Treatment Recommendations for COVID-19 Patients along with Robust Explanations

Panagiotis Symeonidis
University of the Aegean
Samos, Greece
psymeon@aegean.gr

Christos Andras
International Hellenic University
Thessaloniki, Greece
andras@ihu.gr

Markus Zanker
Free University of Bozen-Bolzano
Bolzano, Italy
mzanker@unibz.it

2021 IEEE 34th International Symposium on Computer-Based Medical Systems (CBMS) | 978-1-6654-4121-6/21/$31.00 ©2021 IEEE | DOI: 10.1109/CBMS52027.2021.00020

***Abstract*—The global response to the pandemic introduced by COVID-19 is unprecedented. Scientists develop methods, which analyze data to identify an effective treatment that uncovers possible responses to the SARS-COV-2 virus. However, our global response should be based on knowledge exchange and collaboration among countries. In this paper, we present a recommender system for treatment recommendations, which exploits similar patterns among patients of different clinical studies, and recommends them health interventions (such as to provide oxygen therapy) and drugs (e.g., Remdesivir) based on their symptoms' or diseases' similarity with patients of other similar clinical studies. Our approach can also provide explanations along with recommended treatments to assist doctors in understanding the reasons behind a suggested drug or health intervention. We also perform experiments to identify the effectiveness of our system in terms of recommendation accuracy. Our results demonstrate that our system is able to minimize the false positive and false negative prediction rates. Finally, we provide web links to download both (i) our program's setup file and (ii) our Neo4j database file.**

***Index Terms*—recommender systems, graph-based algorithms, meta path**

I. Introduction

World-leading scientists and health experts are developing approaches to treat and protect people, and to prevent the COVID-19 disease from spreading. Continuously new treatments are developed to minimise symptoms in coronavirus patients or to reduce the duration of treatments in hospitals.

In this paper, we therefore present a recommender system which collaboratively identifies similar patterns between patients of different clinical studies and recommends suitable treatments (i.e., drugs and health interventions) for patients. The data used for demonstrating this approach is derived from the COVID-19 pandemic. Treatment recommendations are based on similarities between patients in different clinical studies and explanations are derived thereof to justify treatment suggestions.

While Artificial Intelligence (AI) could assist medical doctors in many ways, it is unlikely to replace them in the foreseeable future. To this end, our recommender system tries to assist doctors by providing them supplementary drug and health intervention recommendations along with explanations based on similar clinical studies. Obviously, medical doctors may not be always aware of all different medication and health interventions taking place around the Globe, due to rapid evolving of the pandemic. In contrast, our system can be continuously updated with medical data from recent clinical studies.

Our recommendation system runs on multi-dimensional graph data consisting of different participating entities (i.e. patients, drugs, diseases, etc.). In particular, our system runs on a Heterogeneous Information Network (HIN) and uses meta paths to infer similarities among entities. A meta path is a sequence of different node and edge types, which capture a specific relation among graph entities. For example, the meta path (Disease-Patient-Disease or else DPD) can be used to infer similarities among diseases based on the patients' co-disease Electronic Health Records (EHRs). Our system is capable of providing explanations along with the recommended health treatments based on the aforementioned meta paths. Thus, it can better assist doctors to understand the hidden health correlations behind a suggested drug or intervention.

The paper is organized as follows. Section II presents the characteristics of the COVID-19 medical data set [2]. Section III and IV present our proposed algorithm, and recommender system, respectively. Experimental results are presented in Section V. Section VI discusses the main challenges for our system and how we can deal with them. Finally, Section VII concludes the paper.

II. The COVID-19 Graph Data

We have imported the data from the covidanalytics.io[1] website [2] into the Neo4j database as shown in Figure 1. You can download our Neo4j database file here.

