Unsupervised Learning and Multipartite Network Models: A Promising Approach for Understanding Traditional Medicine

Mohieddin Jafari, Yinyin Wang, Ali Amiryousefi and Jing Tang *

Research Program in Systems Oncology, Faculty of Medicine, University of Helsinki, Helsinki, Finland

The ultimate goal of precision medicine is to determine right treatment for right patients based on precise diagnosis. To achieve this goal, correct stratification of patients using molecular features and clinical phenotypes is crucial. During the long history of medical science, our understanding on disease classification has been improved greatly by chemistry and molecular biology. Nowadays, we gain access to large scale patient-derived data by high-throughput technologies, generating a greater need for data science including unsupervised learning and network modeling. Unsupervised learning methods such as clustering could be a better solution to stratify patients when there is a lack of predefined classifiers. In network modularity analysis, clustering methods can be also applied to elucidate the complex structure of biological and disease networks at the systems level. In this review, we went over the main points of clustering analysis and network modeling, particularly in the context of Traditional Chinese medicine (TCM). We showed that this approach can provide novel insights on the rationale of classification for TCM herbs. In a case study, using a modularity analysis of multipartite networks, we illustrated that the TCM classifications are associated with the chemical properties of the herb ingredients. We concluded that multipartite network modeling may become a suitable data integration tool for understanding the mechanisms of actions of traditional medicine.

Keywords: unsupervided learning, network modelling, precision medicine, traditional medicine, systems medicine

INTRODUCTION

Classification and clustering are our fundamental learning process to understand human biology and diseases. To achieve the ultimate goal of precision medicine, i.e., the right intervention for a patient at the right time (Stefano and Kream, 2015), there has been a long history of symptom-based diagnosis that utilizes available information to classify patients, diseases, and drugs (**Figure 1**). In the early days of traditional medicine, physicians tried to characterize diseases using empirical terms, such as temperament and meridian (Rezadoost et al., 2016; Arji et al., 2019; Wang Y. et al., 2019), based on which they prescribed corresponding herbs that are known to target them (Xu, 2011; Li and

FIGURE 1 | A brief history of medicine and its relation with other branches of science. Traditional medicine as the first era of medicine was mainly built on the physical characterization of diseases and patient biographical data. The modern medicine was established by including more chemical and physical characterizations. Defining biomolecule using biochemistry and molecular biology revealed more details of diseases and pathological processes. This eventually led to the development of diagnosis codes and the pharmaceutical industry. Recently, precision medicine has emerged with the advances of data science, which involves more holistic analyses in order to understand human medicine at the systems level.

Weng, 2017). With increasing knowledge on biochemistry, the era of modern medicine has started, further advancing our understanding of human diseases to the molecular level. Molecular profiles, along with clinical phenotypes, are leveraged to formally characterize diseases, disorders, and symptoms (Steindel, 2010).

One of the greatest challenges in precision medicine is to integrate all available patient-derived data for accurate diagnosis and treatment, which would require novel data-driven approaches rather than more conventional hypothesis-driven approaches (Rouillard et al., 2015). Genomic information for patients, albeit fundamental and often necessary, may not be sufficient due to the fact that the human genome is dynamically adjusting its functions by epigenetic regulations (Jafari et al., 2017; Nussinov et al., 2019). Therefore, interactions among other molecular features including transcriptome, proteome, and metabolome should also be considered to obtain a more systematic characterization of the diseases (Eric, 2014). On the other hand, phenotypic data including cell and tissue images have been utilized for illustrating the impact of molecular alternations in human diseases (Langlois et al., 2011; De Fauw et al., 2018). Likewise, to improve our understanding on human diseases, we may also investigate sources of clinical, phenotypic, and pharmacological data that are derived from traditional medicine (Ma et al., 2010; Zhao et al., 2014). A systematic integration of all of these available information may provide a promising approach to turn precision medicine into a reality ultimately.

Here, we started by reviewing the application of clustering analysis in high-throughput biological studies in modern and traditional medicine. Next, we described the application of clustering in network modeling for the stratification of drugs or patients. We focused on the advantages and promises of a particular network modelling approach called multipartite networks which can inherently integrate heterogeneous data types at multiple levels. In a case study, a multipartite network was developed to model traditional medicine herbs. We showed that this modeling approach provides novel insights on the rationale of herb classifications, which may facilitate the drug discovery in TCM, such as discovering herb combinations or prioritization of active ingredients.

USING CLUSTERING TO IMPROVE PATIENT STRATIFICATIONS

Thanks to advanced experimental and computational technologies, we are able to collect, standardize, and integrate a variety of cell-based patient-derived datasets. For example, the LINCS program (Keenan et al., 2018) is one of the recent multi-center studies to facilitate the understanding of cancer biology by providing transcriptional and morphological changes of multiple cancer cell lines in responding to a variety of pharmaceutical agents. Moreover, there are national and international efforts to sequence patients' genomic features. For example, UK Biobank and FinnGen focused on the contributions of genetic predisposition and environmental exposure to the occurrence of common diseases for over half a

million subjects (The Finngen Research Project Takes Finns to a Discovery Trip to Genome Data; Manolio, 2018). On the other hand, with the development of computational methods such as text mining technologies, researchers are able to standardize data resources in traditional medicine, making them easily accessible and reusable (Zhou et al., 2005; Mirzaeian et al., 2019). For example, the SymMap database was constructed to provide a mapping relationship between 499 natural products, 19,595 ingredients, 1,717 clinical symptoms, and 5,235 diseases (Wu et al., 2019). This work showed the potential to integrate traditional and modern medicine at both phenotypic and molecular levels toward phenotype-based drug discovery. Another example was the UNaProd database which contains information concerning 3,411 natural products used in Iranian traditional medicine (ITM) (Naghizadeh et al., 2020).

Clustering has been commonly used to identify subpopulations of patients with distinctive genetic variants or gene expression profiles. For example, Naval et al. showed how clustering analysis helped identifying single nucleotide polymorphisms (SNPs) associated with skin properties (Naval et al., 2014). Combining transcriptomic data with images, Voineagu et al. showed how the clustering methods characterize distinct complex disease subtypes in autism spectrum disorder (Voineagu et al., 2011). Additionally, in clinical proteomics, clustering analysis also identifies a group of proteins as functional modules in pathogenesis. For example, Baldelli et al. clustered non-small cell lung cancer tumors according to the expression, activation, and phosphorylation levels of 26 signaling proteins (Baldelli et al., 2015). Implementing clustering methods in the context of precision medicine is not only applicable to omics data, but also to physiological data. For example, Xu et al. developed human stress management using clustering of physiological signals during series of task-rest cycles (Xu et al., 2015). On the other hand, image data as a major part of health records of individuals is commonly utilized (Hsu et al., 2013). For example, Enguehard et al. presented a strategy of integrating neural network and clustering analysis for automatic magnetic resonance imaging data analysis (Enguehard et al., 2019). Furthermore, it has been shown that utilizing biomedical annotations can potentially improve clustering analysis to obtain more biologically relevant disease categories (Futschik and Carlisle, 2005; Bandyopadhyay et al., 2007; Lee, 2011). Therefore, the integration of existing biomedical annotations, such as gene ontology or pathway enrichment, is also expected to improve patient disease clustering with refined distance functions (Handl et al., 2005).

