medRxiv preprint doi: https://doi.org/10.1101/2022.11.21.22282571; this version posted November 22, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC 4.0 International license .

Under Review 1–9, 2022

Graph-Based Clinical Recommender: Predicting Specialists Procedure Orders using Graph Representation Learning

Sajjad Fouladvand SAJJADF@STANFORD.EDU
Biomedical Informatics Research, Stanford University, Stanford, CA

Federico Reyes Gomez FRG100@ALUMNI.STANFORD.EDU
Hamed Nilforoshan HAMEDN@CS.STANFORD.EDU
Computer Science Department, Stanford University, Stanford, CA

Morteza Noshad NOSHAD@STANFORD.EDU
Biomedical Informatics Research, Stanford University, Stanford, CA

Jiaxuan You JIAXUAN@CS.STANFORD.EDU
Rok Sosic ROK@CS.STANFORD.EDU
Jure Leskovec JURE@CS.STANFORD.EDU
Computer Science Department, Stanford University, Stanford, CA

Jonathan Chen JONC101@STANFORD.EDU
Biomedical Informatics Research, Stanford University, Stanford, CA

Abstract

An automated medical procedure order recommender can facilitate patient referrals and consultations from primary care providers to specialty care sites. Here, we propose to solve this task using a novel graph representation learning approach. We develop a heterogeneous graph neural network to model structured electronic health records and formulate the procedure recommender task as a link prediction problem. Our experimental results show that our model achieves a 14% improvement in personalized recommendations over state-of-the-art neural network models and existing clinical tools including referral guidelines and checklists.

Keywords: Graph Neural Network, Endocrinology, Medical Consultation.

1. Introduction

Access to medical specialty care is often delayed due to growing limitations in clinicians time and resources leading to higher mortality rates (Prentice and Pizer, 2007). Early prediction of procedures to be ordered during initial outpatient specialty consultation care can facilitate specialist consultations as well as decision making (Chiang et al., 2020; Kim-Hwang et al., 2010). Leveraging artificial intelligence (AI) to solve this task is largely unexplored; even though AI has shown successful application in solving real-world problems in healthcare (Yu et al., 2018).

To this end, Noshad et al. (2021) have proposed an endocrinology procedure recommender using an ensemble of multi-layer perceptron neural networks and collaborative filtering. However, the heterogeneity and structured nature of electronic health records (EHR) can be captured more effectively using graphical models (Park et al., 2022; Choi et al., 2018, 2017).

Graph Convolutional Transformer (GCT) (Choi et al., 2020) maps encounters into a fully connected graph and infers the underlying structure by computing self-attentions

© 2022 S. Fouladvand, F.R. Gomez, H. Nilforoshan, M. Noshad, J. You, R. Sosic, J. Leskovec & J. Chen.

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

medRxiv preprint doi: https://doi.org/10.1101/2022.11.21.22282571; this version posted November 22, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC 4.0 International license .

on the graph connection. Liu et al. (2020) addressed the high visibility (Li et al., 2018) of hub nodes such as demographic nodes and showed the effectiveness of modeling EHR data into heterogeneous graphs. Further, heterogeneous graph neural networks (GNN) have been utilized in drug pairs side effect prediction (Zitnik et al., 2018), medical diagnosis prediction (Liu et al., 2021) and medical concept representations (Wu et al., 2021; Vretinaris et al., 2021).

Motivated by Hamilton et al. (2017), Zitnik et al. (2018), and Veličković et al. (2018) we propose a novel GNN based framework to provide personalized procedure order recommendations prior to or during patients initial specialty care visits. Note, here we use the terminology 'order' to refer to the surgical/follow-up procedures ordered by physicians. This work is part of a larger body of work to develop digital specialty consultation systems to expand the access to quality medical expertise. In this paper, we focus on referrals to endocrinology as one of the highest demand and use patients historical structured EHR data.

