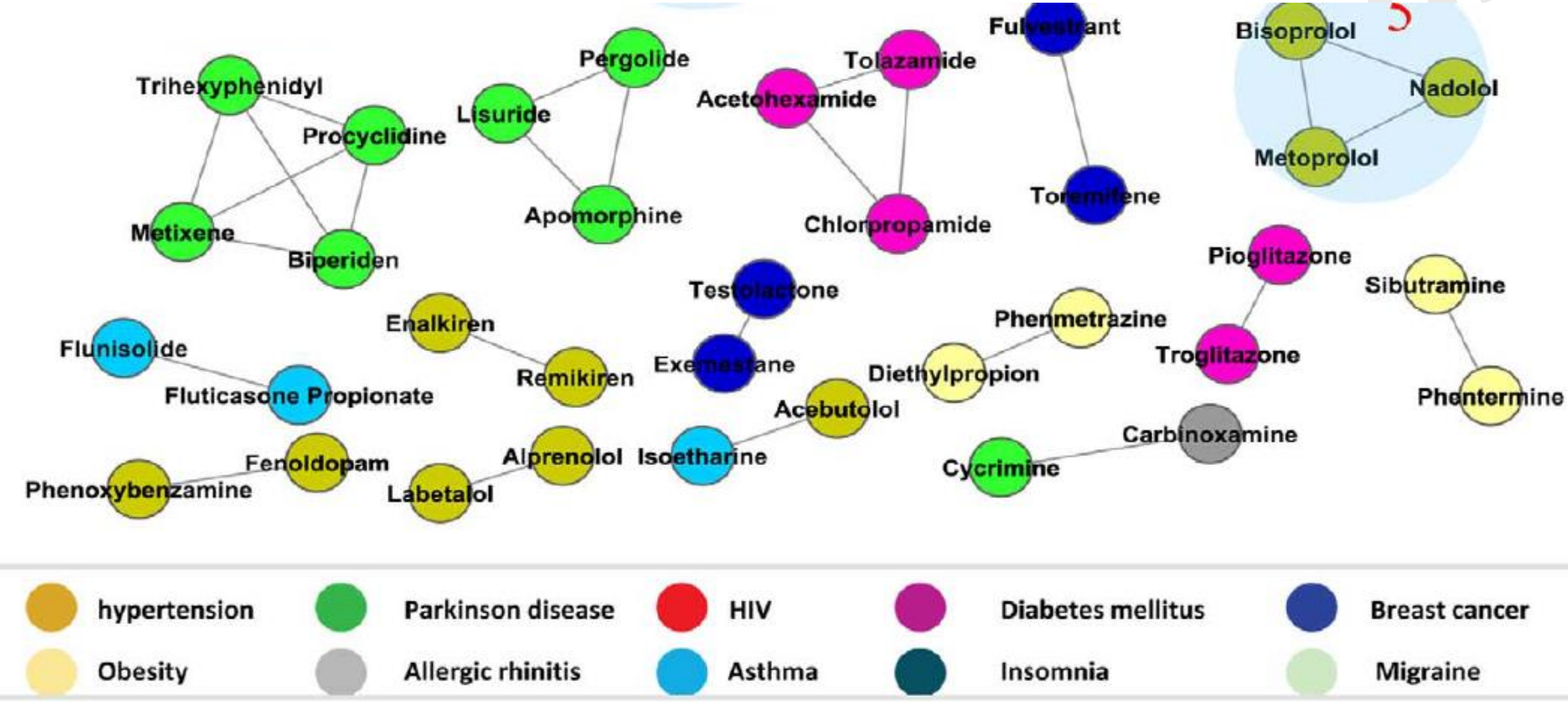
Participating entities

1. Patient
2. Drug
3. Disease
4. Procedure
5. Side Effect

Patients take Drugs to cure Diseases, but may also have side Effects.

Similarity Search and Relevance in Graphs

- Different graph-based algorithms (SimRank, RWR, etc.) can be used to infer Drug-Drug Similarity Network.