The covidanalytics dataset contains the following data columns:

- **Clinical study** : Id, Sub-id, Country, Population
- **Prevention-life style** : Smoking history, Current drinker BMI, Obesity, Any Comorbidity.
- **Symptoms** : Fever (temperature 37·38 celcius), Average temperature (celsius), Max temperature (celsius), Respiratory rate $>$ 24 breaths per min, Cough, Shortness of Breath (dyspnoea), Headache, Sputum (/Expectoration), Myalgia (Muscle Pain), Fatigue Upper air-way congestion, Diarrhoea, Nausea or Vomiting, Loss of Appetite/Anorexia, Sore Throat/Stuffy, Nose Chills, Chest Pain, Loss of smell/taste.

[1] https://www.covidanalytics.io/dataset

DOI 10.1109/CBMS52027.2021.00020

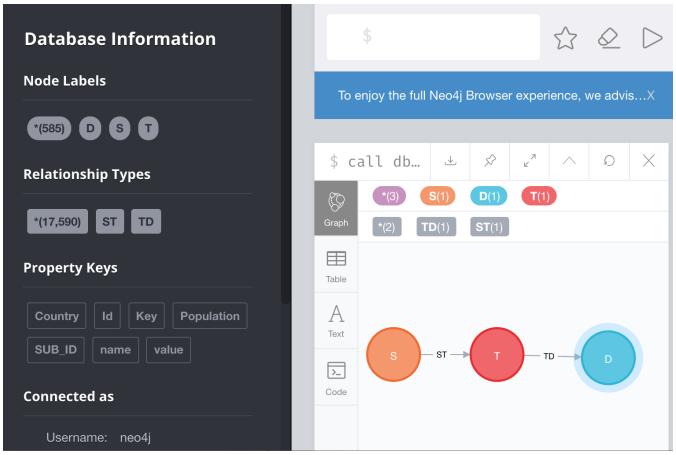

Fig. 1: The network schema of our COVID-19 tripartite graph.

- **Diagnoses** : Hypertension, Diabetes, Cardiovascular, Disease (incl. CAD), Chronic obstructive lung (COPD), Cancer (Any), Liver Disease (any), Cerebrovascular Disease, Chronic kidney, Renal Disease, Other.
- **Drugs** : Antibiotic, Antiviral (Any), Uses Kaletra, (lopinavir/ritonavir), Uses Favipiravir, Uses Tamiflu (oseltamivir), Uses Remdesivir, Uses Umifloprit, Uses Arbidol (umifenovir), Uses Hydroxychloroquine and/or Chloroquine, Corticosteroid (including Glucocorticoid, Methylprednisolone).
- **Health Interventions** : Nasal Cannula High-flow Nasal Cannula, Oxygen therapy, Noninvasive Mechanical Ventilation, Invasive Mechanical Ventilation, ECMO, Renal Replacement Therapy, Interferon Alpha-1b, Thymalfasin and/or Thymosin.
- **Unwanted side effects** : Sepsis, Respiratory Failure or ARDS, Respiratory Failure ARDS, Hypoxemia, Heart Failure, Septic Shock / Shock, Liver Lysfunction, Coagulopathy, Acute Cardiac Injury, Acute Kidney Injury (AKI), Secondary Infection/ Bacterial Infection, Hypoproteinaemia, Acidosis.
- **Survivorship**: ICU admission, discharged (%), ICU length of stay (days), hospital length of stay (days), days to viral clearance (Median), mortality.

Henceforth, we group some categories from the aforementioned ones as follows: (i) "Symptoms" and "Diagnoses" are together considered as "Diseases" and (ii) "Drugs" and "Health Interventions" are together considered as "Treatment".

