Evidence Based Disease Network Construction towards Drug Repositioning

Liwei Wang, MD[1]
[1] Department of Medical Informatics
Jilin University
Changchun, China

Jiabei Wang, PhD[2]
[2]School of Pharmacy
University of Maryland
Maryland, USA

Qian Zhu, PhD*[2]
[2] Department of Information Systems
University of Maryland Baltimore County
Maryland, USA

***Abstract*—Drug repositioning is one of emerging approaches dedicated to find alternative usages of existing drugs efficiently and economically, especially with the advance in computational technology. The current progress made for computational drug repositioning is primarily focusing on informatics approach development/improvement or exploration on different type of data in order to identify possible drug candidates. Comparing to the existing studies, we proposed a novel method for constructing the disease based network by applying data extracted from the Semantic MEDLINE. Phenotypical associations (disease-disease associations) can be identified from this network, which can drive drug repositioning study by targeting on specific domain. In this paper, we successfully demonstrated the capability of the disease based network in hidden phenotypical association discovery to support drug repositioning in case studies.**

Keywords—drug repositioning; phenotypical association; network analysis; semantic MEDLINE

I. INTRODUCTION

Drug repositioning (DR) is an emerging approach of finding alternative indications for existing drugs. It is one of the most cost- and time-effective methods, with the potential to push low-risk drug development. With the advance in computational technology, the pace of DR has been accelerated dramatically in an automated manner. Many computational approaches have been applied towards DR and demonstrated their contributions from different angles. Wang Y et al. [1] used a support vector machine (SVM) to computationally predict novel drug-disease interactions in a bipartite graph representing known interactions between drugs and diseases. Molineris I et al. [2] developed a novel approach based on known drug targets showing conserved anti-correlated expression profiles for human disease genes. Liu HY et al. [3] identified 54 novel small molecular drugs through an integrated analysis of Polycystic Ovary Syndrome-involved sub-pathways and drug-affected sub-pathways. Zhao H et al. [4] developed a computational model based on cancer signaling bridges, and derived specific downstream signaling pathways that reveal previously unknown target-disease connections and new mechanisms for specific cancer subtypes. This computational model enabled DR efforts and their preclinical experiments successfully proved two approved drugs, sunitinib and dasatinib, prohibit brain metastases from breast cancer.

Network analysis, as one of computational approaches, has gained much popularity for DR with promising findings. The current scenario in network analysis for DR is either focusing on biomedical association discovery, or most likely on revealing unexpected/hidden associations. Cheng F et al. [5] inferred new targets for known drugs through a drug-target network based on 12,483 FDA-approved and experimental drug-target binary links. Yildirim MA et al. [6] built a network to connect most drugs into a highly interlinked giant component, and the analyses showed an overabundance of 'follow-on' drugs, that is, drugs that target already targeted proteins. Sun K et al. [7] constructed an Integrated Disease Network (IDN) by integrating different types of biological data including genome-wide association studies data, disease-chemical associations, biological pathways and Gene Ontology annotations. This network provided a powerful approach for exploring connections between diseases, but no insights into DR. Recently Zhang Y et al. [8] demonstrated some potential drug repositioning capability for Alzheimer's Disease based on the association patterns between diseases and drugs from the Semantic MEDLINE (SM), and they explored network motif detection in the drug-disease-gene network based on direct associations among drug, disease and gene presented in literatures.

Lots of biomedical knowledge is buried in a large amount of literatures from biomedical databases. SM is one of repositories for semantic predictions extracted from the MENDLINE and maintained by National Library of Medicine (NLM). A number of studies have explored SM for biomedical data mining and illustrated the capability of SM for supporting such research consequently. For example, Zhang R et.al [9] demonstrated SM, presented as structured knowledge in the form of relationships from the biomedical literature, can support the discovery of potential drug-drug interactions occurring in patient clinical data. Cairelli M et al. [10] identified potential biomarkers for mild traumatic brain injury by mining semantic relations from SM. Cairelli MJ et al. [11] elucidated the paradox that obesity is beneficial in critical care despite contributing to disease generally by using semantic predications from SM and the literature-based discovery paradigm. But none of them attempted to apply or extend their findings identified from SM for DR, except for [8] showing some preliminary work by exploring SM for DR.