2. Materials and Methods

2.1. Data

Our data includes all outpatients referred by Stanford primary care providers to the Stanford endocrinology clinic between January 2008 and December 2018. Use of this data for this study was approved as an exempt protocol by Stanford Institutional Review Board. We only included patients' first visit with their endocrinologist within four months of their referral dates. Our final data set include 6,821 referrals.

We denote the list of patient referrals as $\boldsymbol{P} = \{p_1, \ldots, p_n\}$ in which n is the number of patient referrals. Each patient referral p_i constitutes a tuple $(t_i, \boldsymbol{D}^i, \boldsymbol{O}^i, \boldsymbol{L}^i, \boldsymbol{Y}^i)$, where t_i is referral's date and $\boldsymbol{D}^i \in \mathbb{R}^{10}$, $\boldsymbol{O}^i \in \mathbb{R}^{60}$, and $\boldsymbol{L}^i \in \mathbb{R}^{300\times 3}$ are multi-hot encoded vectors representing diagnoses codes, procedure orders, and lab results for p_i prior to t_i. We used a two month look back window for lab results and procedures. Each lab result was converted to a vector with three elements indicating (a) if p_i has had the lab result, (b) if the result was high, and (c) if the result was low. $\boldsymbol{Y}^i$ is a multi-hot encoded vector representing the procedures ordered by the specialist during patient's special care visit.

Our final feature set includes 370 features: 10 most common endocrinology related diagnoses codes, 300 lab result features, and 60 procedures. The target set includes 60 procedure orders (see variable names in Appendix A).

2.2. Proposed Method

2.2.1. Graph Structure

We modeled patients EHR data set into a heterogeneous graph neural network $G = (V, E)$ (see Figure A1 in Appendix A). V contains two node types: patient referral nodes $\{g_1^p, ..., g_{|\boldsymbol{P}|}^p\}$, and procedure order nodes $\{g_1^o, ..., g_{|\boldsymbol{O}|}^o\}$. Each patient node g_i^p is assigned a 310-dimensional feature vector consisting of concatenation of $\boldsymbol{D}^i$ and $\boldsymbol{L}^i$ and each procedure order node g_i^o are associated with one-hot encoding of the entity IDs (I_{s_i} and I_{o_i}, respectively).

Edge set E contains two edge types. 'ordered-with' edges with edge labels set to 0 that are edges between patient nodes and the procedures they have done before t_i, and 'ordered-with' edges with edge label set to 1 that connect the patients with the procedures that their specialist ordered during the specialty care visit after t_i. Note, 'ordered-with' edges with edge labels equal to 1 that represent specialist's orders after referral date were not used during training and were only used in the prediction phase as we are aiming to predict procedure orders

medRxiv preprint doi: https://doi.org/10.1101/2022.11.21.22282571; this version posted November 22, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC 4.0 International license .

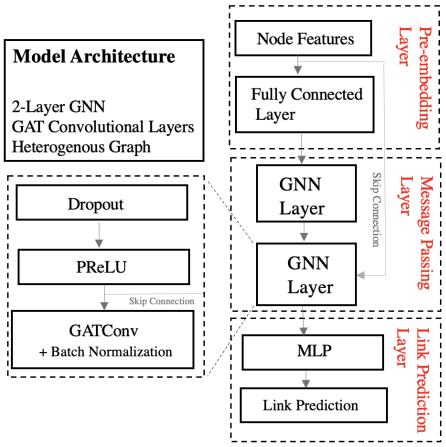

Figure 1: General architecture of our data, graph, and our proposed model.

after t_i. We formulate this task as binary link prediction of the existence of 'ordered-with' edges between a patient and an order. Further, node degree, node clustering coefficient and centrality transformations were applied to add synthetic features to each node feature vector. While the model can learn those features on its own, we added them to help the model focus on learning other features. We apply a different Graph Convolutional layers with independent parameters to each message type of (head, relation, tail) and aggregate embeddings across all node types. The same graph attention mechanism was applied to all node types.