As abovementioned, exploring subclasses of diseases and drugs is a prevalent task in precision medicine, and traditional medicine is no exception. For example, Liu et al. studied the gene expression signature of breast cancer cell lines for an herbal formula Si-Wu-Tang (SWT). This analysis showed that the effect of SWT is comparable to β-estradiol treatment on estrogen-responsive genes (Liu et al., 2013). Ruan et al. proposed a clustering algorithm called THCluster that can effectively discover meaningful categorization of herbs and their potential clinical indications (Ruan et al., 2017). Zhang et al. validated TCM syndrome types using a clustering method based on latent tree models, based on which they proposed a standard for syndrome

differentiation in TCM which was then validated successfully in a study of kidney deficiency (Zhang et al., 2008). Likewise, Zhao et al. proposed a top-down subspace clustering for improving the precision of syndrome differentiation. Considering 5,600 symptoms and 150 syndrome elements of AIDS (acquired immune deficiency syndrome) patients, they showed that their method identified clusters of patients more precisely, compared to conventional clustering algorithms such as k-means (Zhao et al., 2014).

While identifying the heterogeneity of patients is critical, understanding the driving molecular mechanisms of such heterogeneity shall provide more rational on the design of precision medicine. To understand the underlying factors that are shared by patients with similar diseases, more information about the interaction of biological entities including genes, proteins, and drugs is required. By introducing network models, such a complex layer of information can be systematically evaluated, for which clustering analysis may further help infer the distinctive disease patterns. In the following we focused on the combination of network modeling and clustering analysis and showed that how they may contribute to the understanding for precision medicine.

EXPLORING NETWORK MODULES AS A BASIS FOR CLASSIFICATION

A simple network consists of a set of elements called nodes or vertices which are connected by a set of links or edges (Jafari et al., 2013). Depending on the definition of node and edge sets, numerous types of biological networks can be constructed and used for further analysis. For example, the degree of a node, which is defined as the number of links attached to the node, suggests the importance of node and helps detect global and provincial hubs within the network. The heterogeneity of a network which is defined as the root of the variance of degrees divided by their mean, also explains the overall topology of the network and organization of relationships among the nodes (Dong and Horvath, 2007).

One of the major network modeling approaches is the modularity analysis or community detection which is the intersection of clustering analysis and network science (Fortunato, 2010; Fortunato and Hric, 2016). In this analysis, exploring the local densely connected nodes, i.e., networks community structure is the main aim. In other words, a community within a given network includes nodes with high intra-relationship and low inter-relationship with the other nodes outside the community (Girvan and Newman, 2002). Therefore, finding network modules is important to elucidate and understand the complex topology of networks by discriminating dense and sparse local structures. The network topology determines the adjacency matrix, which can be utilized for clustering analysis alone or in combination with the similarity matrix derived from the node properties (Von Luxburg, 2006; Fortunato, 2010). Since in real biological and disease networks, there are multifunctional nodes belonging to more than one group, soft clustering is also recommended (Yang and Leskovec,

2015). In the soft clustering, overlapping communities, also called covers, are detectable because the multiple memberships for a node are allowed. There are other methods including dynamical clustering (Jeub et al., 2015). For example, the Markov Clustering algorithm (MCL) is one of the commonly used dynamic clustering algorithms based on biological annotations (Jafari et al., 2015). After exploring the modules of a given network, it is common to convert the network into its reduced version, where a node of the reduced network corresponds to a module of the original network, and an edge is inferred from the number of interactions between the modules (**Figure 2**).

Utilizing molecular information in conjunction with the module detection to predict biological functions is a common task (Vespignani, 2003; Jafari et al., 2015). The main assumption is that members of the same cluster tend to be involved in the same biological process which is known as guilt-by-association. For example, in protein-protein interaction networks, clique-based clustering was used to detect protein complexes (Altaf-Ul-Amin et al., 2006; Phan and Sternberg, 2012; Jafari et al., 2013). Hierarchical clustering was utilized to identify signaling cascades or metabolic pathways (Guimera and Amaral, 2005; Koch and Ackermann, 2013; Azimzadeh Jamalkandi et al., 2016). Also, according to the local topological features of biological networks, clustering methods are commonly used to predict cellular co-localization and co-expressed gene regulatory mechanisms (Dittrich et al., 2008; Amiri et al., 2013; Mitra et al., 2013). At the phenotypic level, network modeling linking phenotypes to molecular components of a biological system, e.g., disease-causing genetic variations is also one of the exciting research areas (Goh et al., 2007; Loscalzo and Barabasi, 2011; Goh and Choi, 2012; Emmert-Streib et al., 2013). In the context of traditional medicine, Huang et al. highlighted how network pharmacology modeling allows us to integrate concurrent and traditional knowledge of herbal medicines for the development of new drugs for complex human diseases (Huang et al., 2013). Using a network-based integration of chemical structure and omics data, they inferred novel drug-disease interactions *via* molecular targets and pathways. Similarly, Li et al. introduced a distance-based mutual information model to score herb interactions based on their frequencies and distances, and thus identify the rationale of herb combinations (Li et al., 2010).

Network biology approaches have also shown potential for exploring disease subcategories and patient subclasses in TCM. For instance, Zhou et al. constructed a clinical phenotype network to investigate the underlying mechanisms of TCM diagnosis and treatment (Zhou et al., 2014). Wang et al. proposed a co-occurrence network approach to identify the TCM symptoms as biomarkers for the fatty liver disease (Wang W. et al., 2019). Interestingly, Jiang et al. also demonstrated the association between the TCM symptoms and tongue-coating microbiome using co-occurrence networks (Jiang et al., 2012). Network modeling of cold and hot syndromes of traditional medicine has also been developed. For example, Ma et al. provided a gene expression signature of the cold syndrome in TCM associated with the neuroendocrine-immune system. By analyzing the protein interaction networks, they showed that the genes related to the cold syndrome are involved in pathways of energy metabolism, neurotransmitters, hormones, and cytokines (Ma et al., 2010). Likewise, Lu et al. provided distinctive molecular signatures in CD4-positive T cells of Rheumatoid Arthritis patients associated with the cold and heat patterns in TCM respectively (Lu et al., 2012a; Lu et al., 2012b).