978-1-4799-7294-4/14/$31.00 ©2014 IEEE

TABLE I. SAMPLE DATA EXTRACTED FROM THE PA TABLE

No.	PMID	Predicate	s_cui	s_name	s_type	o_cui	o_name	o_type
1	6004517	CAUSES	C0593465	Potassium Chloride Injectable Solution	clnd	C0542052	Coronary artery insufficiency	dsyn
2	20436698	COMPLICATES	C0742342	CHEST SYNDROME	dsyn	C0242770	Bronchiolitis Obliterans Organizing Pneumonia	dsyn
3	20351359	AFFECTS	C0032285	Pneumonia	dsyn	C0029235	Organism	orgm

In this presented study, the effort we made was to identify phenotypical associations (disease-disease associations) from SM, which is the key step of computational drug repositioning candidate screening pipeline we proposed for DR[12]. We included inferred disease specific associations to build a disease based network (DBN). In this paper, we described a novel method for constructing DBN with data from SM along with materials being applied in this study in Materials and Methods section. Corresponding results for DBN construction are included in Results section, followed by Case Studies section to demonstrate the feasibility and validity of our method. We then discussed our findings along with the related issues.

II. MATERIALS AND METHODS

In this paper, we introduced an approach to build a disease network on the basis of phenotypical associations extracted from SM towards drug repositioning. Fig. 1 presents an overview of the workflow for the DBN construction that contains three steps: phenotypical association extraction, phenotypical association prioritization and DBN constrution. Those steps are further described in detail below.

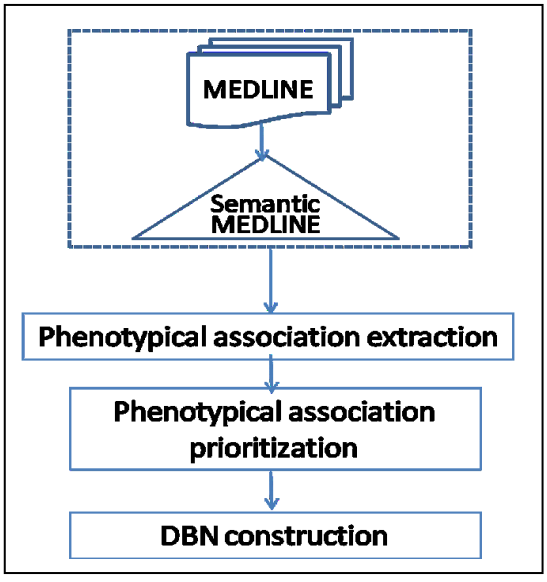

Fig. 1. Workflow for disease network construction

A. Phenotypical association extraction from SM

SM has more than 56 million semantic associations that were extracted from MEDLINE via a natural language processing tool called SemRep on the basis of Unified Medical Language System (UMLS). It consists of eight tables, of which the PREDICATION_AGGREGATE (PA) table gathers salient information from all other tables for efficient access to semantic predications. PA table contains comprehensive information on semantic associations, including subject concepts, object concepts, predicate, and semantic types of subject concepts and object concepts, as well as their associated PubMed IDs (PMIDs). Table I shows sample records extracted from the PA table, where the predicate shows the association between subject (s) and object (o), e.g., the association between the subject “Potassium Chloride Injectable Solution” and the object “Coronary artery insufficiency” is “CAUSES”. “s_type” and “o_type” are semantic types related to the subject concept and object concept, e.g., the s_type of “Potassium Chloride Injectable Solution” is “clnd”(Clinical Drugs), and the o_type of “Coronary artery insufficiency” is “dsyn” (Disease and Syndrome). In this study, we extracted data from the PA table with the MAR 31, 2014 version[13]. We used 2014 May version of UMLS semantic type file MRSTR.RRF [14] to identify the concepts with according semantic types from the PA table.