2.2.2. Message Passing and Graph Attention

Figure 1 shows our proposed architecture. A fully connected layer with hidden size of 128 was used to map each node feature vector to pre-embedding vectors. Distinct fully connected layers were used for each node type. Two message passing layers was used each consisting of a dropout layer, a PReLU activation function, and a graph convolutional layer.

A custom heterogeneous graph attention layer was used using 1-head attention mostly following the structure of the original graph attention networks (Veličković et al., 2018), with the following modifications: 1) we applied fully connected layers with batch normalization to the node embeddings and the neighbor embeddings,and 2) we aggregated neighbor embeddings using the attention mechanism and concatenated the aggregated embedding to the current node's embedding. This is then passed into a fully connected layer that reduces this down to a single output embedding followed by a batch normalization operation. Equation (1) shows our massage passing function

$$\begin{aligned} x_v^{(1)} &= \mathrm{MLP}(x_v^{(0)}) \\ x_v^{(2)} &= \mathrm{GATConv}(\mathrm{PReLU}(\mathrm{Dropout}(x_v^{(1)}))) \\ x_v^{(3)} &= \mathrm{GATConv}(\mathrm{PReLU}(\mathrm{Dropout}(x_v^{(2)})) \\ &\quad + x_v^{(0)}) \\ x_v^{(4)} &= \mathrm{MLP}(x_v^{(3)}) \end{aligned} \tag{1}$$

and Equation (2) shows the GATConv update function

$$\begin{aligned} \mathrm{aggr} &= \Sigma_{v_o \in \mathcal{N}(v)} \alpha_{v_o} * \mathrm{MLP}(x_{v_o}^{(k)}) \\ x_v^{(k+1)} &= \mathrm{MLP}(\mathrm{aggr} + \mathrm{MLP}(x_v^{(k)})) \end{aligned} \tag{2}$$

Where α_{v_o} is the 1 head GAT attention score calculated for v_o, $\mathcal{N}(v)$ is neighbors of v, and $x_v^{(0)}$ represents the node features of node v.

The final predictions on existence of an 'ordered-with' edge e_{ij} between nodes g_i^p and g_j^o is inferred by concatenating their node embeddings and passing that through a fully connected two-layer perceptron, a batch normalization, a ReLU activation, and a final fully connected layer that outputs 2-dimensional logit vectors that are converted to final binary predictions using a softmax function.

medRxiv preprint doi: https://doi.org/10.1101/2022.11.21.22282571; this version posted November 22, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC 4.0 International license .

The formula for the link prediction head is as follows:

$$p = \mathrm{FC}(\mathrm{ReLU}(\mathrm{BN}(\mathrm{MLP}([x^{(4)}_{g^p_i}; x^{(4)}_{g^o_j}])))) \in \mathbb{R}^2 \quad (3)$$

where BN refers to Batch Normalization and the first value corresponds to the probability that the edge exists and the second that it doesn't.

3. Experimental Results

We used transductive disjoint training with a 1:4 positive:negative sampling using PyG (Fey and Lenssen, 2019; pyg). Adam optimizer with a learning rate of 1e-3, weight decay of 5e-4, and 400 epochs were used to train the model. Further, dropout of 0.2, and pre-embedding sizes, hidden sizes, and final embedding sizes of 128 were used. Our GNN model was tested on predictions made on all 'ordered-with' edges between a patient and an order placed during specialty visit.

Table 1 compares prediction results of our proposed model (CR-GNN) with the baselines presented by Noshad et al. (2021) including fully connected multi-layer neural network (Diagnostic Model), a collaborative filtering auto-encoder (AE), singular value decomposition (SVD), probabilistic matrix factorization (PMF), an aggregate neural networks (Aggregated ANN), and an ensemble model (Ensemble Model) that uses a multi-layer neural network to combine the outputs of the diagnostic model, the collaborating filtering auto-encoder and the specialists identifiers as a separate input signal.

