Link prediction in disease networks

Hankyu Jang
University of Iowa
hankyu-jang@uiowa.edu

ABSTRACT

Predicting potential links between nodes in a network is one of the major research areas in social network analysis. These top-notch techniques have been adopted in various types of networks in different fields such as social media, transportation, or co-authorship networks, but not much in the medical domain. Our objective is to explore link prediction methods on disease networks to find disease pairs that are similar to each other. Specifically, we generated a disease-disease network from clinical trials and connected the diseases if there exists an overlapping treatment between the two diseases. In this setting, the predicted edges represent that the two diseases could potentially be treated with a similar set of drugs. We propose this novel approach of drug repositioning opportunities on 165 pairs of diseases from the result of the link prediction.

KEYWORDS

link prediction, disease networks, clinical trials, drug repositioning

ACM Reference Format:
Hankyu Jang. . Link prediction in disease networks. In . ACM, New York, NY, USA, 6 pages.

1 INTRODUCTION

We are living in an era flooded by a large amount of data. Unlike treating data instances separately from other instances in the dataset, network science assumes that there are relationships between instances as well as features in the dataset. There are many different kinds of networks: social network, citation network, transportation network, protein interaction network, and disease networks to name a few. In a social network, perhaps Facebook, each Facebook user is connected to other users if they are friends with each other. In this social network, there is a high chance of two users who are not friends with each other to know each other in real life if they share many mutual friends. This kind of closure property in a network is called a triad and the clustering coefficient, a frequently used network property, captures the proportion of the three nodes that form the triad in the network.

Link prediction is a widely studied area in social network analysis. Link prediction algorithms fall into two categories: 1) unsupervised learning and 2) supervised learning. Methods in unsupervised learning category use one measurement to measure some similarity between pair of nodes. Then, these methods give rank on all possible pair of nodes in the network regardless of the existence of an edge between the pair of nodes. Then, the methods choose the top k disconnected node pairs based on the ranking as the prediction of possible links.

Supervised learning methods also use measurements that capture some similarity between node pairs. However, unlike unsupervised methods, supervised learning uses these measurements as features of node pairs. In other words, these methods generate a set of features for each node pair and then set a binary label for each node pair. If there is a link between the two nodes, there is a True label, otherwise False.

Link prediction methods as mentioned earlier had widely been used to predict links in networks. However, as far as we are concerned, a limited number of studies have utilized these link prediction methods in disease networks. A large number of researchers have conducted studies in disease-related data, but not on the topological information in disease networks. In this study, we propose a novel approach of predicting links in disease networks.

The interpretation of the predicted links in disease networks would vary by the context. If links in the disease network represent the similarity of drugs that are used to treat the disease, then the node pair of newly predicted links represent the two disease that could be treated with similar medications. For instance, assume that disease A and disease B do not have an edge between each other, meaning that they do not have common drugs for treatment. If our link prediction model predicts an edge between disease A and disease B, it would mean that the two diseases would have a potential of being treated with a common drug; either a drug that is used for treating disease A could be used for disease B or vice versa. Our results would be useful for pharmaceutical companies that tried to find a novel use of drugs in treating other diseases.

2 RELATED WORKS

Recently, there have been studies utilizing clinical trials to find drug repositioning opportunities. Su and Sanger [11] proposed an approach of drug repositioning from randomized clinical trials. For each arm, they extracted the total number of patients in the arm and those that are affected by serious events. If the ratio of the infected patients in the control arm exceeds the number of patients in the drug arm, they assumed that the drug might have some positive effect on treating the adverse event. They used this logic to discover drugs that may be used for treating other diseases.

Some researchers generated disease networks from the information captured in clinical trials. Haslam and Perez-Breva [6] proposed a method to generate disease-disease network by using the whole clinical trials up to March 27, 2014, comprised of 163,764 trials. For each target disease in clinical trials, they extracted drugs that are used as an experimental treatment of the disease to create a disease-drug bipartite network. Then, if any two diseases share the same drugs, they connected the two diseases thereby generating a disease-disease network. For a given disease in a test set, they made predictions about each drug among 7349 possible drugs, using collaborative filtering. They used a cosine similarity metric to calculate the similarity of the two diseases with regards to the

drugs that used to treat the conditions. They showed the potential of suggesting drugs for diseases using disease-drug pairs.