1) Phenotypical association extraction

To obtain existing phenotypical associations from the PA table, we extracted associations between subject and object concepts both labeled with the semantic type as "Disease or Syndrome" or its subtypes of “Neoplastic Process” and “Mental or Behavioral Dysfunction”. For example, a disease pair of “CHEST SYNDROME” - “Bronchiolitis Obliterans Organizing Pneumonia” shown in the table I was extracted as both of them had the semantic type “Disease or Syndrome”. It is worthy to note that we excluded the phenotypical associations between the same disease terms, such as, “Abscess” – “Abscess” (C0000833)”. In parallel, we counted the number of PMIDs for suchexisting phenotypical associations for further analysis.

2) Inferred phenotypical association identification

In order to infer non-exisiting phenotypical associations, two-step inference algorithm has been designed. 1) We extracted associations between subject and object concepts in the PA table with either of them having the semantic type as "Disease or Syndrome" or its child types “Neoplastic Process” and “Mental or Behavioral Dysfunction”. 2) We applied a

978-1-4799-7294-4/14/$31.00 ©2014 IEEE

simple but efficient inference rule shown below to identify indirect phenotypical associations.

Inference rule: IF Disease A is associated with Concept B, and Concept B is associated with Disease C, THEN Disease A is associated with Disease C. Where the Concept B could have any semantic types, such as "Disease or Syndrome" or "Clinical Drugs".

Followed by the inference rule, new phenotypical associations have been inferred. One example is shown in the Fig. 2, we generated three novel phenotypical associations, including "Hyperglycinemia" – "Myopathy", "Myopathy" – "Neuromuscular inhibition" and "Hyperglycinemia" – "Neuromuscular inhibition" based on the above inference rule. For those inferred associations same as the existing associations that exist in the PA table, we moved them to the existing phenotypical association list, not in the inferred one.

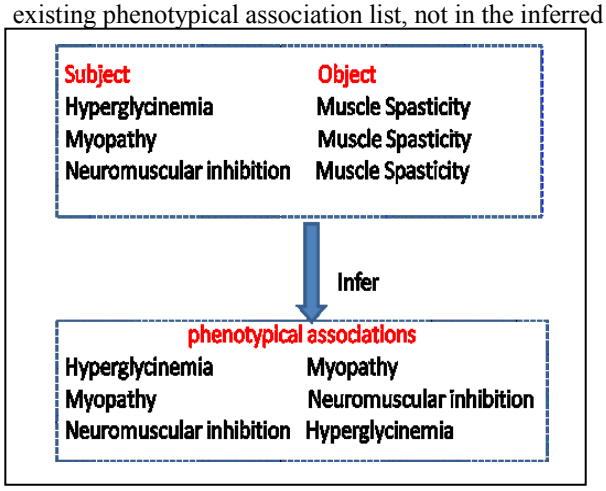

Fig. 2. An example of phenotypical association inference

B. *Phenotypical association prioritization*

With a large number of phenotypical associations, especially for the inferred ones extracted from the SM, we performed prioritization steps to screen the most significant phenotypical associations via three steps, 1) based on the semantic types of Concept B defined in the inference rule, 2) based on the predicates associated with the inferred associations, and 3) based on the weights assigned to the phenotypical associations. Noted that there are two semantic types for concept B and two predicates for the inferred associations (disease A- disease B) as they are derived from two original associations (disease A- Concept B, and disease B – Concept B) with two individual semantic types and predicates.

1) *Prioritization based on semantic types*

We retrieved semantic types for Concept B, and ranked them according to their occurring times through the entire set of associations extracted from the SM. The highest occurrences for one semantic type "human" is up to 290 million times, and the lowest one is "Health Care Related Organization" with 6 occurrences. To exclude less popular semantic types, we set a cutoff occurrence value as 10,000 to remove those phenotypical associations with occurrences less than 10,000. In the meantime, we manually reviewed the rest of semantic types with higher occurrence rate and excluded semantic types that are too general. For example, the semantic type "Clinical Attribute" is applied for the concept B "Ocular Accommodation, C0000936", that is associated with a wide range of different types of disease showing no significant relatedness among those diseases, such as "malaria, C0024530" and "Obesity, C0028754". Another example is "education". The concept "education, C0013621" also connects many diseases without reasonable associations, such as "Cardiovascular Diseases, C0007222" and "Leprosy, C0023343".