Our proposed model can predict endocrinology specialty procedures more effectively (ROC-AUC=0.91) compared to all models proposed by Noshad et al. (2021) (best ROC-AUC=0.80). Further, our model showed significantly higher precision at recalls 0.5, 0.4 and 0.3 compared to all baseline models. Note, we used the same data as the data that were used in Noshad et al. (2021) except we removed features related to the specialists that patients were referred to, because specialists' information can add bias to the model. Additionally, we trained and tested our proposed model using all features used in Noshad et al. (2021) including specialists' information as well. This didn't significantly affect our model's performance. The model's ROC-AUC using all features including specialists' information was 0.912 and precisions at recalls 0.5, 0.4 and 0.3 were 0.60, 0.65, and 0.70, respectively.

Table 1: Performance of endocrinologist procedure order prediction models.

Model	**AUC ROC**	**P@R 0.50**	**P@R 0.40**	**P@R 0.30**
Diagnostic Model	0.65	0.33	0.42	0.46
AE	0.73	0.23	0.33	0.49
PMF	0.62	0.22	0.31	0.43
SVD	0.74	0.23	0.33	0.50
Aggregated NN	0.73	0.31	0.41	0.53
Ensemble Model	0.80	0.37	0.47	0.57
CR-GNN	**0.91**	**0.62**	**0.68**	**0.73**

Figure 2 compares precision-recall curves for our proposed method with the baselines. The precision for the baseline models are higher than our proposed model toward the tail of the curves. However, our proposed model has higher precision compared to all baselines over a wide range of recalls including recalls 0.3, 0.4 and 0.5. This can provide clinicians with the ability to adjust our model based on their preferred precision-recall trade off. Further, our proposed model has significantly higher precision than all baseline precisions at recalls close to the recall of existing clinical guideline and checklist.

medRxiv preprint doi: https://doi.org/10.1101/2022.11.21.22282571; this version posted November 22, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC 4.0 International license .

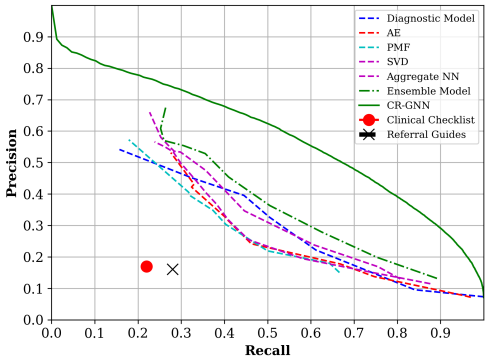

Figure 2: Comparing precision-Recall curve of our proposed model with the baselines.

4. Conclusion

In conclusion, embedding graph neural network models into clinical care can improve digital specialty consultation systems and expand the access to quality medical expertise.

There are some limitations in this work that should be considered before using our proposed model. The proposed model is limited to trandsuctive learning and patients structured data.

References

PyG. https://www.pyg.org/. Accessed: 2022-09-31.

Jonathan Chiang, Andre Kumar, David Morales, Divya Saini, Jason Hom, Lisa Shieh, Mark Musen, Mary K Goldstein, and Jonathan H Chen. Physician usage and acceptance of a machine learning recommender system for simulated clinical order entry. *AMIA Summits on Translational Science Proceedings*, 2020:89, 2020.

Edward Choi, Mohammad Taha Bahadori, Le Song, Walter F Stewart, and Jimeng Sun. Gram: graph-based attention model for healthcare representation learning. In *Proceedings of the 23rd ACM SIGKDD international conference on knowledge discovery and data mining*, pages 787–795, 2017.

Edward Choi, Cao Xiao, Walter Stewart, and Jimeng Sun. Mime: Multilevel medical embedding of electronic health records for predictive healthcare. *Advances in neural information processing systems*, 31, 2018.