There are many other sources of disease information where researchers could generate disease networks. Zhou et al. [13] constructed human symptoms-disease network from biomedical literature databases. They extracted disease-symptom relationships from Medical Subject Headings (MeSH) from Pubmed. The network they generated is reliable since the network shows high similarity to the Human Phenotype Ontology (HPO) [10] which is the manually curated disease-disease network based on the symptoms the diseases share.

Although disease-related networks are abundant, there is not much of work been conducted in analyzing these networks in network science point of view. On 2017, Lu et al. [8] utilized network topological information in drug repositioning. They generated a chemical-protein interaction network from the data retrieved from Manually Annotated Target and Drug Online Resource (MATADOR) database. They used link prediction in this network and showed the predicted links by using network topological features yielded higher precision compared to other link prediction models that utilize additional information of the chemicals and proteins.

More recently on 2018, Davazdahemami and Delen [4] applied link prediction to predict adverse drug events (ADE) for drugs. They generated a drug-ADE network from MEDLINE biomedical articles and enriched the network by adding links between drugs based on the information on drugs' target proteins. They generated similarity features on the drug-ADE links based on network topology and applied machine learning techniques to predict drug-ADE patterns. Their work is by far the most similar approach to the methodology we are presenting in our paper.

3 METHODS

In our paper, we extracted disease-drug pairs from clinical trials, generated disease network, and then applied link prediction on the network. We started by generating a disease-drug bipartite network from the data that Haslam and Perez-Breva [6] provided in their supplemental materials. We used two data tables that Haslam and Perez-Breva provided: 1) trial disease data which contains the clinical trial number, target disease, and the MESH term of the disease, and 2) trial drugs data that has the same format but with that of drugs. By joining these two tables by the clinical trial number (72,066 trials), we generated a disease-drug bipartite network.

3.1 Disease-Drug network

The bipartite network has two types of nodes: diseases and drugs. There are 8168 diseases (2,788 unique) treated with 60,668 drugs (5,388 unique) and 167,172 disease-drug pairs. Among those pairs, 20,293 had placebo as the drug node. Since we are not interested in placebos, we removed the placebo node from the bipartite network, which left us with 146,879 (58,921 unique) drug-disease pairs.

Table 1 shows the top 20 most observed disease-drug pairs in the disease-drug network. Each row represents how many times a disease-drug pair appeared in different clinical trials. For instance, 1) Metformin was used to treat Diabetes Mellitus, Type 2 in 454 different clinical trials and 2) Ritonavir was used to treat HIV Infections in 405 clinical trials.

Table 1: Top 20 disease-drug pairs from clinical trials

Disease	Drug	Trials
Diabetes Mellitus, Type 2	Metformin	454
HIV Infections	Ritonavir	405
Pancreatic Neoplasms	gemcitabine	379
Breast Neoplasms	Paclitaxel	376
Breast Neoplasms	Cyclophosphamide	374
Influenza, Human	Influenza Vaccines	372
Breast Neoplasms	trastuzumab	371
HIV Infections	Zidovudine	358
Hepatitis C, Chronic	Ribavirin	348
Breast Neoplasms	docetaxel	334
Lymphoma	Cyclophosphamide	330
Multiple Myeloma	Dexamethasone	312
Colorectal Neoplasms	Fluorouracil	306
HIV Infections	Lamivudine	289
Leukemia	Cyclophosphamide	280
Lymphoma	rituximab	280
Multiple Myeloma	bortezomib	277
Colorectal Neoplasms	oxaliplatin	268
Colorectal Neoplasms	Leucovorin	266
Breast Neoplasms	Doxorubicin	255

Among these 20 disease-drug pairs, Breast Neoplasms, Lymphoma, and Leukemia are treated using Cyclophosphamide. In this paper, we regard the diseases to be similar if there are the same drugs that are used to treat those diseases. Hence, from this bipartite network, we generated a disease-disease network if any of the two diseases shared at least one drug in their treatment.