2) *Prioritization based on predicates*

Predicates indicate specific associations between two concepts (subject and object). In this study, we included predicates, such as "CAUSES" and "COEXISTS_WITH" indicating descriptive associations among original associations. For example, the predicate "CAUSES" assigned for one association specifies "Multiple Sclerosis" causes "Muscle Spasticity", and the predicate "COEXISTS_WITH" specifies relationship between "Hyperglycinemia" and "Muscle Spasticity".

Meanwhile, we manually removed predicates with less correlation with disease terms, such as, "PROCESS_OF", focusing on biological classification, and "LOCATION_OF", dealing with the body location of disease.

3) *Weights for Phenotypical Associations*

Once we filtered phenotypical associations based on semantic types and predicates, a two-step approach has been applied to calculate and assign weights to phenotypical associations. 1) We assigned a weight to each inferred association. More specifically, in this study, we highly relied on the number of PMIDs related with each association to express the popularity of relevant research interest. We counted the number of PMIDs for the association between disease A and Concept B as C1, and the number of PMIDs for the association between Concept B and disease B as C2, then we calculated the average value of C1 and C2 and assigned this average number as a weight to the inferred association (disease A and disease B). We setup a cutoff value as 100 for the weight to filter out less popular associations. 2) We gave higher weight to the existing phenotypical associations based on the number of PMIDs associated by adding the highest score of the above weight (12922). Ultimately, we had a ranked list of phenotypical associations with corresponding weights for the further network analysis.

C. *Disease based network (DBN) construction*

All resulted phenotypical associations have been used for constructing the DBN, where the nodes correspond to the disease terms extracted from SM, the edges correspond to the associations, and the weight of the edges.

978-1-4799-7294-4/14/$31.00 ©2014 IEEE

III. RESULTS

A. Phenotypical associations extraction

There were 160,090 existing phenotypical associations involving 11,896 concepts from the PA table. Meanwhile, 16,003 concepts were extracted through the inference rule, consisting of 80,713,255 phenotypical associations.

Of total 136 semantic types from the SM, 115 semantic types for concept B were identified. The top 10 semantic types of concept B, including "Disease or Syndrome", are listed in Table II.

Of total 50 different predicates occurring in the SM, 44 predicates for the phenotypical associations have been identified in this study. We ranked those 44 different predicates based on their occurrence, and the top 10 most occurring predicates are shown in Table III.

TABLE II. TOP 10 SEMANTIC TYPES OF CONCEPT B FOR EXTRACTED PHENOTYPICAL ASSOCIATIONS

Order	Semantic type abbreviation	Full name of semantic types	Occurrence times
1	humn	Human	290, 735, 132
2	dsyn	Disease or Syndrome	120, 315, 176
3	topp	Therapeutic or Preventive Procedure	90, 437, 803
4	patf	Pathologic Function	45, 040, 997
5	podg	Patient or Disabled Group	38, 343, 393
6	mamm	Mammal	37, 376, 417
7	hlca	Health Care Activity	37, 247, 413
8	sosy	Sign or Symptom	26, 767, 805
9	phsu	Pharmacologic Substance	25, 563, 120
10	fndg	Finding	25,539,925

TABLE III. TOP 10 PREDICATES BETWEEN CONCEPT B AND PHENOTYPICAL ASSOCIATIONS

Order	Predicates	Occurrence times
1	PROCESS_OF	299, 938, 467
2	TREATS	137, 511, 428
3	AFFECTS	126, 916, 221
4	COEXISTS_WITH	112, 977, 777
5	ASSOCIATED_WITH	52, 645, 970
6	CAUSES	46, 706, 770
7	LOCATION_OF	41, 453, 463
8	DIAGNOSES	32, 501, 936
9	ISA	31, 118, 683
10	OCCURS_IN	29, 846, 571

B. Prioritization of Phenotypical Associations

By excluding 22 less-relevant semantic types, 15,977 unique disease terms and corresponding 80,661,805 unique phenotypical associations were included for further prioritization.