Edward Choi, Zhen Xu, Yujia Li, Michael Dusenberry, Gerardo Flores, Emily Xue, and Andrew Dai. Learning the graphical structure of electronic health records with graph convolutional transformer. In *Proceedings of the AAAI conference on artificial intelligence*, volume 34, pages 606–613, 2020.

Matthias Fey and Jan Eric Lenssen. Fast graph representation learning with pytorch geometric. *arXiv preprint arXiv:1903.02428*, 2019.

Will Hamilton, Zhitao Ying, and Jure Leskovec. Deepsnap. *Advances in neural information processing systems*, 30, 2017.

Judy E Kim-Hwang, Alice Hm Chen, Douglas S Bell, David Guzman, Hal F Yee, and Margot B Kushel. Evaluating electronic referrals for specialty care at a public hospital. *Journal of general internal medicine*, 25(10):1123–1128, 2010.

Qimai Li, Zhichao Han, and Xiao-Ming Wu. Deeper insights into graph convolutional networks for semi-supervised learning. In *Thirty-Second AAAI conference on artificial intelligence*, 2018.

Zheng Liu, Xiaohan Li, Hao Peng, Lifang He, and S Yu Philip. Heterogeneous similarity graph neural network on electronic health records. In *2020 IEEE International Conference on Big Data (Big Data)*, pages 1196–1205. IEEE, 2020.

medRxiv preprint doi: https://doi.org/10.1101/2022.11.21.22282571; this version posted November 22, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC 4.0 International license .

Zheng Liu, Xiaohan Li, Zeyu You, Tao Yang, Wei Fan, and Philip Yu. Medical triage chatbot diagnosis improvement via multi-relational hyperbolic graph neural network. In *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 1965–1969, 2021.

Morteza Noshad, Ivana Jankovic, and Jonathan H Chen. Clinical recommender algorithms to simulate digital specialty consultations. In *PACIFIC SYMPOSIUM ON BIOCOMPUTING 2022*, pages 290–300. World Scientific, 2021.

Sungjin Park, Seongsu Bae, Jiho Kim, Tackeun Kim, and Edward Choi. Graph-text multi-modal pre-training for medical representation learning. In *Conference on Health, Inference, and Learning*, pages 261–281. PMLR, 2022.

Julia C Prentice and Steven D Pizer. Delayed access to health care and mortality. *Health services research*, 42(2):644–662, 2007.

Petar Veličković, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Liò, and Yoshua Bengio. Graph Attention Networks. *International Conference on Learning Representations*, 2018. URL https://openreview.net/forum?id=rJXMpikCZ.

Alina Vretinaris, Chuan Lei, Vasilis Efthymiou, Xiao Qin, and Fatma Özcan. Medical entity disambiguation using graph neural networks. In *Proceedings of the 2021 International Conference on Management of Data*, pages 2310–2318, 2021.

Tong Wu, Yunlong Wang, Yue Wang, Emily Zhao, and Yilian Yuan. Leveraging graph-based hierarchical medical entity embedding for healthcare applications. *Scientific reports*, 11(1):1–13, 2021.

Kun-Hsing Yu, Andrew L Beam, and Isaac S Kohane. Artificial intelligence in healthcare. *Nature biomedical engineering*, 2(10): 719–731, 2018.

Marinka Zitnik, Monica Agrawal, and Jure Leskovec. Modeling polypharmacy side effects with graph convolutional networks. *Bioinformatics*, 34(13):i457–i466, 2018.

Appendix A. Features

Table A1: Diagnosis features.

Diagnosis name
Diabetes mellitus Type I or II
Hypercalcemia
Hyperlipidemia
Hypothyroidism
Hyperthyroidism
Osteopenia
Osteoporosis
Thyroid cancer
Thyroid nodule
Obesity

medRxiv preprint doi: https://doi.org/10.1101/2022.11.21.22282571; this version posted November 22, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC 4.0 International license .

Table A2: Procedure features.

