Received: 14 October 2022 | Revised: 8 March 2023 | Accepted: 30 May 2023

DOI: 10.1002/wsbm.1623

ADVANCED REVIEW

WIREs MECHANISMS OF DISEASE WILEY

Network medicine for patients' stratification: From single-layer to multi-omics

Manuela Petti | **Lorenzo Farina**

Department of Computer, Control and Management Engineering, Sapienza University of Rome, Rome, Italy

Correspondence
Manuela Petti, Department of Computer, Control and Management Engineering, Sapienza University of Rome, Via Ariosto, 25, 00185 Rome, Italy.
Email: manuela.petti@uniroma1.it

Funding information
Sapienza Università di Roma, Grant/Award Number: RM12117A8B3C3033

Edited by: Emily Frieben, Executive Editor

Abstract

Precision medicine research increasingly relies on the integrated analysis of multiple types of omics. In the era of big data, the large availability of different health-related information represents a great, but at the same time untapped, chance with a potentially fundamental role in the prevention, diagnosis and prognosis of diseases. Computational methods are needed to combine this data to create a comprehensive view of a given disease. Network science can model biomedical data in terms of relationships among molecular players of different nature and has been successfully proposed as a new paradigm for studying human diseases. Patient stratification is an open challenge aimed at identifying subtypes with different disease manifestations, severity, and expected survival time. Several stratification approaches based on high-throughput gene expression measurements have been successfully applied. However, few attempts have been proposed to exploit the integration of various genotypic and phenotypic data to discover novel sub-types or improve the detection of known groupings.

This article is categorized under:

Cancer > Biomedical Engineering
Cancer > Computational Models
Cancer > Genetics/Genomics/Epigenetics

KEYWORDS

health-related data, multidimensional, network medicine, patient similarity network, patient stratification, precision medicine

1 | INTRODUCTION

Precision medicine (considered analogous to and sometimes synonymous with personalized medicine or stratified medicine; Erikainen & Chan, 2019) is defined as an emerging approach for disease treatment and prevention that takes into account individual variability in genes, environment, and lifestyle for each person (König et al., 2017). As such, it recognizes the complex nature of diseases as the result of the wide variety of direct and indirect interactions between genes and the environment (Craig, 2008). In the last decades, there has been a growing proliferation of research studies in this

field, which suggest going beyond the traditional classification of diseases based on symptoms (phenotypes), anatomical site, or histology. Indeed, precision medicine aims to characterize diseases at a higher resolution using layered data extracted at different scales (from molecules to system level).

Recent innovations and developments in molecular biology and biotechnology have made it possible to acquire and store large health-related datasets. However, this huge amount of information is not knowledge: rather, it represents a valuable potential that requires computational resources and capabilities to be exploited. The main aim of computational efforts in this field is to propose new methods for data analysis and integration to identify biomarkers that open the possibility of shifting from the traditional therapeutic approach of the *average model patient* to the modern approach of patient-specific and data-driven therapies.

Recently, network science has been established as one of the most successful computational approaches in dealing with the precision medicine challenge, so it can be considered an effective common "language" between physicians and data analysts to implement network medicine in a clinical setting (Farina, 2021). Examples are: the concept of hubs ("essential genes tend to be associated with hubs, or highly connected proteins") and the concept of network module ("disease genes have a high propensity to interact with each other, forming disease modules") (Barabási et al., 2011). These assumptions, as well as other network measures, have been applied with great success in many different tasks, such as: disease gene prediction (Ghiassian et al., 2015; Petti et al., 2020, 2021; Wang & Loscalzo, 2018; Yin et al., 2017), drug repurposing (Cheng et al., 2018, 2019; Paci et al., 2022; Zhou et al., 2020), gene expression analysis (Langfelder & Horvath, 2008; Paci et al., 2017), protein–protein interaction prediction (Colonnese et al., 2021).

In this review, we focus on network-based computational approaches for patients' stratification, which is the key challenge of precision medicine. We discuss its potential and related open problems, survey the analytical methods based on network analysis for integrating health-related data and consider knowledge gaps and future directions for this developing field. Finally, we will review successful applications of patient similarity networks.