3.2 Disease-Disease Network

Table 2 show the graph statistics for the network we used for link prediction. The disease network generated from clinical trials has 2,746 nodes and 514,189 links. Note that this graph has a clustering coefficient of 0.745, which is an unusually high value for a network. This high value indicates that triads are common property in this network. This network has a negative degree assortativity of -0.254, meaning that high degree nodes tend to form edges with low degree nodes. From the 2,746 nodes in the network, 2,730 nodes are connected. For our future analysis, we only used the nodes and edges in this giant connected component.

3.3 Feature Set

For each node pair in the network, we need to generate features that capture some similarity between the two nodes. From the topological features of the network, we computed the following five features: Common Neighbor, Jaccard Coefficient, Preferential Attachment, Adamic Adar, and Resource Allocation.

3.3.1 Common Neighbor. The most direct way of using a similarity measure of nodes x and y is to find the number of neighbors that the two nodes have in common. Newman used this measure in 2001 on the collaboration network [9]. For a node x, let $\tau(x)$ denote the set of the neighboring nodes of x. Let $s(x, y)$ denote the similarity measure of the node pair (x, y).

Table 2: Graph statistics on disease-disease network

Properties	Disease Network
num vertices	2746
num edges	514189
mean degree	374.5
max_degree	2200
std. dev. degree dist.	375.254
clust. Coeff	0.745
num. components	16
assortativity	-0.254
num. vertices giant comp	2730
num edges giant comp.	514188

$$s_{xy} = |\tau(x) \cap \tau(y)| \tag{1}$$

3.3.2 Jaccard Coefficient. Jaccard Coefficient proposed this index in the early 1900s [5]. This index captures the number of common neighbors of x and y divided by the number of total neighbors of x and y.

$$s_{xy} = \frac{|\tau(x) \cap \tau(y)|}{|\tau(x) \cup \tau(y)|} \tag{2}$$

3.3.3 Preferential Attachment. This measure is adapted in the link prediction by Chen, Li, and Huang on 2005 [3] where the original idea is from the scale-free network by Barabasi and Albert [2]. In the formula below, $k(x)$ denote the degree of the node x.

$$s_{xy} = k(x) \times k(y) \tag{3}$$

3.3.4 Adamic Adar. Adamic and Adar proposed Adamic Adar measure in 2003 [1]. This measure holds the sum of inverse-log degree of the common neighbors of x and y by assigning more weights to the neighbors with a lower degree.

$$s_{xy} = \sum_{z \in \tau(x) \cap \tau(y)} \frac{1}{\log k(z)} \tag{4}$$

3.3.5 Resource Allocation. Zhou, Lu, and Zhang introduced the Resource Allocation algorithm in 2009 [12] which the idea is parallel with the Adamic Adar index but without the log.

$$s_{xy} = \sum_{z \in \tau(x) \cap \tau(y)} \frac{1}{k(z)} \tag{5}$$

3.4 Prepare Dataset from Network

Link prediction could be viewed as a binary classification problem: if there is a link between two nodes, the label of the node pair is True, otherwise False. In other words, edges that are present in the network would have a True label. For these node pairs, we can compute the features mentioned previously. Node pairs that are not neighbors in the network would have a False label. However, since networks are sparse, the dataset generated in this manner would be highly imbalanced.

3.4.1 Class Imbalance. There are several ways to handle this class imbalance problem in link prediction. One way is to only focus on nodes that are 2-hops or 3-hops away and only include these node pairs in the network. One downside of this approach is that this method limits the prediction capability to nodes that are at most 3-hops away from each other. Also, even if we only consider node pairs that are close to each other, the dataset using this approach could still be imbalanced.

Another way is to handle class imbalance is to randomly sample label 0 node pairs so that the class would be balanced. We used this approach in our experiment.

3.4.2 Dataset for Link Prediction. Table 3 shows a few instances of the dataset that we have processed for link prediction. When computing these features, we treated the graph as unweighted graph regardless of the edge weights that denote the number of drugs that overlap in treating two diseases. Note that the disease names are stored as the Mesh terms. Looking at the first pair, 'D001017' is a Mesh term for Aortic Coarctation and 'D009369' is a Mesh term for Neoplasms. These diseases have a True label, meaning that they have common drugs that are used for the treatment. There are about one million disease pairs in the dataset. Among those pairs, roughly half of the pairs have True label, and the remaining half have False label.