MULTIPARTITE NETWORK MODELS FOR INTEGRATING HETEROGENEOUS DATA

With the development of high-throughput technologies, precision medicine has been made more plausible with increasingly diversified data sets. These data sets range from gene expression profiles to medical images, where the scales, characteristics, and formats are different since they are gathered at the different levels of biological systems (Lee, 2011). The integration of information from these heterogeneous biological and clinical data sets need to be applied in order to discover new mechanistic insights of systems medicine. For example, to predict more effective disease
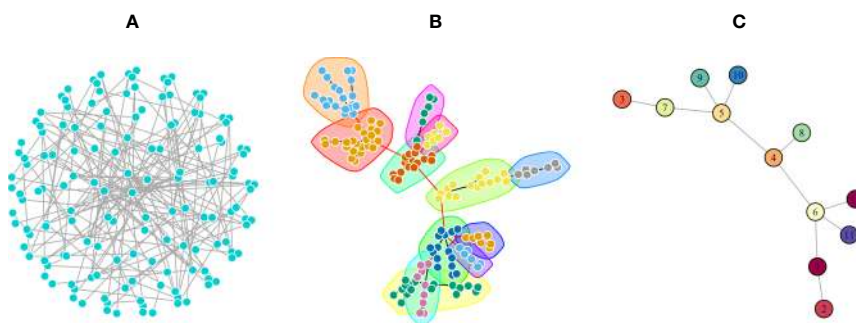

FIGURE 2 | A Poisson-distributed random network is represented in three formats. **(A)** A complex view of the network by placing nodes on a sphere layout that obscure the complexity of the network topology. **(B)** The same network using the force-directed layout algorithm in which communities were identified *via* a greedy optimization of modularity score. This representation is usually used to show the modular structure of the network. **(C)** The simplified network in which each node represents a community in the original network, and each edge denotes the interaction of communities.

treatment options using multi-targeted drug combinations (Tang and Aittokallio, 2014), we need to gather multiple data types such as *in vitro* drug response of cancer cells and *in vivo* response of patients including symptoms and molecular profiles. To predict the effectiveness of drug combinations, understanding about the signaling pathways and drug target interactions along with the pathophysiological states is essential. There is a major type of network models called multipartite networks which are commonly used in systems medicine (Junker and Schreiber, 2008). This kind of network modeling is crucial due to its flexibility to integrate mixed datasets and discover complex hidden relationships which are required for understanding precision medicine. Unlike ordinary uni-partite networks which contain single sets of nodes and edges represented by an adjacency matrix, a multipartite network constitutes of multiple sets of nodes and edges which are exemplified by incidence matrix (Agnarsson and Greenlaw, 2007). Depending on the data types, the network can represent gene-disease (Bauer-Mehren et al., 2011; Barneh et al., 2016; Chen et al., 2018), drug-target (Barneh et al., 2016), protein-cell localization (Mirzaei Mehrabad et al., 2018), drug-disease (Lamb, 2007) and drug-side effect associations (Luo et al., 2014), as well as associations at the patient level including patient-drug interactions and patient-symptom interactions (Bhavnani et al., 2010).
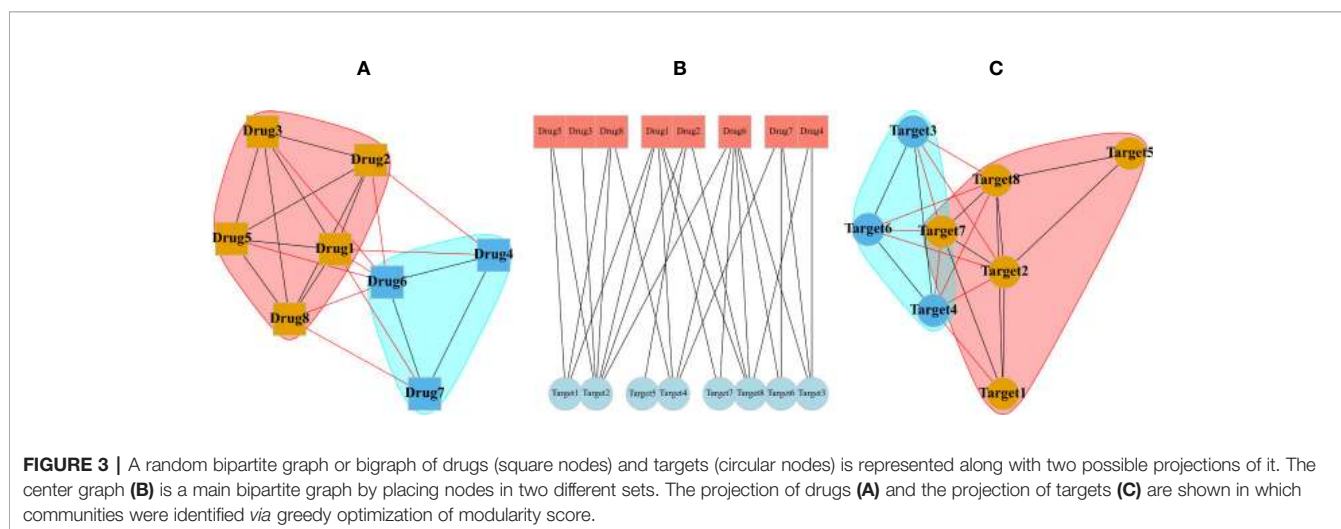
Based on the constructed multipartite networks, different kinds of clustering algorithms can be applied to identify the hidden subnetwork structures for each node set. For instance, Long et al. proposed a clustering method by a combination of co-clustering and probabilistic hidden Markov models (Long et al., 2007). Also, Hartsperger et al. developed a fuzzy multipartite clustering to decompose the nodes of multiple types in tripartite networks (Hartsperger et al., 2010). They showed that the fuzzy clustering algorithm was able to identify functionally correlated modules of a tripartite gene-disease-protein complex network for the identification of biologically meaningful clusters. Duan et al. identified two major subtypes of breast cancer by reconstructing a tripartite graph of drug-cell line-patient tumors. They showed how drug response data helped discover dysregulated pathways

for breast cancer (Duan et al., 2013). A multipartite network can be also utilized as a visualization tool, with which one can navigate efficiently the high-throughput drug response data from public databases including Cancer Cell Line Encyclopedia (CCLE) and Genomics of Drug Sensitivity in Cancer (GDSC) (Duan et al., 2014).

Typical multipartite network analyses involve network projection, which aims for simplifying the network topology from the viewpoint of each node set separately. In **Figure 3**, the projection of a schematic bipartite network and module detection analysis is briefly presented. Constructing the projected networks facilitates the exploration of hidden relationships among each set of nodes in a multipartite network. For example, Barneh et al. constructed a drug-target network and further developed its projected version called Drug Similarity Network (DSN) and Target Similarity Network (TSN), which can be used for drug-target prediction (Barneh et al., 2016). Recently, they have applied the method to predict drug combinations, and confirmed them experimentally (Barneh et al., 2018; Barneh et al., 2019). To facilitate drug repositioning, network topological similarity-based inference (NTSIM) and its classification-equipped version, i.e., NTSIM-C methods were also proposed to unveil novel drug-disease associations (Zhang et al., 2018).

NETWORK ANALYSIS RATIONALIZES TCM CLASSIFICATIONS: A CASE STUDY

The idea of utilizing multipartite networks in traditional medicine is potentially feasible, as the data standardization and annotation has been increasingly pursued. However, to the best of our knowledge, the models are not yet utilized to provide more profound insights on traditional medicine, although some tools such as SymMap (Wu et al., 2019) may provide an appropriate dataset to build such multipartite networks. In the following, we conducted a case study to reconstruct a bipartite network of natural products and ingredients of TCM to show the potential

FIGURE 3 | A random bipartite graph or bigraph of drugs (square nodes) and targets (circular nodes) is represented along with two possible projections of it. The center graph **(B)** is a main bipartite graph by placing nodes in two different sets. The projection of drugs **(A)** and the projection of targets **(C)** are shown in which communities were identified *via* greedy optimization of modularity score.

of this modeling for understanding TCM rationale for disease treatment and drug discovery.