2 | PATIENTS' STRATIFICATION: COMPUTATIONAL APPROACHES AND OPEN CHALLENGES

Patient stratification is an important goal of biomedical research and one of the main challenges aimed at developing personalized treatments. Despite being based on specific patient selection criteria, clinical trials often reveal unexpected heterogeneity in response and survival rates. This is related to the well-established hypothesis that cohorts of patients with the same disease contain subgroups with similar clinical features and/or similar omics profiles which determine the occurrence of similar within-group clinical outcomes. For example, complex common diseases such as cancer are *heterogeneous diseases* with diverse pathogeneses and clinical features. In particular, to date, many cancers are known to contain subtypes associated with different outcomes and therapeutic treatments (Russnes et al., 2017; West et al., 2012).

Patient stratification allows the identification of subgroups by their genetic, genomic, physiological, and clinical profiles. In the past, risk stratification assays have focused mainly on response prediction to existing treatment regimens. Still, today there is awareness about the need for robust subtypes in epidemiologic and functional studies to learn more about mechanisms in tumor development and evolution during treatment with a focus on response and resistance (Russnes et al., 2017). Indeed, more recently, molecular stratification (based on genetics and omics datasets) allowed to build gene signatures and to model group-specific underlying pathological mechanisms. However, most risk models are based on one or few features, with limited use of integrated clinical and omics data. New evidence suggests that multiple biomarkers must be considered at the same time (Alfano et al., 2023). Most computational approaches based on machine learning recommend selecting the most accurate prognostic factors and their combination (Ali & Aittokallio, 2019; Jalali-najafabadi et al., 2021). Instead, in this review, we highlight the need to address the multidimensionality of health-related data and the potential of multiple and integrated biomarkers exploiting network and systemic approaches, as described in detail in the following sections.

3 | PATIENT SIMILARITY NETWORK PARADIGM

Patients with similar clinical signatures and/or similar omics profiles are expected to show similar clinical outcomes. Pairwise similarities between patients have a natural representation as networks. Indeed, a network is commonly

defined as a series of interconnected components, systems or entities. A graph $G = (V, E)$ is the abstract representation of a network: it consists of the set V of vertices (or nodes) linked by means of edges (set E), indicating the presence of some relationship (similarity, dependency) between the vertices. Patient Similarity Network (PSN) is a recent paradigm aimed at patient stratification where nodes represent patients and edges represent the similarity between patients calculated using their clinical and/or biomolecular features (Figure 1).

Each patient data feature (e.g., age, sex, mutation status, gene expression profile) can be used to obtain a network of pairwise patient similarities. Several distance/similarity measures can be adopted to measure patient similarity, and in Table 1 we reported the most used and their application in patient similarity frameworks.

Once the PSN is built, the patient subtypes can be identified investigating the community structure of the obtained network. In network science, a community is defined as a subset of nodes that are densely connected to each other and loosely connected to the nodes in different communities. In the case of weighted networks, a community is instead a group of nodes strongly connected to each other with weak connections with nodes of other communities. In PSN, communities are thus densely/strongly connected groups of nodes and could reveal patients' subtypes. Among the methods proposed in literature (Fortunato, 2010), the most used are the algorithms based on modularity (Newman, 2006), spectral algorithms (von Luxburg, 2007) and information theory algorithms such us Markov Clustering Algorithm.

The PSN framework is characterized by important advantages such as generalizability, accuracy, interpretability, ability to integrate heterogeneous data, and the natural management of missing values.

In the next section we show network-based methods for health-related data integration describing approaches already used for this aim, but also other approaches potentially exploitable.

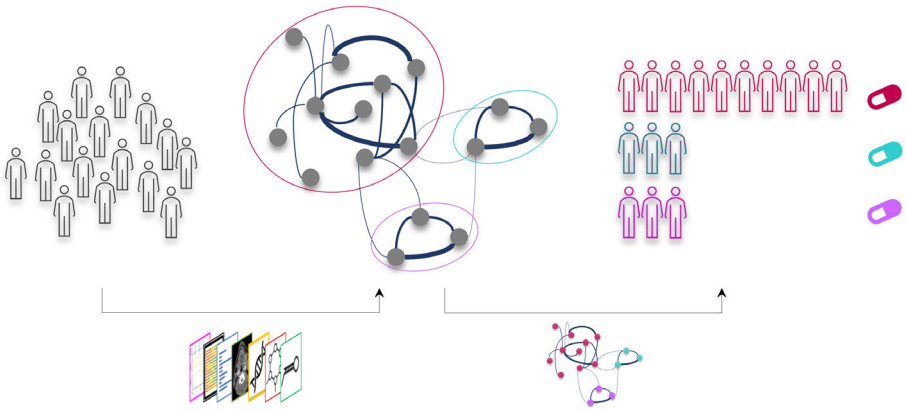

FIGURE 1 Overview of patient stratification based on health-related data integration and Patient Similarity Network paradigm.