Katz Status Index

We capture the similarity score between two nodes v_x and v_y by aggregating the paths of different lengths ℓ connecting v_x to v_y.

$$Katz_{\beta} = \sum_{\ell=1}^{\infty} \beta^{\ell} \, | \, paths^{\ell}_{V_x,V_y} \, |$$

- $paths^{\ell}_{V_x,V_y}$: the set of all paths of length ℓ from node V_x to V_y, computed from the adjacency table A
- β: a damping coefficient, which weights the paths of different lengths according to their contribution to the final similarity score of the nodes. Can take values such as $\beta < 1/\lambda$, where λ is the largest eigenvalue of the adjacency matrix

SimRank for Unipartite Graphs

According to SimRank, two nodes of a graph are similar if they are referenced by similar nodes. And defined as follows:

$$s(a,b) = \begin{cases} 1, & If\ a = b; \\ \dfrac{C}{|I(a)||I(b)|} \sum_{i=1}^{|I(a)|} \sum_{j=1}^{|I(b)|} s\left(I_i(a), I_j(b)\right), & If\ a \neq b, and\ (I(a) \neq \emptyset\ or\ I(b) \neq \emptyset); \\ 0, & If I(a) = \emptyset\ or\ I(b) = \emptyset, \end{cases}$$

Where $C \in [0,1]$ is the attenuation coefficient or else, damping factor, $I(a)$ are the incoming neighbor nodes a, and $I(b)$ are the incoming neighbor nodes of b

Random Walk with Restart (RWR) algorithm

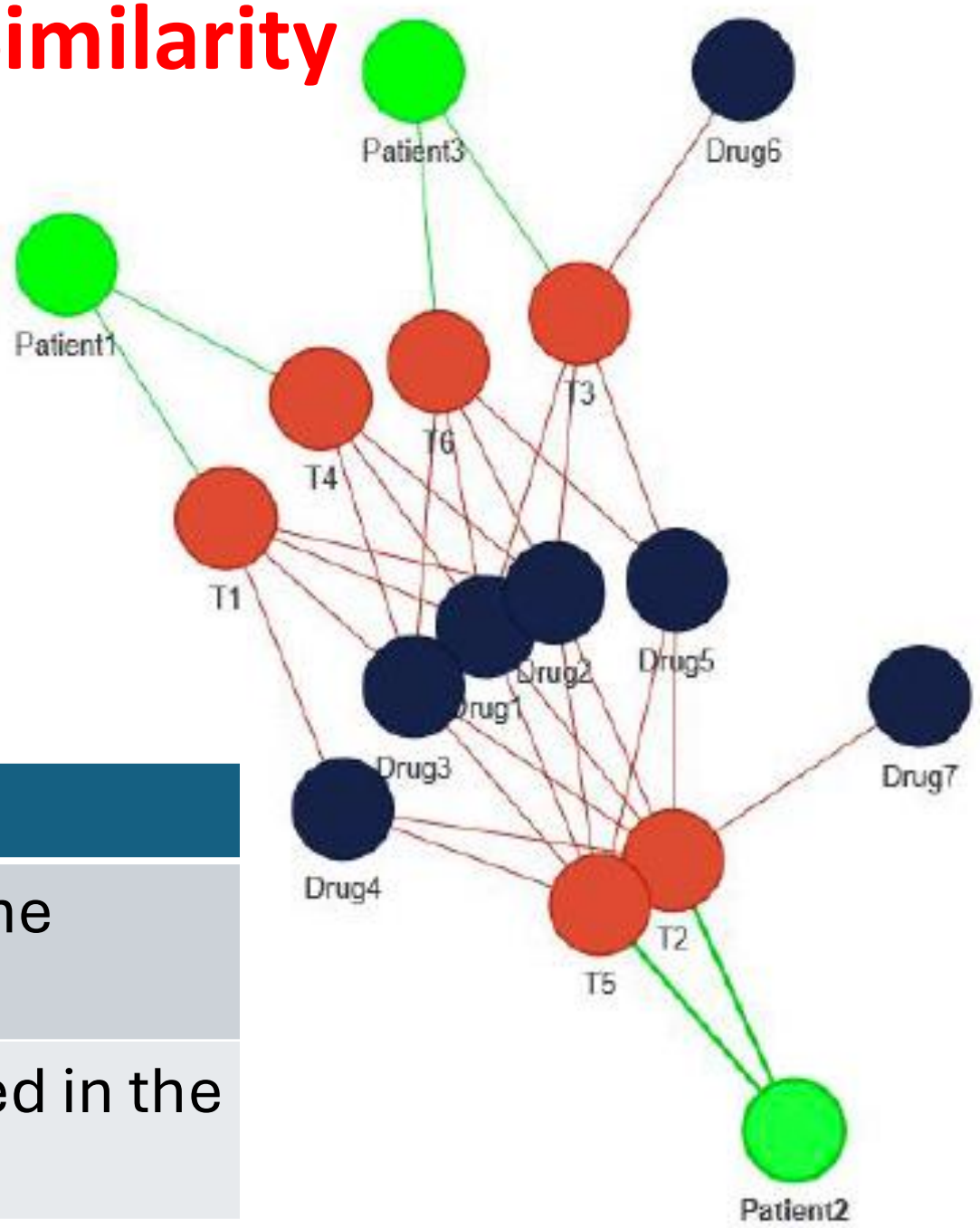
- Running RWR on a graph with n nodes, we obtain a similarity matrix V of dimension *n x n*
- Matrix V' can be computed iteratively using Equation:

$$V' = \beta \cdot M \cdot V + (1 - \beta) \cdot I$$

- Where:
  - V': relevance matrix between any pair of nodes
  - β : the probability for a "random walker" to randomly continue its "walk"
  - M : transition probability matrix of a unipartite graph
  - $(1 - \beta)$:the probability for a "random walker" to interrupt his "walk" and return at the origin node
  - I : identity matrix of dimension *n x n*

Meta Path-based Similarity

- A meta path is a sequence of relations between different node types in a graph, which carries a specific semantic meaning



Meta Path	Semantic Meaning
PTDTP	Patients who took the same drugs
DTD	Drugs which are prescribed in the same treatment

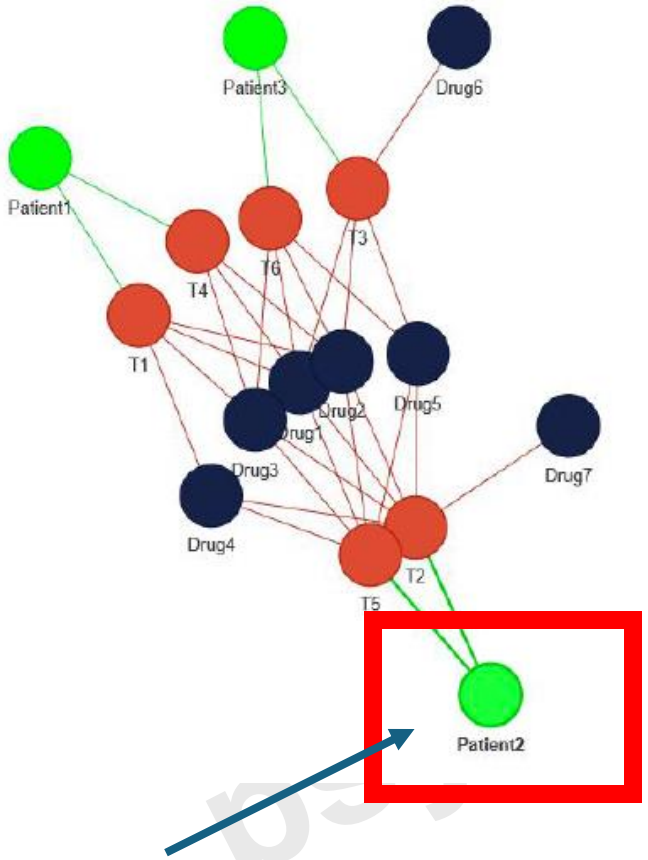
PathSim Algorithm

- Given a symmetric meta-path, the similarity measure is defined as:

$$s(x,y) = \frac{2 * |p_{x\sim y} : p_{x\sim y} \in P|}{|p_{x\sim x} : p_{x\sim x} \in P| + |p_{y\sim y} : p_{y\sim y} \in P|}$$

- where:
  - $s(x,y)$: Similarity measure between node x and node y.
  - $p_{x\sim y}$: is a path that starts from the origin node x and ends at the destination node y,
  - $p_{x\sim x}$: is a path that starts from node x and returns back to the same node,
  - $p_{y\sim y}$: is a path that starts at node y and returns back to the same node.