TCM-related databases provide a large set of information about the TCM herbs including their classifications and disease indications, as well as molecular characterization, such as ingredient profiles and molecular targets. Recently, this information has been successfully applied for developing computational models to understand the TCM classifications (Wang Y. et al., 2019). As a case study to show the potential of multipartite networks in the integration of heterogonous data, we obtained a list of 4,485 natural products consisting of 2,857 chemical ingredients from the TCMID database (Huang et al., 2018). We used the second version of the TCMID database, as the largest dataset in this field, which contains richer experimental data originating from ingredient-specific and herbal mass spectrometry spectra. The natural products and ingredients are considered to be the two parts of a bipartite network. After removal of disconnected nodes, we extracted a giant component of the graph consisting of 7,004 nodes and 17,555 edges. Following the projection of this bigraph as outlined in **Figure 3**, two projected graphs called the natural product similarity network (NSN) and ingredient similarity network (ISN) were reconstructed, such that each edge indicates at least one common ingredient or natural product in NSN and ISN, respectively. The NSN contains 4,308 natural products and 204,807 edges, while the ISN consists of 2,696 nodes and 78,228 edges. The community detection was subsequently done for both similarity networks *via* optimizing a modularity score (Clauset et al., 2004), resulting in 42 and 24 communities for NSN and ISN, separately. The fast greedy algorithm outperformed compared to the other high-performance algorithms, i.e., infomap and walktrap (Labatut and Balasque, 2013; Wagenseller et al., 2018) according to the highest average of modularity index in the NSN and ISN (**Supplementary file 2**). These communities reflect the internal similarity of herbs and ingredients which could be investigated further. For example, a community of NSN indicates a set of natural products with similar profiles of ingredients. Therefore, the members of the same natural product cluster can be used for therapeutic interchanging due to the similarity of ingredient profiles in the cluster. Also, members of different clusters can be candidates for new drug combinations as they are expected to affect distinctive biological pathways. Similarly, the cluster of active ingredients in ISN can be used to predict the mechanism of action of newly discovered or synthesized compounds based on TCM classifications. In other words, a functionally-unknown molecule with high structural similarity to any of active ingredient clusters indicate the analogous TCM properties and implications. Therefore, any follow-up experimental analysis can be prioritized to disclose therapeutic hits of the new molecules based on known properties and implications of the corresponding cluster. Also, the priority of active ingredients for treatment can be redefined in each cluster independently using availability, and the relevant protein targets characterizations (code and data set for this case study can be found in **Supplementary File 1**).

We sought to validate our prediction about the herb and ingredient communities, i.e., whether the herbs or active ingredients that are clustered in the same community tend to share similar features. Four types of features for the natural products and their ingredients including meridians and properties were extracted from TCMID. Furthermore, the SMILE strings of these ingredients along with the identified or predicted protein targets of them were extracted using PubChem (Kim et al., 2018) and STITCH databases (Szklarczyk et al., 2016). Then, the average of pairwise intersection of meridian and property profiles was computed separately for each cluster in NSN. Likewise, the average similarity of SMILE string using the Dice index and the pairwise intersection of their protein targets in each cluster of ISN were also calculated. We showed in **Figure 4** the average similarity of all the 42 and 24 communities in NSN (**Figures 4A, B**) and ISN (**Figures 4C, D**), as compared to that of random clustering from 100 simulations. We found that the similarity of natural products or active ingredients within a cluster is significantly higher than that for the random clustering. For example, the median of meridian-based similarity of random grouping is 0.56, while the median similarity of the 42 clusters found in NSN is 0.96 (p-value = 1.99e-05, Wilcoxon test). Similarly, in ISN, the median of the Dice similarity of SMILE strings in random groups is 0.22, while the median of the 24 clusters in ISN is 0.36 (p-value = 5.84e-05, Wilcoxon test). Our findings suggested that the clusters of ISN and NSN consist of similar ingredients or natural products, and thus validating the feasibility of bipartite modeling in analyzing TCM data.

Interestingly, these network analyses also suggested a molecular basis of TCM classifications, which originated from the physical features of natural products or empirical knowledge about the disease indications. Although the chemical and molecular characteristics of the natural products, i.e., chemical structures and protein targets have only been available recently, the TCM classification according to meridians were indeed associated with them. The same observation was found for the property classification in TCM in our findings, as the natural products in a given cluster based on ingredient profiles are associated with their property profiles. On the other hand, our approach promises to bridge a gap between pharmaceutical chemistry and traditional pharmacology in TCM. For example, we can use attributes of active ingredient profile of natural products as a rich training set, and newly discovered, or synthesized molecules can be characterized accordingly as a test set. To summarize, this bipartite network analysis provides novel insights for the understanding of molecular evidence of traditional classification in TCM. Using the bipartite network modeling, we may integrate phenotypes of different types, i.e., signs and symptoms, with the chemical knowledge of drug molecules in order to provide a formal framework for phenotype-based drug discovery in TCM.

SUMMARY AND OUTLOOK

Nowadays, TCM, along with the other traditional medical schools, was modernized and expanded by the molecular shreds of evidence provided by experimental biology (Xue et al., 2013). These
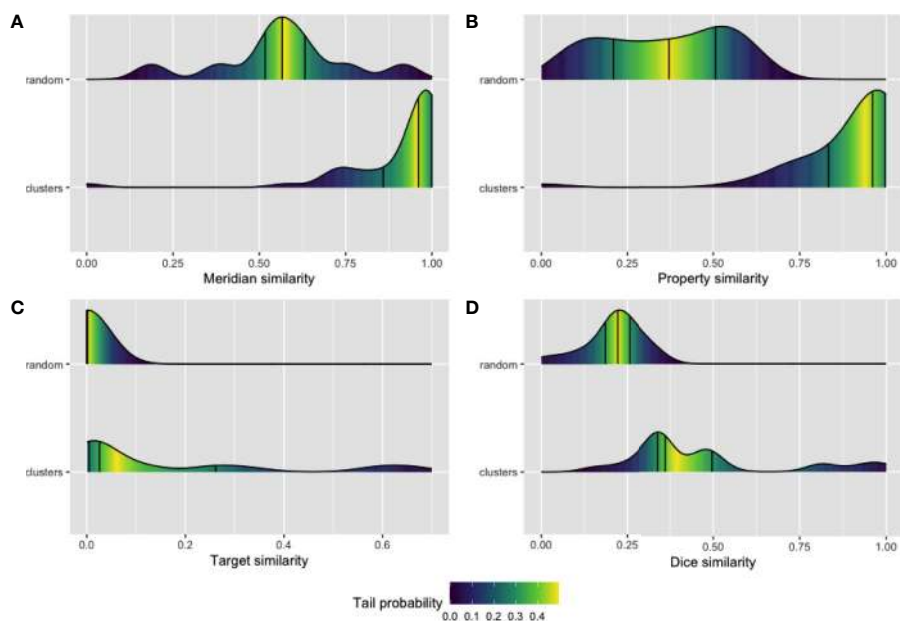
FIGURE 4 | Validation of the network communities in TCM herbs. The distribution of Meridian similarity **(A)** and Property similarity **(B)** within the communities of NSN (natural product similarity network) compared to random groups; The distribution of drug target similarity **(C, D)** the Dice index similarity within the communities of ISN (ingredient similarity network) compared to random groups. Lines indicate the quartiles values and gradient color represent the probability of values within the empirical cumulative density function for the distributions.