In our graph, we have 538 (main- and sub-clinical Studies) (S), 19 different Treatments (drug or health interventions) (T) for COVID-19 and 28 Diseases (symptoms or underlying chronic diseases) (D). As it is shown in Figure 1, we have imported all relationships among the aforementioned three entities (S, T, D), which results in 17,590 relationships of type S $\rightarrow$ T (the treatment that was provided to patients of a clinical study) or type T $\rightarrow$ D (the treatment protocol which is followed for a given list of diseases/symptoms). Thus, the network structure of our graph, as shown in Figure 1, is Clinical Study – Treatment – Disease, i.e., S $\rightarrow$ T $\rightarrow$ D, as can be shown on the bottom right of Figure 1, with the orange, red, and blue node, respectively. Our graph consists of 585 nodes (538 main- and sub-clinical study nodes + 19 treatment nodes + 28 patient disease and symptom nodes). There are also two types of links (ST and TD). ST represents the fact that the patients of a clinical study were prescribed a list of drugs (e.g., Remdesivir, Hydroxochloroquine, etc.) or were provided a list of health interventions (e.g., to get an oxygen therapy, admission in ICU, etc.), whereas TD represents the treatment that should be followed given specific patients' symptoms and their possible underlying chronic diseases (diabetes, cancer, etc.).

III. The PathSim Similarity Measure

In this Section, we describe the PathSim similarity measure used in our system that is inspired by the work of Sun et al. [10]. They proposed the novel idea of measuring similarities between network objects by analysing meta-paths, through which objects are connected. In a heterogeneous network, two objects can be connected through different paths as defined in the following:

Information Network. *[10] An information network is defined as a directed graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ with an object type mapping function $\phi : \mathcal{V} \rightarrow \mathcal{Q}$ and a link type mapping function $\psi : \mathcal{E} \rightarrow \mathcal{R}$, where each object $v \in \mathcal{V}$ belongs to one particular object type $\phi(v) \in \mathcal{Q}$, and each link $e \in \mathcal{E}$ belong to a particular relation $\psi(e) \in \mathcal{R}$.*

For instance, in the network schema depicted in Figure 1 two treatments can be connected through the path "Treatment-Disease-Treatment" , or "Treatment-Disease-Treatment-Disease-Treatment" (collaborative filtering similarity). Using different paths, different similarities are observed. These paths are called *meta paths* and are formally defined as follows:

Meta Path. *[10] A meta path $\mathcal{P}$ is a path defined on the graph of network schema $T_G = (\mathcal{Q}, \mathcal{R})$, and is denoted in the form of $Q_1 \xrightarrow{R_1} Q_2 \xrightarrow{R_2} \ldots \xrightarrow{R_l} Q_{l+1}$, which defines a*

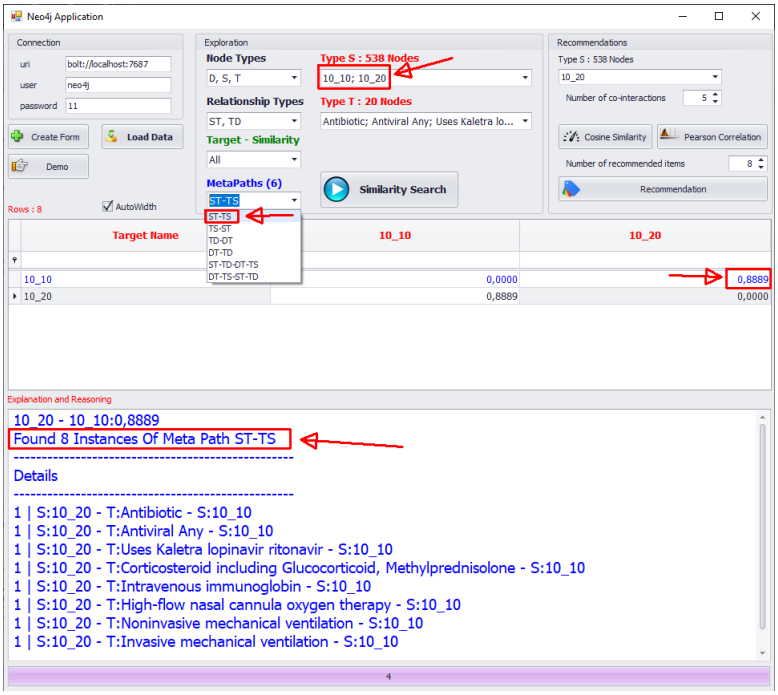

Fig. 2: The similarity search and exploration task.