Meta path-based Explanations



Which drug should we recommend for patient 2?

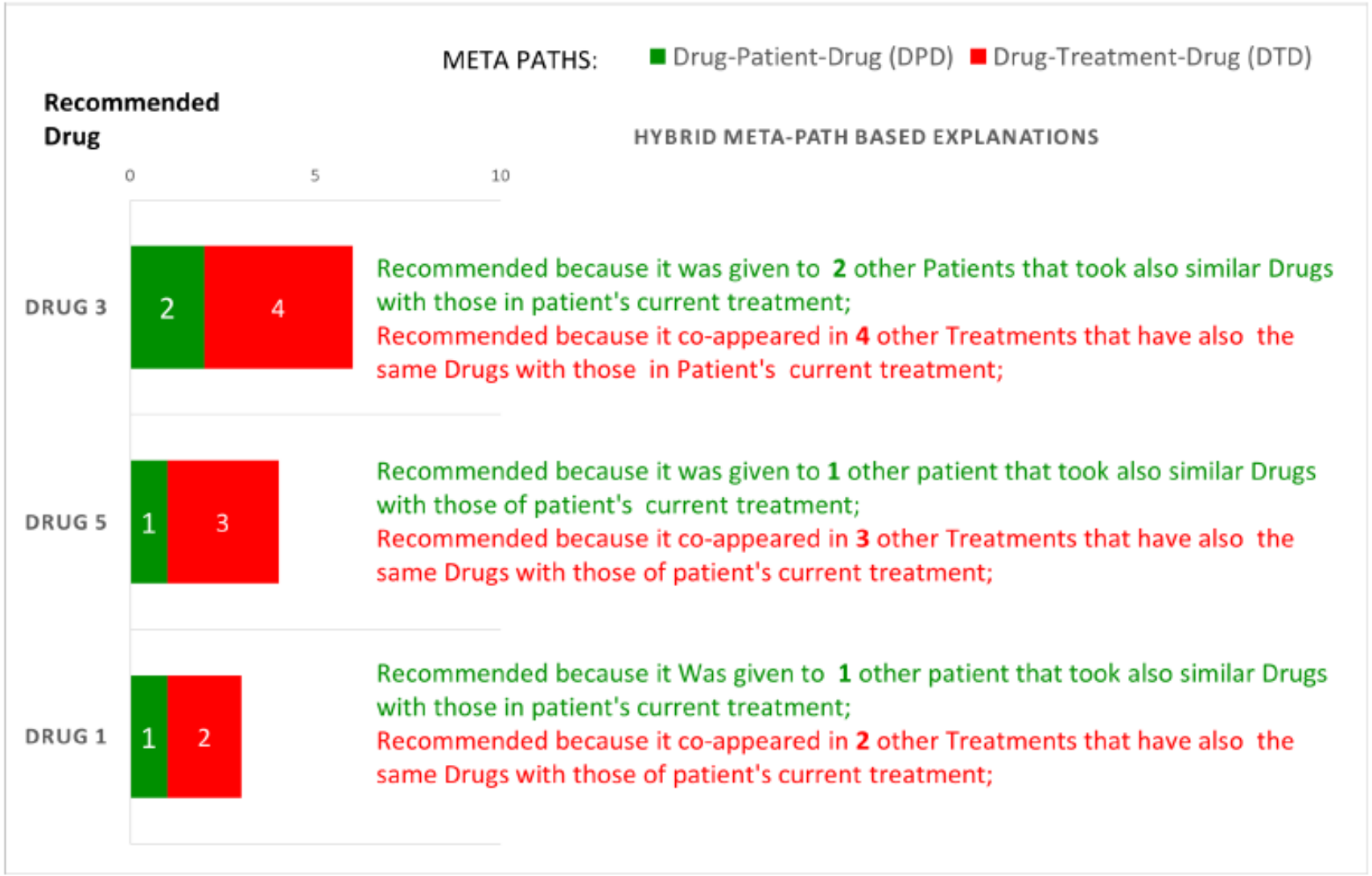


Fig. 4: Drug Recommendations along with explanation for patient 2. Drug 3 is recommended because it has the most frequent meta paths supporting it.