experiments usually are started by extraction and fractionation techniques such as chromatography, and followed by identification methods such as mass spectrometry to determine a comprehensive profile of ingredients within the natural products (Jafari et al., 2016; Kabiri et al., 2017). The challenging part of this experimental design is identifying the ingredients responsible for the bioactivities of natural products. To further explore the drug-target interactions of these ingredients, high-throughput omics is now a preferable approach to study the effects on gene expression and protein activity. Through these experimental techniques, large volumes of molecular features related to disease indications can be disclosed, including antimicrobial, antiviral, antioxidant, anti-inflammatory, and neurological activities (Neghabi-Hajiagha et al., 2016; Pourramezan et al., 2018). In the next step, we always face the challenge of applying appropriate data integration methods to associate these molecular features with the phytochemical and pharmacological properties of the TCM ingredients.

A remarkable portion of biological studies deal with generating, organizing, and retrieving patient data, which is usually large scale and noisy. Data mining algorithms such as clustering and classification are being applied. Furthermore, the integration of heterogeneous biological data is imperative. Rigorous and efficient analysis tools are required for the integration of different data characteristics and formats as standard statistical inference techniques may be limited (Lee, 2011; Eric, 2014). Harnessing the network modeling in computational biology becomes a feasible strategy for data integration to navigate the complex space of

biological systems (Barabási et al., 2011). Here, we highlighted the application of the multipartite network reconstruction for data integration in biomedical researches, particularly in traditional medicine. More specifically, we demonstrated how we can combine current chemical knowledge of ingredients and TCM classifications of natural products to bridge the gap between traditional and modern medicine. We provided a case study to show its potentials for uncovering TCM concepts and discovering potential treatments.

Although network science, and more precisely, network medicine is on its developing stage, using multipartite network modeling may provide more rational on the therapies in traditional medicine. Generally in traditional medicine, much efforts are spent on collecting the symptoms of patients, while in modern medicine, biochemical profiles and image data are more relied on. Providing a framework for integrating all these data using multipartite network model shall facilitate the interchange of knowledge from traditional and modern medicine. Reconstructing multipartite networks is a convenient way to characterize patient similarity, which serves the basis for further explorations on their diseases mechanisms (Pai and Bader, 2018). Depending on the nature of the node sets, a multipartite network can be utilized to investigate complex interactions, that might be critical for understanding diseases with high-level patient heterogeneity such as cancer (Yaffe, 2019). Considering all available data from cellular behaviors to patient responses using multipartite network modeling can play a significant role

in the integration of these heterogeneous datasets, a successful application of which may make precision medicine a reality ultimately.

AUTHOR CONTRIBUTIONS

Study design: MJ, YW, JT. Data analysis: YW, MJ. Writing: MJ, YW, AA, JT.

FUNDING

ACKNOWLEDGMENTS

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fphar.2020.01319/full#supplementary-material

SUPPLEMENTARY FILE 1 | R code and dataset for section "Network analysis rationalizes TCM classifications: a case study".

SUPPLEMENTARY FILE 2 | Modularity index of the clustering algorithms used for the present case study.

REFERENCES

Agnarsson, G., and Greenlaw, R. (2007). *Graph theory: Modeling, applications, and algorithms* (Upper Saddle River, New Jersey: Pearson/Prentice Hall).

Altaf-Ul-Amin, M., Shinbo, Y., Mihara, K., Kurokawa, K., and Kanaya, S. (2006). Development and implementation of an algorithm for detection of protein complexes in large interaction networks. *BMC Bioinf.* 7, 207. doi: 10.1186/1471-2105-7-207

Amiri, M., Jafari, M., Azimzadeh Jamalkandi, S., and Davoodi, S. M. (2013). Atopic dermatitis-associated protein interaction network lead to new insights in chronic sulfur mustard skin lesion mechanisms. *Expert Rev. Proteomics* 10, 449–460. doi: 10.1586/14789450.2013.841548

Arji, G., Safdari, R., Rezaeizadeh, H., Abbassian, A., Mokhtaran, M., and Hossein Ayati, M. (2019). A systematic literature review and classification of knowledge discovery in traditional medicine. *Comput. Methods Prog. Biomed.* 168, 39–57. doi: 10.1016/j.cmpb.2018.10.017

Azimzadeh Jamalkandi, S., Mozhgani, S. H., Gholami Pourbadie, H., Mirzaie, M., Noorbakhsh, F., Vaziri, B., et al. (2016). Systems biomedicine of rabies delineates the affected signaling pathways. *Front. Microbiol.* 7, 1688. doi: 10.1101/068817

Baldelli, E., Haura, E. B., Crino, L., Cress, D. W., Ludovini, V., Schabath, M. B., et al. (2015). Impact of upfront cellular enrichment by laser capture microdissection on protein and phosphoprotein drug target signaling activation measurements in human lung cancer: Implications for personalized medicine. *Proteomics Clin. Appl.* 9, 928–937. doi: 10.1002/prca.201400056

Bandyopadhyay, S., Mukhopadhyay, A., and Maulik, U. (2007). An improved algorithm for clustering gene expression data. *Bioinformatics* 23, 2859–2865. doi: 10.1093/bioinformatics/btm418

Barabási, A. L., Gulbahce, N., and Loscalzo, J. (2011). Network medicine: a network-based approach to human disease. *Nat. Rev. Genet.* 12, 56–68. doi: 10.1038/nrg2918

Barneh, F., Jafari, M., and Mirzaie, M. (2016). Updates on drug-target network; facilitating polypharmacology and data integration by growth of DrugBank database. *Briefings Bioinf.* 17, 1070–1080. doi: 10.1093/bib/bbv094

Barneh, F., Salimi, M., Goshadrou, F., Ashtiani, M., Mirzaie, M., Zali, H., et al. (2018). Valproic acid inhibits the protective effects of stromal cells against chemotherapy in breast cancer: Insights from proteomics and systems biology. *J. Cell. Biochem.* 119, 9270–9283. doi: 10.1002/jcb.27196

Barneh, F., Mirzaie, M., Nickchi, P., Tan, T. Z., Thiery, J. P., Piran, M., et al. (2019). Integrated use of bioinformatic resources reveals that co-targeting of histone deacetylases, IKBK and SRC inhibits epithelial-mesenchymal transition in cancer. *Briefings Bioinf.* 20, 717–731. doi: 10.1093/bib/bby030