By excluding 2 less-relevant predicates, 15804 unique diseases and 51,129,608 unique disease pairs were retained for weight assignment.

After we conferred the weights to the phenotypical associations and applied the cutoff value of 100 for filter, 2970 unique disease terms and 300,843 phenotypical associations were obtained for further network analysis. Among these inferred associations, 42,982 associations proved to have direct relationship in PA table, and the remaining 257,168 pairs revealed the potential new associations between diseases.

C. DBN construction

The DBN contains 2970 unique nodes corresponding to the resulted diseases and 300,843 edges corresponding to the resulted associations in total, where the weight for edges can be retrieved. We will release the DBN when it is ready.

IV. CASE STUDIES

The DBN provides plentiful information on disease related associations. we will be able to identify potential phentypical associations from this network to ultimately support drug repositioning. More specifically, such application can be used in two folds, 1) the drugs for the disease directly associated with the interested disease in the DBN maybe the candidates to be repositioned for the interested disease, or 2) we will be able to explore the data indirectly relevant to the associated disease, then from there, we can identify possible drug repositioning candidates for the interested disease. The below two case studies illustrate the feasibility of our method and the capability of the network for drug repositioning.

A. Case study 1

Infective endocarditis (IE) is a rare, potentially fatal disease with susceptible endocardium or a prosthetic heart valve being infected by microorganisms such as streptococci, staphylococci and candida[15]. It primarily affects people over the age of 55 though it was once a disease of early adulthood, partly because the incidence of rheumatic heart disease in the under 55 age group decreased and rates of cardiac damage with age increased[16].

In this case study, we were interested in identifying any other diseases that are potentially related to IE from the DBN. Especially the novel inferred associations may not only offer some novel insights to assist diagnosis upon the new disease pattern for clinicians, but also support chemists/pharmacists to explore new drug repositioning candidates for IE.

We first searched IE with the UMLS identifier "C1541923" against the DBN, and obtained 610 unique inferred phenotypical associations to IE. Table IV shows the top 10 inferred associations .

Of the possible phenotypical associations related to IE extracted from the DBN, one association "Infective Endocarditis, obesity" has been examined further by manual literature searching. Only one studys [17] has been found,

978-1-4799-7294-4/14/$31.00 ©2014 IEEE

where they reported possible association between IE and obesity, i.e., people with a risk of obesity may have a major and significant weight gain after a six-week intravenous IE treatment by vancomycin plus gentamycin, especially males older than 65 who have not undergone surgery. This case shows the capability of DBN to uncover potential associations, especially the association that has not yet been well studied. There will be probabilities that the exploration of data related to obesity may identify drug repositioning candidates for IE.

We also performed further network analysis to identify other indirect associations that can be derived from the DBN via network inference, i.e., disease A linking to disease B and disease B linking to disease C→ disease A linking to disease C. In this study, we only identified the inferred relations via one intermediate layer, with no farther inference results through more than one intermediate layers allowed. For this case, the associations, "Infective Endocarditis, Diabetes (C1541923, C0011847)" and "Diabetes, Uremia (disorder) (C0011847, C0041948)" with the weight of 1117 listed in Table IV had been extracted. Thus, the novel association between IE and Uremia can be generated through network inference. This finding has been proved by the literature evidence, such as Miyake M et al [18] reported a patient with IE accompanied by renal failure, that suggests uremia can develop as an initial manifestation of IE. While currently not many attentions have been put on the relationships between uremia and IE, this case study using network analysis approach revealed the potential association in the DBN for further study.