Table 3: Dataset for link prediction

pair	RA	JC	AA	PA	CN	label
(D001017 D009369)	0.214	0.078	25.240	380600	172	True
(D001017 D009765)	0.214	0.087	25.239	341675	172	True
(D001017 D000163)	0.214	0.138	25.230	215385	172	True
(D001017 D015658)	0.214	0.079	25.240	375410	172	True
(D001017 D009103)	0.214	0.105	25.235	282509	172	True
(D018318 D007870)	0.006	0.020	1.218	33948	9	False
(D018318 D010148)	0.104	0.142	15.944	199998	113	False
(D018318 D010255)	0.196	0.324	24.428	115866	167	False
(D018318 D014591)	0.241	0.395	28.199	112545	191	False
(D018318 D000347)	0.006	0.021	1.111	7011	8	False

3.5 Classification Algorithms

For classification algorithms, we used multi-layer perceptron (MLP) and random forest (RF). For MLP, we trained the model with one hidden layer with 32 neurons on the dataset using mini-batch gradient descent with a batch size of 20 for ten epochs by setting aside 20 percent of the data per epoch. When training RF classifier, we varied two parameters: 1) the number of trees from 2 to 32 and 2) minimum split nodes from 2 to 32. For each parameter setting, we trained the model using 5-fold cross-validation and selected the parameter setting that produces the highest AUC value. After training the model, we predicted the links on the whole dataset.

4 RESULTS

Table 4 shows the result of link prediction using MLP and RF models are both very high. The accuracy, precision, recall and AUC value for MLP range in 0.945 to 0.96 whereas the metrics for RF is over 0.999.

Table 4: Performance of link prediction

Metrics	MLP	RF
Accuracy	0.9518	0.9995
Precision	0.9451	0.9997
Recall	0.9594	0.9993
AUC	0.9518	0.9995

5 DISCUSSION

There were 165 false positives in the link prediction result of RF and all the disease pairs that yield false positives in RF also yield false positive in MLP. These false positives indicate the prediction of links that currently do not exist in the network. Since the prediction accuracy of the link prediction is reasonably high, we can view these false positives as potential candidates of diseases that could be treated with the same drugs.

The table in the appendix contains disease pairs that yield 165 false positives in the link prediction. Each row is a disease pair with the name of the disease and its MESH term. The pairs denote diseases that have high similarity values computed from the disease graph but the drugs that are used to treat them do not overlap. Hence, these disease pairs pose an opportunity for drug-repositioning; the drugs that are used to treat one disease may as well be used to treat the other disease.

The disease pair in the first row of the table is “mycosis fungoides" and “Desmoplastic Small Round Cell Tumor." Both of these diseases are categorized as rare cancers by NIH[1]. Mycosis fungoides is a disease where a type of white blood cell becomes cancerous and affect the skin, whereas Desmoplastic Small Round Cell Tumor is a type of cancer that starts in the abdomen.

The disease pair in the second row of the table is “Pulmonary Embolism" and “Asphyxia." When we queried the articles that contain both of these two diseases, 40 papers were retrieved[2]. Asphyxia is a condition which could lead to unconsciousness or even death due to the deprived of oxygen whereas Pulmonary Embolism causes shortness of breath, chest pain, and cough due to the blood clots blocking arteries in the lung. Hence, there is some correlation between the two diseases [7].

6 LIMITATION

One of the major limitations of our work is that the network is constructed based on the information we gathered in the clinical trials. The disease networks we generated does not incorporate other information about diseases but only whether there were the same drugs that were tested among them.

The predicted links show that the two diseases are similar, but are not capable of recommending specific disease-drug pair for drug repositioning. When preparing the dataset for link prediction, we did not divide the network into training and test set. Due to this, the RF model may have overfitted to the dataset.

The similarity features are computed by treating the graph as an unweighted graph, and this may have affected the prediction results. However, by treating the edges as unweighted, our prediction models may have been able to predict links for underrepresented nodes; both the diseases that are not tested much in the clinical trials and those that do not have much of the drugs that were treated for the disease.

[1]https://rarediseases.info.nih.gov/diseases/diseases-by-category/1/rare-cancers

[2]https://www.ncbi.nlm.nih.gov/pubmed/?term=%22Pulmonary+Embolism%22+%22Asphyxia%22

7 CONCLUSION

In this paper, we generated disease-disease network from clinical trials that represent whether two diseases were treated with the same drugs. Then we applied link prediction on this network and found 165 disease pairs that have the potential of being treated using the drugs that were tested on them.