Bauer-Mehren, A., Bundschus, M., Rautschka, M., Mayer, M. A., Sanz, F., and Furlong, L. I. (2011). Gene-disease network analysis reveals functional modules in mendelian, complex and environmental diseases. *PloS One* 6, e20284–e20284. doi: 10.1371/journal.pone.0020284

Bhavnani, S. K., Bellala, G., Ganesan, A., Krishna, R., Saxman, P., Scott, C., et al. (2010). The nested structure of cancer symptoms. *Methods Inf. Med.* 49, 581–591. doi: 10.3414/ME09-01-0083

Chen, X., Xie, D., Wang, L., Zhao, Q., You, Z. H., and Liu, H. (2018). BNPMDA: bipartite network projection for MiRNA–disease association prediction. *Bioinformatics* 34, 3178–3186. doi: 10.1093/bioinformatics/bty333

Clauset, A., Newman, M. E., and Moore, C. (2004). Finding community structure in very large networks. *Phys. Rev. E.* 70, 066111. doi: 10.1103/PhysRevE.70.066111

De Fauw, J., Ledsam, J. R., Romera-Paredes, B., Nikolov, S., Tomasev, N., Blackwell, S., et al. (2018). Clinically applicable deep learning for diagnosis and referral in retinal disease. *Nat. Med.* 24, 1342. doi: 10.1038/s41591-018-0107-6

Dittrich, M. T., Klau, G. W., Rosenwald, A., Dandekar, T., and Müller, T. (2008). Identifying functional modules in protein–protein interaction networks: an integrated exact approach. *Bioinformatics* 24, i223–i231. doi: 10.1093/bioinformatics/btn161

Dong, J., and Horvath, S. (2007). Understanding network concepts in modules. *BMC Syst. Biol.* 1, 24. doi: 10.1186/1752-0509-1-24

Duan, Q., Kou, Y., Clark, N., Gordonov, S., and Ma'ayan, A. (2013). Metasignatures identify two major subtypes of breast cancer. *CPT: Pharmacometr. Syst. Pharmacol.* 2, 1–10. doi: 10.1038/psp.2013.11

Duan, Q., Wang, Z., Fernandez, N. F., Rouillard, A. D., Tan, C. M., Benes, C. H., et al. (2014). Drug/Cell-line Browser: interactive canvas visualization of cancer drug/cell-line viability assay datasets. *Bioinformatics* 30, 3289–3290. doi: 10.1093/bioinformatics/btu526

Emmert-Streib, F., Tripathi, S., R.d.M. Simoes, A. F., and Dehmer, M. (2013). The human disease network: Opportunities for classification, diagnosis, and prediction of disorders and disease genes. *Syst. Biomed.* 1, 20–28. doi: 10.4161/sysb.22816

Enguehard, J., O'Halloran, P., and Gholipour, A. (2019). Semi-Supervised Learning With Deep Embedded Clustering for Image Classification and Segmentation. *IEEE Access* 7, 11093–11104. doi: 10.1109/ACCESS.2019.2891970

Eric, J. (2014). Topol, Individualized Medicine from Prewomb to Tomb. *Cell* 157, 241–253. doi: 10.1016/j.cell.2014.02.012

Fortunato, S., and Hric, D. (2016). Community detection in networks: A user guide. *Phys. Rep.* 659, 1–44. doi: 10.1016/j.physrep.2016.09.002

Fortunato, S. (2010). Community detection in graphs. *Phys. Rep.* 486, 75–174. doi: 10.1016/j.physrep.2009.11.002

Futschik, M. E., and Carlisle, B. (2005). Noise-robust soft clustering of gene expression time-course data. *J. Bioinf. Comput. Biol.* 3, 965–988. doi: 10.1142/S0219720005001375

Girvan, M., and Newman, M. E. (2002). Community structure in social and biological networks. *Proc. Natl. Acad. Sci.* 99, 7821–7826. doi: 10.1073/pnas.122653799

Goh, K. I., and Choi, I. G. (2012). Exploring the human diseasome: the human disease network. *Briefings Funct. Genomics* 11, 533–542. doi: 10.1093/bfgp/els032

Goh, K. I., Cusick, M. E., Valle, D., Childs, B., Vidal, M., and Barabási, A.-L. (2007). The human disease network. *Proc. Natl. Acad. Sci.* 104, 8685–8690. doi: 10.1073/pnas.0701361104

Guimera, R., and Amaral, L. A. N. (2005). Functional cartography of complex metabolic networks. *nature* 433, 895. doi: 10.1038/nature03288

Handl, J., Knowles, J., and Kell, D. B. (2005). Computational cluster validation in post-genomic data analysis. *Bioinformatics* 21, 3201–3212. doi: 10.1093/bioinformatics/bti517

Hartsperger, M. L., Blöchl, F., Stümpflen, V., and Theis, F. J. (2010). Structuring heterogeneous biological information using fuzzy clustering of k-partite graphs. *BMC Bioinf.* 11, 522. doi: 10.1186/1471-2105-11-522

Hsu, W., Markey, M. K., and Wang, M. D. (2013). Biomedical imaging informatics in the era of precision medicine: progress, challenges, and opportunities. *J. Am. Med. Inf. Assoc.* 20, 1010–1013. doi: 10.1136/amiajnl-2013-002315

Huang, C., Zheng, C., Li, Y., Wang, Y., Lu, A., and Yang, L. (2013). Systems pharmacology in drug discovery and therapeutic insight for herbal medicines. *Briefings Bioinf.* 15, 710–733. doi: 10.1093/bib/bbt035

Huang, L., Xie, D., Yu, Y., Liu, H., Shi, Y., Shi, T., et al. (2018). TCMID 2.0: A comprehensive resource for TCM. *Nucleic Acids Res.* 46, D1117–D1120. doi: 10.1093/nar/gkx1028

Jafari, M., Sadeghi, M., Mirzaie, M., Marashi, S. A., and Rezaei-Tavirani, M. (2013). Evolutionarily conserved motifs and modules in mitochondrial protein-protein interaction networks. *Mitochondrion* 13, 668–675. doi: 10.1016/j.mito.2013.09.006

Jafari, M., Mirzaie, M., and Sadeghi, M. (2015). Interlog protein network: an evolutionary benchmark of protein interaction networks for the evaluation of clustering algorithms. *BMC Bioinf.* 16, 319–319. doi: 10.1186/s12859-015-0755-1

Jafari, M., Mirzaie, M., Khodabandeh, M., Rezadoost, H., Ghassempour, A., and Aboul-Enein, H. Y. (2016). Polarity-based fractionation in proteomics: hydrophilic interaction vs reversed-phase liquid chromatography. *Biomed. Chromatogr.* 30, 1036–1041. doi: 10.1002/bmc.3647

Jafari, M., Ansari-Pour, N., Azimzadeh, S., and Mirzaie, M. (2017). A logic-based dynamic modeling approach to explicate the evolution of the central dogma of molecular biology. *PloS One* 12, e0189922–e0189922. doi: 10.1371/journal.pone.0189922