composite relation $R = R_1 \circ R_2 \circ ... \circ R_l$ *between type* Q_1 *and* Q_{l+1}*, where* $\circ$ *denotes the composition operator on relations.*

There are various meta paths that can be built based on the clinical Study-Treatment-Disease (STD) network structure of Figure 1. If we start from a treatment node, we can build the following paths: TDT, TST, etc. If we start from a disease node, we can build meta paths: DTD, DTSTD, etc. to recommend treatments based on similar diseases.

The meta path framework provides a powerful mechanism for a doctor to select the appropriate similarity semantics by choosing a proper meta path. A well-known similarity measure that is able to capture the semantics of similarity among network objects is **PathSim** [10]. Another variation of PathSim was proposed by Sun et al. [8] for predicting future links (e.g., future collaborations or co-authorships) among researchers. In particular, they proposed **PathCount** [8], which measures the number of path instances between two objects following a given meta path. Recently, Behrens et al. [1] proposed MetaExp, a system which is based on the PathSim algorithm to infer similarity among entity nodes using meta paths.

A. Meta Path-based Similarity

PathSim: A Single Meta path-based similarity measure [10]. *Given a symmetric meta path P, PathSim between two objects of the same type* x *and* y *is*

$$s(x,y) = \frac{2 * |p_{x \rightsquigarrow y} : p_{x \rightsquigarrow y} \in P|}{|p_{x \rightsquigarrow x} : p_{x \rightsquigarrow x} \in P| + |p_{y \rightsquigarrow y} : p_{y \rightsquigarrow y} \in P|}, \quad (1)$$

where $p_{x \rightsquigarrow y}$ *is a path instance between* x *and* y*,* $p_{x \rightsquigarrow x}$ *is that between* x *and* x*, and* $p_{y \rightsquigarrow y}$ *is that between* y *and* y*.*

PathSim similarity measure overcomes problems of similarity inference in graphs. That is, Jeh and Widom [3] proposed SimRank based on the idea that two nodes are similar if they are referenced by similar nodes. Moreover, the Random Walk with Restart (RWR) (a.k.a. P-PageRank) algorithm [4]–[6], [9], which is a variation of the well-known PageRank algorithm, has properties, that capture also the notion of nodes' similarity in a heterogeneous graph. However, both of the aforementioned algorithms are biased towards those nodes that have high node's degree (i.e., nodes with many links to other nodes). In contrast to SimRank and P-PageRank (RWR), which favour more popular items in the network, PathSim is able to capture the nodes' *visibility* in the network, bringing the nodes that share similar visibility closer.

IV. The Proposed Recommmender System

Our recommender system is designed in the .Net framework 4.8 and connects to a Neo4j database through the official neo4j.driver 1.7.2. It consists of two different components: (i) Similarity Search, Exploration and Reasoning and (ii) Treatment Recommendation. You can download the installation of the desktop application here. You can find step by step guidelines for the installation here.

A. Similarity Search, Exploration and Reasoning Component

In this Section, we describe how the first component of our system can be used as an exploratory tool to assist doctors in identifying similar patterns among different clinical studies. In particular, when a physician clicks the button [Load Data], as it is shown on the left top side of Figure 2, our system reads abstractly from the COVID-19 graph database all the different type of nodes, and relationships.