TABLE 1 List of common distance and similarity measures used in patient similarity frameworks.

Distance/ similarity metrics	Measures	Uses in patient similarity framework
Euclidean distance	Distance in cartesian space	(Pai et al., 2019; Vitali et al., 2018; Wang et al., 2014; Wang & Sun, 2015)
Hamming distance	Distance between equal length string	(Lee et al., 2018)
City Block distance	Absolute difference between coordinates of a pair of vector	(Chen et al., 2018)
Minkowsky distance	Generalization of both the Euclidean distance and the Manhattan distance	(Labellapansa et al., 2016)
Chebychev distance	Greatest differences between two vector along any coordinate dimension	(Li et al., 2020)
Pearson correlation	Linear dependence	(Gliozzo et al., 2020)
Cosine similarity	Similarity in vector space	(Lee et al., 2015; Tashkandi et al., 2018)
Jaccard similarity	Similarity between sets	(Shu et al., 2019)

4 | NETWORK-BASED METHODS FOR DATA INTEGRATION

In this section, we focus on the definitions that can be exploited to address data multidimensionality and data integration.

4.1 | Multipartite networks

A bipartite graph $G = (S, T, E)$ is a network in which nodes can be divided into two disjoint sets of nodes S and T, so that each edge in E connects a node of S with a node of T. From a bipartite network, it is possible to derive 2 projected networks, where each one is composed of only one set of nodes. The projections in the two sets of vertices are obtained by deducing relationships between nodes of the same type. In other words, in order to study the relationships among a particular set of nodes, the projected networks contain links between 2 nodes of S if both are connected to the same node of T. A bipartite graph represents a special case of multipartite graphs or k-partite graphs with $k = 2$.

In Figure 2, we show an example of a phenotype–patient bipartite network composed of the set of patients and the set of phenotypes: the links inform about the phenotypes manifested by the patients. The projection in the set of patients returns a patient similarity network in which pairs of patients are linked if they manifest at least one common phenotype (Figure 2, panel a). From a patient cohort, multiple information can be represented and analyzed using multipartite networks. For example, a phenotype–patient–gene tripartite network can be constructed (Figure 2, panel b): in this case, we can obtain an additional patient similarity network based on the similarity between patients in sharing at least one mutation. Moreover, the multipartite networks built with real-world cohort data can be expanded by integrating publicly available sources. In Figure 2c we show an example of integration with well-known databases: protein–protein interactions (e.g., BioGRID;Oughtred et al., 2021), drug-target interactions (e.g., Drugbank; Wishart et al., 2018), miRNA-target interactions (e.g miRTarBase; Huang et al., 2022), lncRNA–miRNA interactions (e.g., LncCeRBase; Pian et al., 2018), lncRNA-disease associations (e.g., LncRNADisease; Bao et al., 2019) and drug-phenotype associations (e.g., Drugbank).

A multipartite graph also allows to consider indirect associations between nonadjacent layers of nodes: for example, the direct links between the two pairs patient-gene and gene-drug, can be substituted with the indirect and gene-mediated association patient-drug (see Figure 2, panel c).

Multipartite graphs have proven to be a successful tool in representing, integrating and analyzing molecular and, more in general, health-related data (Pavlopoulos et al., 2018; Petti et al., 2023; Wu et al., 2021). It is worth noting that also a temporal bipartite network modeling approach has been proposed to study multimorbidity trajectories in patients with severe mental illnesses (Wang et al., 2022). However, their potential in the PSN framework has not yet been explored also as a consequence of the need for specialized methods and software for studying such networks (Pavlopoulos et al., 2018).