Jeub, L. G., Balachandran, P., Porter, M. A., Mucha, P. J., Mahoney, M. W., and locally, T. (2015). act locally: Detection of small, medium-sized, and large communities in large networks. *Phys. Rev. E.* 91, 012821. doi: 10.1103/PhysRevE.91.012821

Jiang, B., Liang, X., Chen, Y., Ma, T., Liu, L., Li, J., et al. (2012). Integrating next-generation sequencing and traditional tongue diagnosis to determine tongue coating microbiome. *Sci. Rep.* 2, 936. doi: 10.1038/srep00936

Junker, B. H., and Schreiber, F. (2008). *Analysis of biological networks, Wiley Online Library.* (John Wiley & Sons, Inc.). doi: 10.1002/9780470253489

Kabiri, M., Rezadoost, H., and Ghassempour, A. (2017). A comparative quality study of saffron constituents through HPLC and HPTLC methods followed by isolation of crocins and picrocrocin. *LWT* 84, 1–9. doi: 10.1016/j.lwt.2017.05.033

Keenan, A. B., Jenkins, S. L., Jagodnik, K. M., Koplev, S., He, E., Torre, D., et al. (2018). The library of integrated network-based cellular signatures NIH program: system-level cataloging of human cells response to perturbations. *Cell Syst.* 6, 13–24. doi: 10.1016/j.cels.2017.11.001

Kim, S., Chen, J., Cheng, T., Gindulyte, A., He, J., He, S., et al. (2018). PubChem 2019 update: improved access to chemical data. *Nucleic Acids Res.* 47, D1102–D1109. doi: 10.1093/nar/gky1033

Koch, I., and Ackermann, J. (2013). On functional module detection in metabolic networks. *Metabolites* 3, 673–700. doi: 10.3390/metabo3030673

Labatut, V., and Balasque, J. M. (2012). Detection and Interpretation of Communities in Complex Networks: Methods and Practical Application. *Computational Social Networks: Tools, Perspectives and Applications.* (Springer), 81–113. doi: 10.1007/978-1-4471-4048-1_4

Lamb, J. (2007). The Connectivity Map: a new tool for biomedical research. *Nat. Rev. Cancer* 7, 54–60. doi: 10.1038/nrc2044

Langlois, R., Pallesen, J., and Frank, J. (2011). Reference-free particle selection enhanced with semi-supervised machine learning for cryo-electron microscopy. *J. Struct. Biol.* 175, 353–361. doi: 10.1016/j.jsb.2011.06.004

Lee, J. K. (2011). *Statistical bioinformatics: for biomedical and life science researchers* (John Wiley & Sons).

Li, F. S., and Weng, J. K. (2017). Demystifying traditional herbal medicine with modern approach. *Nat. Plants* 3, 1–7. doi: 10.1038/nplants.2017.109

Li, S., Zhang, B., Jiang, D., Wei, Y., and Zhang, N. (2010). Herb network construction and co-module analysis for uncovering the combination rule of traditional Chinese herbal formulae. *BMC Bioinf.* 11, S6. doi: 10.1186/1471-2105-11-S11-S6

Liu, M., Fan, J., Wang, S., Wang, Z., Wang, C., Zuo, Z., et al. (2013). Transcriptional profiling of Chinese medicinal formula Si-Wu-Tang on breast cancer cells reveals phytoestrogenic activity. *BMC Complement. Altern. Med.* 13, 11. doi: 10.1186/1472-6882-13-11

Long, B., Zhang, Z. M., and Yu, P. S. (2007). *A probabilistic framework for relational clustering, Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining* (California, USA: ACM, San Jose), 470–479.

Loscalzo, J., and Barabasi, A. L. (2011). Systems biology and the future of medicine. *Wiley Interdiscip. Rev.: Syst. Biol. Med.* 3, 619–627. doi: 10.1002/wsbm.144

Lu, C., Niu, X., Xiao, C., Chen, G., Zha, Q., Guo, H., et al. (2012a). Network-based gene expression biomarkers for cold and heat patterns of rheumatoid arthritis in traditional chinese medicine. *Evidence-Based Complement. Altern. Med.* 2012, 203043. doi: 10.1155/2012/203043

Lu, C., Xiao, C., Chen, G., Jiang, M., Zha, Q., Yan, X., et al. (2012b). Cold and heat pattern of rheumatoid arthritis in traditional Chinese medicine: distinct molecular signatures indentified by microarray expression profiles in CD4-positive T cell. *Rheumatol. Int.* 32, 61–68. doi: 10.1007/s00296-010-1546-7

Luo, Y., Liu, Q., Wu, W., Li, F., and Bo, X. (2014). "Predicting drug side effects based on link prediction in bipartite network," in *2014 7th International Conference on Biomedical Engineering and Informatics.* (IEEE), 729–733.

Ma, T., Tan, C., Zhang, H., Wang, M., Ding, W., and Li, S. (2010). Bridging the gap between traditional Chinese medicine and systems biology: the connection of Cold Syndrome and NEI network. *Mol. Biosyst.* 6, 613–619. doi: 10.1039/b914024g

Manolio, T. A. (2018). UK Biobank debuts as a powerful resource for genomic research. (Nature Publishing Group). *Nature Med.* 24 (12), 1792–1794.

Mirzaei Mehrabad, E., Hassanzadeh, R., and Eslahchi, C. (2018). PMLPR: A novel method for predicting subcellular localization based on recommender systems. *Sci. Rep.* 8, 12006. doi: 10.1038/s41598-018-30394-w

Mirzaeian, R., Sadoughi, F., Tahmasebian, S., and Mojahedi, M. (2019). Progresses and challenges in the traditional medicine information system: A systematic review. *J. Pharm. Pharmacogn. Res.* 7, 246–259.

Mitra, K., Carvunis, A. R., Ramesh, S. K., and Ideker, T. (2013). Integrative approaches for finding modular structure in biological networks. *Nat. Rev. Genet.* 14, 719. doi: 10.1038/nrg3552

Naghizadeh, A., Hamzeheian, D., Akbari, S., Mohammadi, F., Otoufat, T., Asgari, S., et al. (2020). UNaProd: A Universal Natural Product Database for Materia Medica of Iranian Traditional Medicine. *Evid. Based. Complement. Alternat. Med. Accepted.* 2020. doi: 10.1155/2020/3690781

Naval, J., Alonso, V., and Herranz, M. A. (2014). Genetic polymorphisms and skin aging: The identification of population genotypic groups holds potential for personalized treatments. *Clin. Cosmetic Investigational Dermatol.* 7, 207–214. doi: 10.2147/CCID.S55669

Neghabi-Hajiagha, M., Aliahmadi, A., Taheri, M. R., Ghassempour, A., Irajian, G., Rezadoost, H., et al. (2016). A bioassay-guided fractionation scheme for characterization of new antibacterial compounds from Prosopis cineraria aerial parts. *Iranian J. Microbiol.* 8, 1.