Procedure name	Procedure name	Procedure name	Procedure name
TSH	T4, FREE	Vitamin d, 25-hydroxyvitamin	Metabolic panel, comprehensive
Hemoglobin A1C	Parathyroid hormone	Metabolic panel, basic	Cortisol, serum
ANTI - TPO AB	Phosphorus, serum/plasma	Albumin with creatinine, urine (random)	TSH W/ REFLEX FT4
Prolactin	US thyroid	Creatinine, urine (timed)	Calcium, urine (timed)
FSH	T3, FREE	Adrenocorticotropic hormone (ACTH)	Calcium, serum/plasma
Lab unlisted 1	Lipid panel with direct ldl	Luteinizing hormone	Lipid panel with calculated LDL
T3, total	Bone alkaline phosphatase, serum	IGA ANTI TTG	C - peptide, serum
DXA adult	Testosterone, total, bio, free	HGB A1C W/ EST mean glucose	Magnesium, serum/plasma
Thyroid-stimulating immunoglobulin (TSI)	Collagen type i c-telopeptide (ctx)	Albumin, serum/plasma	TSH and free T4
Thyroglobulin and tgab comprehensive	Vitamin B12	Thyroglobulin ab ultra-sensitive	CBC with differential
Testosterone	Dehydroepiandrosterone, sulfate	Estradiol	CBC W/O DIFF
Creatinine, serum/plasma	Bone density adult	Insulin-like growth factor 1	Metanephrines fractionated free, plasma
Aldosterone	Free cortisol, urine (timed)	Hepatic function panel a	Thyroglobulin
US head neck soft tissue	Cortisol, AM	IGA, SERUM	Urine protein immunofixation electrophoresis
Renin	ALT, serum/plasma	Ferritin	Thyroid stimulating immunoglobulin

medRxiv preprint doi: https://doi.org/10.1101/2022.11.21.22282571; this version posted November 22, 2022. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC 4.0 International license .