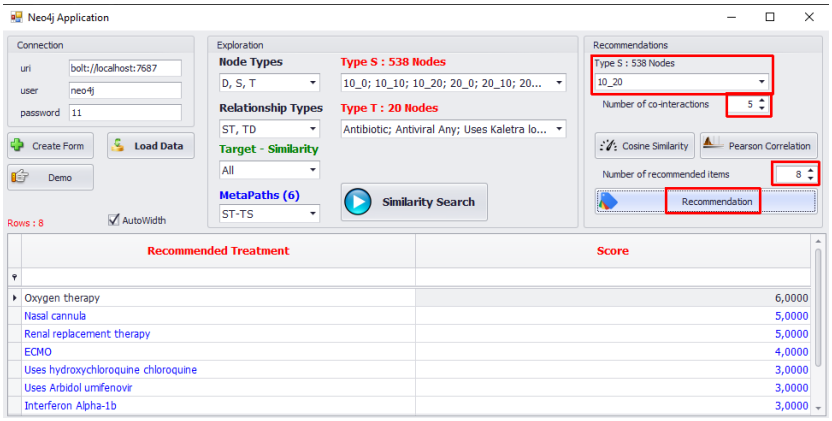

Fig. 3: The Treatment Recommendations for COVID-19.

Then, as shown in the drop box of Figure 2, our system identifies meta paths of length 1 (i.e., STS, TST, TDT, DTD) and meta paths of length 2 (STDS, DTSD) in the graph. When the doctor selects the meta path that he wants to semantically explore (e.g., STS), the system loads all possible instances for each entity that participates in the relationships of the selected meta path.

For instance, as shown in Figure 2, let us assume that we want to analyze the similarity patterns between the patients of the clinical studies with id 10_10 and 10_20 from the covidanalytics website [2] and then click the similarity search button to explore their similarity in terms of their common treatments (i.e., meta path STS). Please note that we consider all the drugs and health interventions that the patients of both clinical studies have in common as "common treatment". Based on the PathSim algorithm [10] (see Figure 2) we compute the predicted similarity between patients of the two clinical studies equal to 0,8889 based on their common treatment (i.e,. the drugs they were prescribed and the health interventions which took place during their hospitalization). Then, by clicking on this value we exploit the predicted degree of similarity as an explanation, where all the instances that connect the patients of the two clinical studies in terms of their treatment are considered (see Figure 2). As shown, the patients of the clinical study 10_10 are connected with the patients of the clinical study 10_20 with 8 instances of the meta path STS (study-treatment-study). As indicated in Figure 2, both clinical studies received medication such as Antibiotics, Antiviral, Kaletra, etc. In addition, patients of both clinical studies had common health interventions such as High-flow Nasal Cannula, Oxygen Therapy and Mechanical Ventilation. Based on the aforementioned meta path-based explanations, we expect that doctors can understand in a more transparent way, what are the latent health correlations between the two - under comparison - or even more clinical studies.

B. Drug and Health Intervention Recommendations Component

In this Section, we describe how the second component of our system can provide treatment recommendations for the patients of a target clinical study. As it is shown in Figure 3, we recommend to the patients of the clinical study with id 10_20, a list of top-8 recommended drugs and health interventions (i.e., treatment recommendations) based on the User-based Collaborative Filtering (UBCF) algorithm [7] (i.e., by using metapath STS) and by taking under consideration the similarity of the target clinical study 10_20 with its neighbor clinical studies (i.e., those clinical studies which have in common with the target clinical study at least 5 drugs or health interventions). That is, given a target clinical study, a Collaborative Filtering algorithm will identify their neighbourhood of similar clinical studies.

In particular, we recommend to the patients of the clinical study to get Oxygen therapy, Nasal cannula, Renal replacement therapy, and ECMO. We also recommend them to get drug combinations such as Hydroxychloroquine, Arbidol umifenovir, and Interferon Alpha-1b, ranked based on a given score, as it is shown on the bottom left side of Figure 3.

V. Experimental Evaluation

In this Section, we perform experiments with the goal to evaluate the treatment suggestions of our system to the patients' studies. The data set characteristics are described in details in Section II. We have followed a 5-fold cross validation protocol and created 80-20 train-test splits of the treatments (i.e., drugs and health interventions) that where given/followed in each patient study.