TABLE IV. TOP 10 INFERRED PHENOTYPICAL ASSOCIATIONS FOR IE

Rank	Weight	Top 10 Disease Pairs
1	590.5	Infective Endocarditis, Clusters| Symptom (C0039082)
2	560.5	Infective Endocarditis|Unspecified septicemia (C0036690)
3	552.5	Infective Endocarditis|Endocarditis (C0014118)
4	523.5	Infective Endocarditis|Pneumonitis (C0032285)
5	456.5	Infective Endocarditis|Bacteraemia (C0004610)
6	427	Infective Endocarditis|Diabetes (C0011847)
7	422	Infective Endocarditis|Infarction; heart (C0027051)
8	419	Infective Endocarditis|High Blood Pressures (C0020538)
9	408	Infective Endocarditis|Heart failure NOS (disorder) (C0018801)
10	399	Infective Endocarditis|Obesity (C0028754)

B. case study 2

Alzheimer's disease (AD), a neurodegenerative disease characterized by a progressive damage to the brain cells leading to cognitive dysfunction, commonly occurs in the elderly[19]. Using the similar method for case study 1, we first searched AD with the UMLS identifier "C0002395" against the DBN, and obtained 1505 unique inferred phenotypical associations to AD. Table V shows the top 10 associations in part 2.

Of the associations presented in Table V, we manually examined from literatures for the association between AD and diabetes (C0011847). Existing studies provided supporting evidences, [20, 21] have suggested a possible shared pathophysiology between type 2 diabetes mellitus (T2DM) and AD. A hypothesis that AD might be associated with type 3 diabetes has been proposed[22]. Therefore, treatment currently used for Type 2 Diabetes might be effective for AD patients, e.g., GSK-3 inhibition could be a common target treatment of both AD andT2DM[23]. In fact several clinical trials are underway to test the effectiveness of 'antidiabetic' drugs, such as intranasal insulin[24] in AD patients.

Besides the above exploration, we retrieved the associations to "Diabetes (C0011847)" in the DBN and located the association "Diabetes, neuroborreliosis (C0011847, C0948264)" with the weight of 109. Through the network inference, neuroborreliosis can be linked to AD and that can be verified by the latest study [25], where patients with AD and a positive "intrathecal anti-Borrelia antibody index" (AI), specific for neuroborreliosis were identified. This case study suggests AD and neuroborreliosis are co-morbidity, demonstrating the validity and feasibility of our method.

TABLE V. TOP 10 INFERRED PHENOTYPICAL ASSOCIATIONS FOR AD

Rank	Weight	Top 10 Disease Pairs
1	2543.5	Alzheimer's disease|Diabetes(C0011847)
2	2386.5	Alzheimer's disease|Obesity (C0028754)
3	2335	Alzheimer's disease|High Blood Pressures (C0020538)
4	2036	Alzheimer's disease|Cerebrovascular accident unspecified (C0038454)
5	1926.5	Alzheimer's disease|Asthma (C0004096)
6	1865	Alzheimer's disease|Consumption-archaic term for TB (C0220811)
7	1770.5	Alzheimer's disease|seizures syndrome (C0014544)
8	1714.5	Alzheimer's disease|Diabetes mellitus, NOS (C0011849)
9	1482.5	Alzheimer's disease|HTLV WIII LAV INFECTIONS (C0019693)
10	1451.5	Alzheimer's disease|Chronic Illnesses (C0008679)

V. DISCUSSION

In this study, we introduced a disease network built by integrating phenotypical associations extracted/inferred from literature in SM. Disease orientated network was designed not only to provide comprehensive information specific to diseases that will strongly support epidemiology related study, e.g., finding treatment paradigm upon the new disease pattern identified from the DBN, but also to lead a novel direction towards drug repositioning. In the case studies presented in this paper, we have successfully demonstrated the capability of the DBN for existing/novel phenotypical association discovery to support drug repositioning.

Two tags can be highlighted for the DBN, 1) evidence supported, all phenotypical associations were extracted from the SM. All associations presented in the DBN can be tracked back to the relevant published manuscripts accordingly. Each association has associated with one or more PMID, as there are at least two PMIDs assigned to the inferred association (at least one PMID associated with one original association). 2) concepts included in the DBN are labeled with standardized identifiers, UMLS CUIs. Data standardization will facilitate future data integration with more additional resources and support cross evaluation by other relevant resources.