Nussinov, R., Jang, H., Tsai, C. J., and Cheng, F. (2019). Review: Precision medicine and driver mutations: Computational methods, functional assays and conformational principles for interpreting cancer drivers. *PloS Comput. Biol.* 15, 1–54. doi: 10.1371/journal.pcbi.1006658

Pai, S., and Bader, G. D. (2018). Patient Similarity Networks for Precision Medicine. *J. Mol. Biol.* 430, 2924–2938. doi: 10.1016/j.jmb.2018.05.037

Phan, H. T., and Sternberg, M. J. (2012). PINALOG: a novel approach to align protein interaction networks—implications for complex detection and

function prediction. *Bioinformatics* 28, 1239–1245. doi: 10.1093/bioinformatics/bts119

Pourramezan, Z., Kermanshahi, R. K., Oloomi, M., Aliahmadi, A., and Rezadoost, H. (2018). In vitro study of antioxidant and antibacterial activities of Lactobacillus probiotic spp. *Folia Microbiol.* 63, 31–42. doi: 10.1007/s12223-017-0531-x

Rezadoost, H., Karimi, M., and Jafari, M. (2016). Proteomics of hot-wet and cold-dry temperaments proposed in Iranian traditional medicine: a Network-based Study. *Sci. Rep.* 6, 30133–30133. doi: 10.1038/srep30133

Rouillard, A. D., Wang, Z., and Ma'ayan, A. (2015). Abstraction for data integration: Fusing mammalian molecular, cellular and phenotype big datasets for better knowledge extraction. *Comput. Biol. Chem.* 58, 104. doi: 10.1016/j.compbiolchem.2015.06.003

Ruan, C., Wang, Y., Zhang, Y., Ma, J., Chen, H., Aickelin, U., et al. (2017). "THCluster: Herb supplements categorization for precision traditional Chinese medicine," in *2017 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*. (Kansas City, MO, USA), 417–424.

Stefano, G. B., and Kream, R. M. (2015). Personalized- and one- medicine: bioinformatics foundation in health and its economic feasibility. *Med. Sci. Monit.* 21, 201–204. doi: 10.12659/MSM.893207

Steindel, S. J. (2010). International classification of diseases, 10th edition, clinical modification and procedure coding system: descriptive overview of the next generation HIPAA code sets. *J. Am. Med. Inf. Assoc.* 17, 274–282. doi: 10.1136/jamia.2009.001230

Szklarczyk, D., Santos, A., von Mering, C., Jensen, L. J., Bork, P., and Kuhn, M. (2016). STITCH 5: augmenting protein-chemical interaction networks with tissue and affinity data. *Nucleic Acids Res.* 44, D380–D384. doi: 10.1093/nar/gkv1277

Tang, J., and Aittokallio, T. (2014). Network Pharmacology Strategies Toward Multi-Target Anticancer Therapies: From Computational Models to Experimental Design Principles. *Curr. Pharm. Design* 20, 23–36. doi: 10.2174/13816128113199990470

The Finngen Research Project Takes Finns to a Discovery Trip to Genome Data. (Nature Publisher Group). *Nature Med.* 24 (12), 1792–1794.

Vespignani, A. (2003). Evolution thinks modular. *Nat. Genet.* 35, 118–119. doi: 10.1038/ng1003-118

Voineagu, I., Wang, X., Johnston, P., Lowe, J. K., Tian, Y., Horvath, S., et al. (2011). Transcriptomic analysis of autistic brain reveals convergent molecular pathology. *Nature* 474, 380–384. doi: 10.1038/nature10110

Von Luxburg, U. (2006). A tutorial on spectral clustering. *Max Planck Institute Biol. Cybernet. Tech. Rep.*

Wagenseller, P., Wang, F., and Wu, W. (2018). Size Matters: A Comparative Analysis of Community Detection Algorithms. *IEEE Trans. Comput. Soc. Syst.* 5, 951–960. doi: 10.1109/TCSS.2018.2875626

Wang, Y., Jafari, M., Tang, Y., and Tang, J. (2019). Predicting Meridian in Chinese traditional medicine using machine learning approaches. *PloS Comput. Biol.* 15, e1007249–e1007249. doi: 10.1371/journal.pcbi.1007249

Wang, W., Li, C., Xu, J., and Li, X. (2019). "Bridging Fatty Liver Disease and Traditional Chinese Medicine: A Complex Network Approach," in *2019 IEEE International Symposium on Circuits and Systems (ISCAS)*. (IEEE), 1–5.

Wu, Y., Zhang, F., Yang, K., Fang, S., Bu, D., Li, H., et al. (2019). SymMap: an integrative database of traditional Chinese medicine enhanced by symptom mapping. *Nucleic Acids Res.* 47, D1110–D1117. doi: 10.1093/nar/gky1021

Xu, Q., Nwe, T. L., and Guan, C. (2015). Cluster-based analysis for personalized stress evaluation using physiological signals. *IEEE J. BioMed. Health Inform* 19, 275–281. doi: 10.1109/JBHI.2014.2311044

Xu, Z. (2011). Modernization: one step at a time. *Nature* 480, S90–S92. doi: 10.1038/480S90a

Xue, R., Fang, Z., Zhang, M., Yi, Z., Wen, C., and Shi, T. (2013). TCMID: Traditional Chinese Medicine integrative database for herb molecular mechanism analysis. *Nucleic Acids Res.* 41, D1089–D1095. doi: 10.1093/nar/gks1100

Yaffe, M. B. (2019). Why geneticists stole cancer research even though cancer is primarily a signaling disease. *Sci. Signaling* 12, eaaw3483. doi: 10.1126/scisignal.aaw3483

Yang, J., and Leskovec, J. (2015). Defining and evaluating network communities based on ground-truth. *Knowledge Inf. Syst.* 42, 181–213. doi: 10.1007/s10115-013-0693-z

Zhang, N. L., Yuan, S., Chen, T., and Wang, Y. (2008). Latent tree models and diagnosis in traditional Chinese medicine. *Artif. Intell. Med.* 42, 229–245. doi: 10.1016/j.artmed.2007.10.004

Zhang, W., Yue, X., Huang, F., Liu, R., Chen, Y., and Ruan, C. (2018). Predicting drug-disease associations and their therapeutic function based on the drug-disease association bipartite network. *Methods* 145, 51–59. doi: 10.1016/j.ymeth.2018.06.001

Zhao, Y., Zhang, X., Luo, L., He, L., Zhao, Y., Liu, B., et al. (2014). "TCM syndrome differentiation of AIDS using subspace clustering algorithm," in *2014 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*. (IEEE), 219–224.

Zhou, X., Liu, B., and Wu, Z. (2005). "Text mining for clinical Chinese herbal medical knowledge discovery," in *International Conference on Discovery Science*. (Berlin, Heidelberg: Springer), 396–398.

Zhou, X., Li, Y., Peng, Y., Hu, J., Zhang, R., He, L., et al. (2014). Clinical phenotype network: the underlying mechanism for personalized diagnosis and treatment of traditional Chinese medicine. *Front. Med. China* 8, 337–346. doi: 10.1007/s11684-014-0349-8