A. Balancing data between positive and negative examples

For a fair and coherent evaluation protocol, all link connections between a patient's study and a treatment will act as our positive examples, but we'll also need to create some negative examples. Most real-world networks are sparse, with a small number of relationships among graph entities. In our case, the graph properties are not different. The number of examples between a patient's study and a treatment node that do not have a relationship is much larger than the number that do have a relationship. If we had used all the negative examples to train our model, we would have a severe class imbalance problem. In other words, a model could achieve extremely

high accuracy by predicting that every pair of study-treatment nodes does not have a relationship. To overcome this problem, we have selected a balanced number of positive and negative examples (for the train set data, which corresponds to 1294 edges from a clinical study node to a treatment node, we have created other 1294 clinical study-treatment *negative* edges). In particular, for the train set, for every positive example, we have selected one negative example to train our model, resulting to 1294 instances for the positive and 1294 instances for the negative samples. For the test set, we followed the same procedure resulting to 325 instances for the positive samples and 325 instances for the negative samples.

B. Methods' Comparison

In this Section, we compare the performance in terms of precision, recall and AUC (Area Under Curve) of the following three models: (i) Recommend treatments based on meta path STS (i.e., number of common treatments between two patients' studies), denoted as $\mathrm{PathSim}_{\mathrm{STS}}$. This model is the analogue of User-based Collaborative Filtering [7]. (ii) Recommend treatments based on meta path TSTS (i.e., number of paths that connect a patients' study with a treatment through any other patients' study), denoted as $\mathrm{PathSim}_{\mathrm{TSTS}}$. (iii) Recommend Popular Treatments by taking under consideration the number of direct edges between a specific treatment and all patients' studies, denoted as POP.

Table I presents the methods' performance in terms of accurate recommendations on the Covid-19 data set. As shown in Table I, $\mathrm{PathSim}_{\mathrm{STS}}$ outperforms $\mathrm{PathSim}_{\mathrm{TSTS}}$ in terms of precision, which is equal to 0.99. That is, $\mathrm{PathSim}_{\mathrm{STS}}$ is excellent at predicting almost all the links that exist between a patient study and a treatment (i.e., it predicts an extremely small number of false positives, and does not recommend unneeded drugs for the patients' treatment). The main reason that precision is so high, is the fact that the data set is dense and there are not many treatments. They are just 19 to recommend. In contrast, in recommender systems with thousands of movies the recommendation task is much harder.

Model	Covid-19 Data set		
	Precision	Recall	AUC
$\mathrm{PathSim}_{\mathrm{STS}}$	0.99	0.93	0.97
$\mathrm{PathSim}_{\mathrm{TSTS}}$	0.70	0.96	0.54
POP	0.81	0.10	0.56

TABLE I: Models' comparison in terms of accurate medication recommendations.

$\mathrm{PathSim}_{\mathrm{STS}}$ has also a very high recall equal to 0.93, which means it is also good at predicting negative links i.e., those links that do not exist and thus, it predicts only a few false negatives, which results in recommending some drugs that are not needed for patients' treatment.

Please note that $\mathrm{PathSim}_{\mathrm{TSTS}}$ has the highest recall equal to 0.96. At this point, we highlight the fact that in healthcare, it is very important to be able to correctly identify all those drugs that were prescribed by the doctor. To this end, recall should be considered as the more important metric than precision. For instance, you may get 100% precision by just recommending a single drug, but at the same time you may have missed to identify all other drugs which were also prescribed (in reality) by the doctor for the patient's treatment (so you fail to recommend needed drugs).

Based on the confusion matrix, which counts true positives, true negatives, false positives, and false negatives of drug recommendations, we can also create the ROC curve for the three models, against also the random model. ROC curve is a plot of the recall (true positive rate) against the false positive rate of our predictions. AUC measures the two-dimensional area underneath the ROC curve from an X-Y axis (0,0) to (1,1).