978-1-4799-7294-4/14/$31.00 ©2014 IEEE

To expand the coverage of the DBN and provide more possible associations, we implemented two-fold inference approach, 1) including non-existing phenotypical associations derived from original associations in the SM based on a two-step inference rule, 2) to find indirect association from the DBN via network inference, that is shown in the case studies.

A huge number of phenotypical associations have been included in the DBN, applying prioritizing algorithms is essential to filter out significant associations with interested diseases. To exclude false positive results in this preliminary work, we employed simple but efficient methods, such as occurrence of the particular association in the SM and the number of studies (PMIDs) related to the associations, to prioritize inferred associations. The results shown in the case studies evidently illustrate the efficacy of our approach for supporting drug repositioning. However, human interfere is inevitable during this ranking process, as we have to manually select the most significant associations from the top list generated automatically based on domain knowledge/interests or literature searching. Thus, in the future, we will explore more systematical/intelligent algorithms, such as page rank, topic modeling to assist association selection.

As we know, UMLS semantic types contain a hierarchical structure to capture relationships among them, such as "Disease or Syndrome" has two child semantic types of "Neoplastic Process" and "Mental or Behavioral Dysfunction", and one parent type "Pathologic Function". In this preliminary study, we only selected one UMLS semantic type "Disease or Syndrome" to identify disease related concepts from the SM, because we intended to exclude any possible false positive information that may interfere the final results in the early stage of investigation. However, on another hand, other semantic types related to disease may be missed, such as "Congenital Abnormality". To integrate more phenptypical associations into the DBN, we will include more disease related semantic types to extract more phenotypical associations from the SM.

While we applied semantic types to filter phenotypical associations from the SM, we found the semantic types defined in the SM are inconsistent with those defined by UMLS (UMLS semantic type file MRSTR.RRF with version of 2014 May). For example, Staghorn Calculus (C0333014) was labeled in SM as the type "Body Substance, bdsu" instead of "Disease or Syndrome". One reason may be that the concepts from the SM were defined on the basis of the UMLS 2006AA release[13], which is an early version. We consider that it would be desirable for SM to update its terminology reference in time, thus providing more accurate information to researchers.

VI. CONCLUSION

In this study we have built a DBN through inference in SM. The network offers information on phenotypical associations, supporting medical studies, especially providing a orientation and paradigm to gain insights for drug repurposing. The capability and the potential of the DBN have been successfully demonstrated by case studies.

REFERENCES

[1] Wang Y, Chen S, Deng N, Wang Y: Drug Repositioning by Kernel-Based Integration of Molecular Structure, Molecular Activity, and Phenotype Data. PloS one 2013, 8(11):e78518.

[2] Molineris I, Ala U, Provero P, Di Cunto F: Drug repositioning for orphan genetic diseases through Conserved Anticoexpressed Gene Clusters (CAGCs). BMC bioinformatics 2013, 14(1):288.

[3] Liu H-Y, Liu J-Q, Mai Z-X, Zeng Y-T: A Subpathway-Based Method of Drug Reposition for Polycystic Ovary Syndrome. Reproductive Sciences 2014:1933719114542025.

[4] Zhao H, Jin G, Cui K, Ren D, Liu T, Chen P, Wong S, Li F, Fan Y, Rodriguez A: Novel modeling of cancer cell signaling pathways enables systematic drug repositioning for distinct breast cancer metastases. Cancer research 2013, 73(20):6149-6163.

[5] Cheng F, Liu C, Jiang J, Lu W, Li W, Liu G, Zhou W, Huang J, Tang Y: Prediction of drug-target interactions and drug repositioning via network-based inference. PLoS computational biology 2012, 8(5):e1002503.

[6] Yıldırım MA, Goh K-I, Cusick ME, Barabási A-L, Vidal M: Drug—target network. Nature biotechnology 2007, 25(10):1119-1126.

[7] Sun K, Buchan N, Larminie C, Pržulj N: The integrated disease network. Integrative Biology 2014.