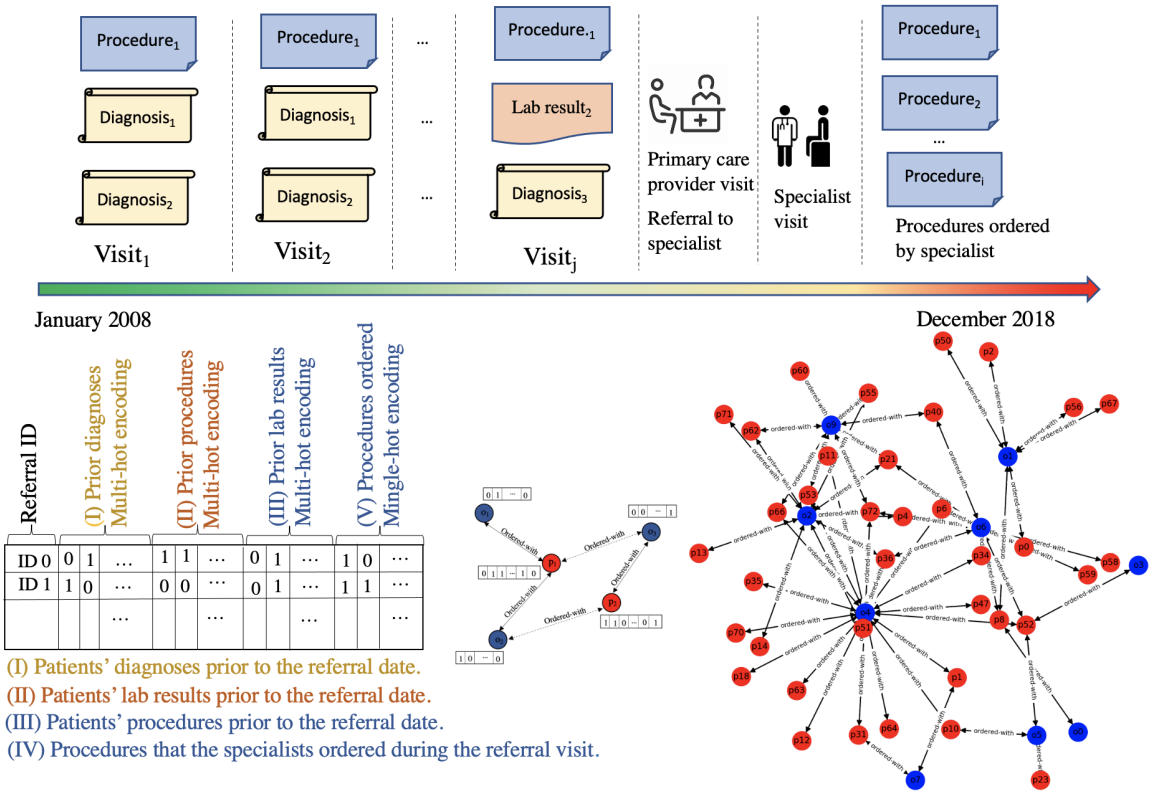

Figure A1: Graph construction.

medRxiv preprint doi: https://doi.org/10.1101/2022.11.21.22282571; this version posted November 22, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC 4.0 International license .

Table A3: Lab result features.

Lab Test	**Lab Test**	**Lab Test**	**Lab Test**
Creatinine, Ser/Plas	Calcium, Ser/Plas	Potassium, Ser/Plas	Sodium, Ser/Plas
Glucose, Ser/Plas	Chloride, Ser/Plas	CO2, Ser/Plas	ALT (SGPT), Ser/Plas
Albumin, Ser/Plas	WBC	RBC	AST (SGOT), Ser/Plas
Hematocrit	Platelet count	Hemoglobin	MCV
RDW	MCHC	MCH	Alk P TASE, Total, Ser/Plas
Protein, Total, Ser/Plas	Globulin	Anion Gap	eGFR
TSH	Total Bilirubin	Triglyceride, Ser/Plas	BUN, Ser/Plas
HDL Cholesterol	Urea Nitrogen,Ser/Plas	Hemoglobin A1c	Total Bilirubin, Ser/Plas
EOS, ABS	Cholesterol/HDL Ratio	Monocyte, Absolute	Lymphocyte, Absolute
Eosinophil, Absolute	Neutrophil, Absolute	eGFR for African American	NEUT, ABS
MONO, ABS	LYM, ABS	Non-HDL Chol, Calc	Cholesterol, Total
Magnesium, Ser/Plas	LDL (Calculated)	INR	Prothrombin Time
Glucose by Meter	25-Hydroxy D, Total	pH	Direct LDL Chol
Phosphorus, Ser/Plas	Part. Thromboplastin Time	Conjugated Bili	C-Reactive Protein
Unconjugated Bilirubin	HCO3	tCO2	Glucose, Whole Blood
Chloride, Whole Bld	Potassium, Whole Bld	Sodium, Whole Blood	Hct (Est)
PCO2 (v), ISTAT	HCO3 (v), ISTAT	TCO2 (v), ISTAT	O2 Saturation, ISTAT (Ven)
PO2 (v), ISTAT	Calcium Ionized	LDH, Total, Ser/Plas	pCO2 (a)
pO2 (a)	pH (a)	ctO2 (a)	Base Excess (vt)
Lymphocytes	Calcium, Ionized	HgB	HCO3 (a), ISTAT
pCO2 (a), ISTAT	PO2 (a), ISTAT	PH (a), ISTAT	O2 Saturation, ISTAT
Hgb(Calc), ISTAT	pCO2 (v)	O2 Saturation (v)	pO2 (v)
ctO2 (v)	tHB	TCO2 (a), ISTAT	Fibrinogen
Myelocytes	Lym, ABS (man diff)	Seg neutrophils	Base Deficit (vt)
D-Dimer	Hct, ISTAT	Calcium,Ion, ISTAT	TCO2, ISTAT