Figure 4 visualizes the ROC curve performance of each method, such that we can get an idea of the difference that exists between $\mathrm{PathSim}_{\mathrm{STS}}$ and the remaining models in terms of prediction accuracy. As shown in Figure 4, $\mathrm{PathSim}_{\mathrm{STS}}$ gets close to a 93% true positive rate (recall), whereas its false positive rate is 1% (i.e., it predicts almos zero false positives and very few false negatives), which is very close to the ideal case. This is the reason that AUC for $\mathrm{PathSim}_{\mathrm{STS}}$ is so high (AUC = 0.97). The remaining comparison partners are only a little bit better than random guess, which is represented with the dotted diagonal blue line in Figure 4 and attains an AUC equal to 0.5. This is a clear indication that $\mathrm{PathSim}_{\mathrm{STS}}$ is the most successful model and STS is the most informative meta path.

In summary, we note that our aim is to develop models to help doctors screen all possible candidate treatments more comprehensively to identify all relevant medicines for a target patient. To this end, the final decision is always on the medical doctor's side.

VI. Discussion

In our experiments, only one meta-path from the graph is incorporated to benefit the process of similarity calculation. However, someone could argue that the similarity between two drugs should come from more meta paths of the graph. To this end, we could modify the original PathSim algorithm to consider the ratio of the number of connections through which a candidate drug is connected to the target patient compared to the average number of connections of this type of meta path in the whole graph.

Moreover, in our experiments, we have not used a meta path that takes under consideration the "Disease" type of node (i.e., Symptoms and Diagnoses). We note that in our experiments, we further restricted the given information for the patients' studies profiles. That is, we assumed that we are able to predict the next drugs to prescribe for a target patients' study based only on the drugs that were prescribed to them in the past. We assume that since specific drugs are given to cure a list of diseases, then by taking under consideration only the medicines that are included in the Electronic Health Records of the patients, we can easily infer the diseases that they may deal with even if we do not encode into our model

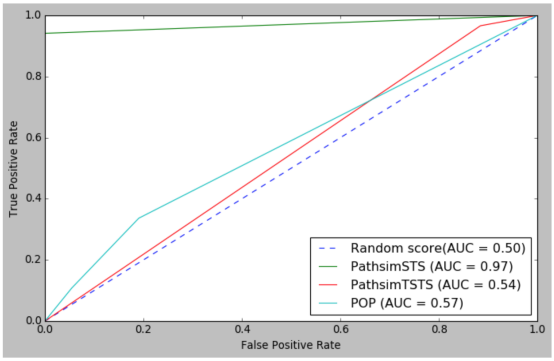

Fig. 4: True positive rate vs. false positive rate on the COVID-19 data set.

the sequence of diseases. Our assumption is verified, as we can effectively derive accurate drug recommendations. However, we should also perform experiments with meta paths that will include the "Diseases" type of node, such that, we gain a better overview of the patients' conditions, which will probably result in an additional improvement of the drug recommendations' effectiveness.

Furthermore, someone could argue that the data set is very small. It is true that for an effective decision making at regional, national, and global levels, more relevant data on patient outcomes are needed. Our system can be easily extended with any new additions in the current database. We hope that in near future, the covid analytics dataset [2], which aggregates data now only from 160 published clinical studies released between December 2019 and April 2020, will be extended with additional medical data information from the COVID-19 pandemic.