[8] Zhang Y, Tao C, Jiang G, Nair AA, Su J, Chute CG, Liu H: Network-based analysis reveals distinct association patterns in a semantic MEDLINE-based drug-disease-gene network. Journal of Biomedical Semantics 2014, 5(1):1-13.

[9] Zhang R, Cairelli MJ, Fiszman M, Rosemblat G, Kilicoglu H, Rindflesch TC, Pakhomov SV, Melton GB: Using semantic predications to uncover drug–drug interactions in clinical data. Journal of biomedical informatics 2014.

[10] Cairelli M, Zhang H, Rindflesch T: Mining Semantic Relations from MEDLINE Citations: Identifying Potential Biomarkers for Mild Traumatic Brain Injury. 2011.

[11] Cairelli MJ, Miller CM, Fiszman M, Workman TE, Rindflesch TC: Semantic MEDLINE for Discovery Browsing: Using Semantic Predications and the Literature-Based Discovery Paradigm to Elucidate a Mechanism for the Obesity Paradox. In AMIA Annual Symposium Proceedings: 2013. American Medical Informatics Association; 2013:164.

[12] Qian Zhu HL, Yuji Zhang: Evidence based computational drug repositioning candidate screening pipeline design: Case Study. to be submitted.

[13] SemMedDB download website [http://skr3.nlm.nih.gov/SemMedDB/dbinfo.html]

[14] UMLS download website [http://www.nlm.nih.gov/research/umls/licensedcontent/umlsknowledgesources.html]

[15] Pallasch TJ, Slots J: Antibiotic prophylaxis and the medically compromised patient. Periodontology 2000 1996, 10(1):107-138.

[16] Tetsas A, Ferguson MM: Concepts for the prophylaxis of infective endocarditis in dentistry. Australian dental journal 2001, 46(3):220-225.

[17] Thuny F, Richet H, Casalta J-P, Angelakis E, Habib G, Raoult D: Vancomycin treatment of infective endocarditis is linked with recently acquired obesity. PloS one 2010, 5(2):e9074.

[18] Miyake M, Hatta K, Kameyama T, Himura Y, Gen H, Kobashi Y, Konishi T: Infective endocarditis developing as uremia. Internal medicine (Tokyo, Japan) 2005, 44(6):598-602.

[19] Luo Y, Vali S, Sun S, Chen X, Liang X, Drozhzhina T, Popugaeva E, Bezprozvanny I: Aβ42-binding peptoids as amyloid aggregation inhibitors and detection ligands. ACS chemical neuroscience 2013, 4(6):952-962.

[20] Ott A, Stolk R, Van Harskamp F, Pols H, Hofman A, Breteler M: Diabetes mellitus and the risk of dementia The Rotterdam Study. Neurology 1999, 53(9):1937-1937.

978-1-4799-7294-4/14/$31.00 ©2014 IEEE

[21] Kroner Z: The relationship between Alzheimer's disease and diabetes: type 3 diabetes? Alternative Medicine Review 2009, 14(4):373.

[22] Suzanne M, Wands JR: Alzheimer's disease is type 3 diabetes—evidence reviewed. Journal of diabetes science and technology 2008, 2(6):1101-1113.

[23] Bhat RV, Budd Haeberlein SL, Avila J: Glycogen synthase kinase 3: a drug target for CNS therapies. Journal of neurochemistry 2004, 89(6):1313-1317.

[24] Reger M, Watson G, Frey Ii W, Baker L, Cholerton B, Keeling M, Belongia D, Fishel M, Plymate S, Schellenberg G: Effects of intranasal insulin on cognition in memory-impaired older adults: modulation by APOE genotype. Neurobiology of aging 2006, 27(3):451-458.

[25] Blanc F, Philippi N, Cretin B, Kleitz C, Berly L, Jung B, Kremer S, Namer IJ, Sellal F, Jaulhac B: Lyme Neuroborreliosis and Dementia. Journal of Alzheimer's Disease 2014.

978-1-4799-7294-4/14/$31.00 ©2014 IEEE