8 FUTURE WORKS

We could strengthen the disease-disease network by incorporating other sources of information on diseases. The relationship between diseases and genes or diseases and proteins could enrich the network as well as their co-appearance in articles. If we generate a disease-disease network with edges having these meaning, then we can add drug nodes to the disease-disease network, compute similarity features on disease-drug pairs, and then apply link prediction on the disease-drug pairs. This approach would be more suitable for drug repositioning given that the approach would output potential disease-drug pairs.

Validation of the predicted links may be precarious since a large number of disease pairs hamper us to validate the results manually. One way of validating the results would be by using information retrieval platforms. Once we index medical articles, we can retrieve documents by querying the predicted disease pairs from the link prediction. Based on the number of documents retrieved, we can validate the strength of those predicted pairs.

REFERENCES

[1] Lada A Adamic and Eytan Adar. 2003. Friends and neighbors on the web. *Social networks* 25, 3 (2003), 211–230.

[2] Albert-László Barabási and Réka Albert. 1999. Emergence of scaling in random networks. *science* 286, 5439 (1999), 509–512.

[3] Hsinchun Chen, Xin Li, and Zan Huang. 2005. Link prediction approach to collaborative filtering. In *Proceedings of the 5th ACM/IEEE-CS Joint Conference on Digital Libraries (JCDL'05)*. IEEE, 141–142.

[4] Behrooz Davazdahemami and Dursun Delen. 2018. A chronological pharmacovigilance network analytics approach for predicting adverse drug events. *Journal of the American Medical Informatics Association* 25, 10 (2018), 1311–1321.

[5] Société Vaudoise des Sciences Naturelles. 1864. *Bulletin de la Société vaudoise des sciences naturelles*. Vol. 7. F. Rouge.

[6] Bryan Haslam and Luis Perez-Breva. 2016. Learning disease relationships from clinical drug trials. *Journal of the American Medical Informatics Association* 24, 1 (05 2016), 13–23. https://doi.org/10.1093/jamia/ocw003 arXiv:http://oup.prod.sis.lan/jamia/article-pdf/24/1/13/8765470/ocw003.pdf

[7] Duan Li, Omar S Mabrouk, Tiecheng Liu, Fangyun Tian, Gang Xu, Santiago Rengifo, Sarah J Choi, Abhay Mathur, Charles P Crooks, Robert T Kennedy, et al. 2015. Asphyxia-activated corticocardiac signaling accelerates onset of cardiac arrest. *Proceedings of the National Academy of Sciences* 112, 16 (2015), E2073–E2082.

[8] Yiding Lu, Yufan Guo, and Anna Korhonen. 2017. Link prediction in drug-target interactions network using similarity indices. *BMC bioinformatics* 18, 1 (2017), 39.

[9] Mark EJ Newman. 2001. Clustering and preferential attachment in growing networks. *Physical review E* 64, 2 (2001), 025102.

[10] Peter N Robinson, Sebastian Köhler, Sebastian Bauer, Dominik Seelow, Denise Horn, and Stefan Mundlos. 2008. The Human Phenotype Ontology: a tool for annotating and analyzing human hereditary disease. *The American Journal of Human Genetics* 83, 5 (2008), 610–615.

[11] Eric Wen Su and Todd M Sanger. 2017. Systematic drug repositioning through mining adverse event data in ClinicalTrials. gov. *PeerJ* 5 (2017), e3154.

[12] Tao Zhou, Linyuan Lü, and Yi-Cheng Zhang. 2009. Predicting missing links via local information. *The European Physical Journal B* 71, 4 (2009), 623–630.
[13] XueZhong Zhou, Jörg Menche, Albert-László Barabási, and Amitabh Sharma. 2014. Human symptoms–disease network. *Nature communications* 5 (2014), 4212.

A DISEASE PAIRS

The table below shows 165 disease pairs that we propose as an opportunity of drug repositioning: the drugs treated for one disease could be tested to treat the other disease in a pair.