Finally, our method is content agnostic and can be applied to different recommendation tasks within the health care domain. That is, our system is not only developed for the COVID-19 pandemic, but the approach is also suitable for different scenarios with medical cases, where different node types are present such as patients, drugs, diseases, health interventions, and even unwanted side effects. For instance, for capturing patient's drug treatment, we would have a graph consisting of Patients (P), who undergo a Treatment (T) using Drugs (D) to target Genes (G) and may have side Effects (E). We can provide a hybrid meta path-based explanation to a medical doctor as follows: "*We recommend for your patient drug D250, because: (i) it was prescribed to 6 other Patients (who have diagnosed the same disease with your patient) and took also similar Drugs with those of your Patient's current treatment (DTPTD), and (ii) It cures/targets similar Genes together with 5 Drugs that your Patient has already taken in his treatment (DGD)*".

VII. Conclusions

We presented a recommender system for the COVID-19 pandemic, which identifies similar patterns among patients of different clinical studies, and recommends them treatments. Our system is also capable to provide explanations along with recommended treatments to assist doctors to understand the reasons behind a treatment recommendation. We performed experiments to identify the most informative meta path with the goal to minimize the false positive and negative prediction rates. As future work, we want to perform a user study to understand the acceptance of our system to medical doctors, and how these methods can actually influence real medical healthcare pathways.

References

[1] F. Behrens, S. Bischoff, P. Ladenburger, J. Rückin, L. Seidel, F. Stolp, M. Vaichenker, A. Ziegler, D. Mottin, F. Aghaei, et al. Metaexp: Interactive explanation and exploration of large knowledge graphs. In *Companion of the The Web Conference 2018 on The Web Conference 2018*, pages 199–202. International World Wide Web Conferences Steering Committee, 2018.

[2] D. Bertsimas, H. Bandi, L. Boussioux, R. Cory-Wright, A. Delarue, V. Digalakis, S. Gilmour, J. Graham, A. Kim, D. Lahlou Kitane, Z. Lin, G. Lukin, M. Li, L. Mingardi, L. Na, A. Orfanoudaki, T. Papalexopoulos, I. Paskov, J. Pauphilet, O. Skali Lami, M. Sobiesk, B. Stellato, K. Carballo, Y. Wang, H. Wiberg, and C. Zeng. An aggregated dataset of clinical outcomes for covid-19 patients, 2020.

[3] G. Jeh and J. Widom. Simrank: a measure of structural-context similarity. In *Proceedings 8th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'2002)*, pages 538–543, Edmonton, Canada, 2002.

[4] S. Lee, S. Park, M. Kahng, and S.-G. Lee. Pathrank: Ranking nodes on a heterogeneous graph for flexible hybrid recommender systems. *Expert Systems with Applications*, 40(2):684–697, 2013.

[5] L. Li, L. Zheng, F. Yang, and T. Li. Modeling and broadening temporal user interest in personalized news recommendation. *Expert Systems with Applications*, 41(7):3168 – 3177, 2014.

[6] J. Pan, H. Yang, C. Faloutsos, and P. Duygulu. Automatic multimedia cross-modal correlation discovery. In *Proceedings of the 10th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD)*, pages 653–658, 2004.

[7] B. Sarwar, G. Karypis, and J. Konstan, J. an Riedl. Item-based collaborative filtering recommendation algorithms. In *Proceedings of the 10th International Conference on World Wide Web (WWW'01)*, pages 285–295, New York, NY, USA, 2001.

[8] Y. Sun, R. Barber, M. Gupta, C. C. Aggarwal, and J. Han. Co-author relationship prediction in heterogeneous bibliographic networks. In *Advances in Social Networks Analysis and Mining (ASONAM), 2011 International Conference on*, pages 121–128. IEEE, 2011.

[9] H. Tong, C. Faloutsos, and J. Pan. Fast random walk with restart and its applications. In *Proceedings 6th International Conference on Data Mining (ICDM'2006)*, pages 613–622, Hong Kong, China, 2006.

[10] S. Yizhou, H. Jiawei, Y. Xifeng, Y. Philip S., and W. Tianyi. Pathsim: Meta path-based top-k similarity search in heterogenuos information networks. In *Proceedings of the VLDB Endowment (VLDB'2011)*, Seattle, Washigton, 2011.