Disease1	MESH1	Disease2	MESH2
Mycosis Fungoides	D009182	Desmoplastic Small Round Cell Tumor	D058405
Pulmonary Embolism	D011655	Asphyxia	D001237
Kidney Neoplasms	D007680	Delirium, Dementia, Amnestic, Cognitive Disorders	D019965
Bronchiolitis	D001988	Sandhoff Disease	D012497
Meningitis, Meningococcal	D008585	Psychotic Disorders	D011618
Hemifacial Spasm	D019569	Eye Diseases	D005128
Neuroectodermal Tumors, Primitive, Peripheral	D018241	Xerostomia	D014987
Hepatitis, Autoimmune	D019693	Choroid Plexus Neoplasms	D016545
Borrelia Infections	D001899	Tuberculosis	D014376
Scoliosis	D012600	dopamine beta hydroxylase deficiency	C535600
Hematologic Neoplasms	D019337	Diabetes, Gestational	D016640
Uveitis, Intermediate	D015867	Keratoconjunctivitis	D007637
Uveitis, Intermediate	D015867	Triple Negative Breast Neoplasms	D064726
Migraine without Aura	D020326	Influenza, Human	D007251
Postoperative Nausea and Vomiting	D020250	Hemifacial Spasm	D019569
Neoplasm, Residual	D018365	Triple Negative Breast Neoplasms	D064726
Endotoxemia	D019446	Depressive Disorder	D003866
Heart Failure, Systolic	D054143	Retinal Vein Occlusion	D012170
Arthritis, Rheumatoid	D001172	Affective Disorders, Psychotic	D000341
Arthritis, Reactive	D016918	Skin Diseases, Infectious	D012874
Rosacea	D012393	Preterm Premature Rupture of the Membranes	C563032
Rosacea	D012393	Renal Hypodysplasia, Nonsyndromic, 1	C563661
Teratoma	D013724	Pulmonary Veno-Occlusive Disease	D011668
Heart Arrest	D006323	Otitis Media	D010033
Heart Arrest	D006323	Tobacco Use Disorder	D014029
Infant, Newborn, Diseases	D007232	Marijuana Abuse	D002189
Headache	D006261	Sexual Dysfunctions, Psychological	D020018
Spondylitis, Ankylosing	D013167	Thyroid Neoplasms	D013964
Acne Vulgaris	D000152	Atrial Flutter	D001282
Lacerations	D022125	Chest Pain	D002637
Rhinitis, Vasomotor	D012223	Parkinson Disease	D010300
Sinusitis	D012852	Febrile Neutropenia	D064147
Intestinal Obstruction	D007415	Amphetamine-Related Disorders	D019969
Cataract	D002386	Urologic Neoplasms	D014571
Adenomatous Polyposis Coli	D011125	Acquired Immunodeficiency Syndrome	D000163
Pancreatic Cyst	D010181	Retinal Neoplasms	D019572
Eye Diseases	D005128	Ischemic Attack, Transient	D002546
Melanoma	D008545	Uterine Cervical Dysplasia	D002578
Endodermal Sinus Tumor	D018240	Ganglioglioma	D018303
Parkinson Disease	D010300	Humeral Fractures	D006810
Carcinoma, Ductal, Breast	D018270	Liver Cirrhosis	D008103
Fatigue	D005221	Learning Disorders	D007859
Metabolism, Inborn Errors	D008661	Carcinoma, Ductal, Breast	D018270
Catheter-Related Infections	D055499	Ulcer	D014456
Clear-cell metastatic renal cell carcinoma	C538445	Chordoma	D002817
Liver Cirrhosis	D008103	Gallbladder Diseases	D005705
Anemia, Iron-Deficiency	D018798	Urinary Incontinence	D014549
Cicatrix	D002921	Lymphoma, Non-Hodgkin	D008228
Pruritus	D011537	Colitis	D003092
Inflammatory Bowel Diseases	D015212	Carcinoma, Neuroendocrine	D018278
Immune Deficiency Disease	C565469	Hepatitis, Alcoholic	D006519
von Hippel-Lindau Disease	D006623	Kidney Diseases	D007674
Myelodysplastic-Myeloproliferative Diseases	D054437	Bacterial Infections and Mycoses	D001423
Breast Neoplasms	D001943	Jaw, Edentulous	D007575
Radiculopathy	D011843	Intracranial Aneurysm	D002532
Acute Coronary Syndrome	D054058	Temporomandibular Joint Disorders	D013705
Pelvic Pain	D017699	Meningitis, Bacterial	D016920
Lymphoma, T-Cell	D016399	Uterine Cervical Diseases	D002577
Fever	D005334	Neurocysticercosis	D020019
Dilatation, Pathologic	D004108	Ocular Hypertension	D009798
Mucopolysaccharidosis VI	D009087	Sarcoma, Kaposi	D012514
Polycythemia Vera	D011087	Arthritis, Gouty	D015210
Polycythemia Vera	D011087	Hypertension, Portal	D006975
Mental Disorders	D001523	Liver Failure	D017093
Mental Disorders	D001523	Rib Fractures	D012253
Nephrosis, Lipoid	D009402	Leukocyte-Adhesion Deficiency Syndrome	D018370
Nephrosis, Lipoid	D009402	Osteogenesis Imperfecta	D010013
Memory	D008568	Peripheral Arterial Disease	D058729
Spondylolisthesis	D013168	Sleep	D012890
Tendinopathy	D052256	Rhinitis, Allergic, Seasonal	D006255
Neoplasm Recurrence, Local	D009364	Thyroid cancer, papillary	C536915
Fibrosis	D005355	Urinary Bladder Diseases	D001745
Dystonia	D004421	Rhinitis, Atrophic	D012222
Dystonia	D004421	Abscess	D000038
Uterine Cervical Neoplasms	D002583	Peripheral Nerve Injuries	D059348
Dermatitis	D003872	Rosacea	D012393
Exanthema	D005076	Pelvic Inflammatory Disease	D000292
Virus Diseases	D014777	Teratoma	D013724
Abscess	D000038	Hepatitis C, Chronic	D019698

Disease1	MESH1	Disease2	MESH2
Diabetic Retinopathy	D003930	Ureteral Neoplasms	D014516
Lung Diseases	D008171	Vulvar Vestibulitis	D054515
Uterine Diseases	D014591	Precursor T-Cell Lymphoblastic Leukemia-Lymphoma	D054218
Pain, Intractable	D010148	Genital Diseases, Female	D005831
Pain, Intractable	D010148	Ankle Fractures	D064386
Biliary Tract Neoplasms	D001661	Carcinoma, Embryonal	D018236
Varicose Veins	D014648	Acute Pain	D059787
AIDS-related Kaposi sarcoma	C554498	Parotid Neoplasms	D010307
Esophageal Squamous Cell Carcinoma	C562729	Pulmonary Fibrosis	D011658
Ureteral Calculi	D014514	Meningitis, Bacterial	D016920
Mild Cognitive Impairment	D060825	Amphetamine-Related Disorders	D019969
Wilms Tumor	D009396	Eye Diseases	D005128
Lichen Planus, Oral	D017676	Staphylococcal Infections	D013203
Leukemia, Myeloid, Acute	D015470	Smoking Cessation	D016540
Subarachnoid Hemorrhage	D013345	Migraine with Aura	D020325
Liposarcoma	D008080	Retroperitoneal Fibrosis	D012185
Carcinoma, Non-Small-Cell Lung	D002289	Rett Syndrome	D015518
Central Serous Chorioretinopathy	D056833	Febrile Neutropenia	D064147
End Stage Liver Disease	D058625	Breast Diseases	D001941
Systemic Inflammatory Response Syndrome	D018746	Headache Disorders	D020773
Fibromyalgia	D005356	Esophageal Motility Disorders	D015154
Combined Pituitary Hormone Deficiency	C580003	Fetal Growth Retardation	D005317
Seizures	D012640	Cognition Disorders	D003072
Blood Platelet Disorders	D001791	Headache	D006261
Craniopharyngioma	D003397	Chronic Disease	D002908
Urinary Bladder Diseases	D001745	Retinal Neoplasms	D019572
Opioid-Related Disorders	D009293	Unconsciousness	D014474
Periodontitis	D010518	Glomerulonephritis, IGA	D005922
Pityriasis	D010915	Musculoskeletal Diseases	D009140
Acute Pain	D059787	Periarthritis	D010489
Stroke	D020521	Jaw, Edentulous	D007575
Glycogen Storage Disease Type II	D006009	Amyotrophic Lateral Sclerosis	D000690
Multiple Organ Failure	D009102	Scleroderma, Systemic	D012595
Hemorrhage	D006470	Multiple Endocrine Neoplasia	D009377
Hemorrhage	D006470	Depressive Disorder	D003866
Lymphoma, Large B-Cell, Diffuse	D016403	Gonorrhea	D006069
DiGeorge Syndrome	D004062	Body Weight Changes	D001836
Communicable Diseases	D003141	Acne Vulgaris	D000152
Communicable Diseases	D003141	Parasitic Diseases	D010272
Fissure in Ano	D005401	Dermatitis, Seborrheic	D012628
Shock, Septic	D012772	Otitis	D010031
Carcinoma, Large Cell	D018287	Hemangioendothelioma, Epithelioid	D018323
Psychoses, Substance-Induced	D011605	Laryngismus	D007826
Prostatic Hyperplasia	D011470	Biliary Tract Diseases	D001660
Biliary Atresia	D001656	Parkinson Disease	D010300
Hyperlipidemias	D006949	Fractures, Bone	D050723
Myocardial Infarction	D009203	Tremor	D014202
Hypertension, Portal	D006975	Disease	D004194
Musculoskeletal Diseases	D009140	Hemangiopericytoma, Malignant	C562740
Revesz Debuse syndrome	C538371	Sarcoidosis	D012507
Retinopathy of Prematurity	D012178	Lung Diseases, Obstructive	D008173
Prehypertension	D058246	Tennis Elbow	D013716
Irritable Bowel Syndrome	D043183	Multiple Sclerosis, Chronic Progressive	D020528
Adenocarcinoma Of Esophagus	C562730	Leukocyte-Adhesion Deficiency Syndrome	D018370
Adenocarcinoma Of Esophagus	C562730	Hutchinson's Melanotic Freckle	D018327
Emphysema	D004646	Intraabdominal Infections	D059413
Urinary Calculi	D014545	Mood Disorders	D019964
Pulpitis	D011671	Peripheral Vascular Diseases	D016491
Gallbladder Diseases	D005705	Diabetic Neuropathies	D003929
Sleep Disorders	D012893	Binge Drinking	D063425
Urinary Incontinence	D014549	Bacteremia	D016470
Vasospasm, Intracranial	D020301	Muscular Dystrophies	D009136
Carcinoma, Basal Cell	D002280	Xerostomia	D014987
Glaucoma	D005901	alpha 1-Antitrypsin Deficiency	D019896
Hyperalgesia	D006930	Borderline Personality Disorder	D001883
Hyperalgesia	D006930	Vascular Diseases	D014652
Disorders of Excessive Somnolence	D006970	Diabetic Neuropathies	D003929
Diabetes Mellitus	D003920	Corneal Ulcer	D003320
Nephrotic Syndrome	D009404	Carcinoma, Transitional Cell	D002295
Preleukemia	D011289	Uterine Cervical Neoplasms	D002583
Choroidal Neovascularization	D020256	Vulvar Neoplasms	D014846
Myalgia	D063806	Motor Activity	D009043
Headache Disorders	D020773	Meningomyelocele	D008591
Tachycardia	D013610	Shoulder Pain	D020069
Tachycardia	D013610	Sexual Dysfunctions, Psychological	D020018
Eye Diseases, Hereditary	D015785	Coronary Disease	D003327
Kidney Diseases	D007674	Cholangitis, Sclerosing	D015209
Dizziness	D004244	Delirium	D003693
Esophageal and Gastric Varices	D004932	Hypercholesterolemia	D006937
Child Development Disorders, Pervasive	D002659	Carpal Tunnel Syndrome	D002349
Lymphoma, Non-Hodgkin	D008228	Anorexia	D000855
Respiratory Sounds	D012135	Hermanski-Pudlak Syndrome	D022861
Obesity	D009765	Papillomavirus Infections	D030361
Leukemia, Mast-Cell	D007946	Bile Duct Neoplasms	D001650
Hepatic Encephalopathy	D006501	Glomerulonephritis, Membranous	D015433
Neurofibroma, Plexiform	D018318	Neoplasms, Unknown Primary	D009382