4.2 | Multi-layer networks

A multi-layer network is described by a set of nodes V, a set of edges E, and a set of layers L. Multilayer networks provide better modeling for biological networks and, more in general, for complex networks. They can be classified in two types: "node-colored graphs" (heterogeneous nodes are distributed in the layers) and edge-colored graphs (heterogeneous links in different layers, that is, each layer is related to a specific type of connection) (Hammoud & Kramer, 2020). In particular, the latter are node-aligned graphs, that is, the same node belongs to different layers simultaneously: such graphs are very suitable for studying different aspects of the same system and thus in modeling patient similarity extracted from multiple features (see Figure 3, panel b). The multi-layer PSN can be obtained by building unimodal PSNs on each data source or data type and then staking or merging the layers (Figure 3).

The analysis of stacked PSNs is possible thanks to some graph measures that have already been extended in the multi-layer space. Among them, there is modularity optimization (Mucha et al., 2010): the authors extended the modularity function for community detection by adapting it to layered network. The key point is to include a meso-structure (called *interslice couplings*) that connects a node (i.e., a patient) of a specific slice (i.e., PSN) to its copy in another slice. This coupling parameter determines the stability of the network partitioning (i.e., patients' subtypes) across the slices

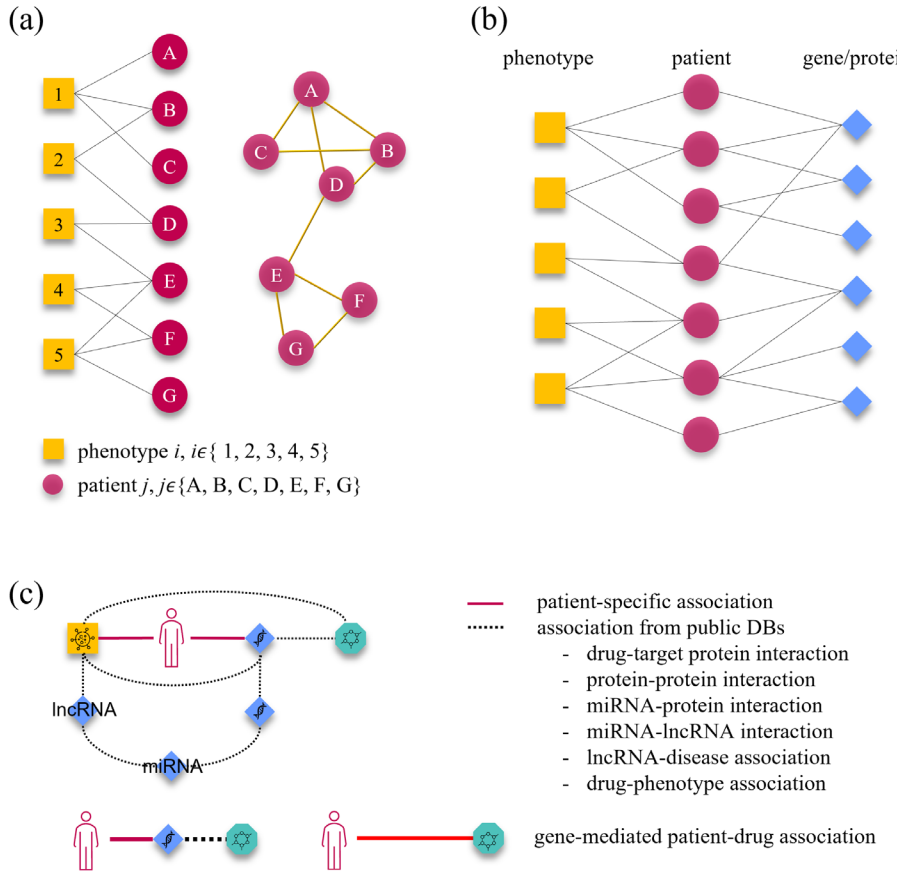

FIGURE 2 Overview of the use of multipartite networks in representing, integrating and analyzing patient data and publicly open sources. (a) Phenotype–patient bipartite network with associated Patient Similarity Network (PSN) obtained with the projection in the set of patient nodes. (b) Phenotype–patient–gene tripartite network. (c) Example of integration between patient data and public databases.

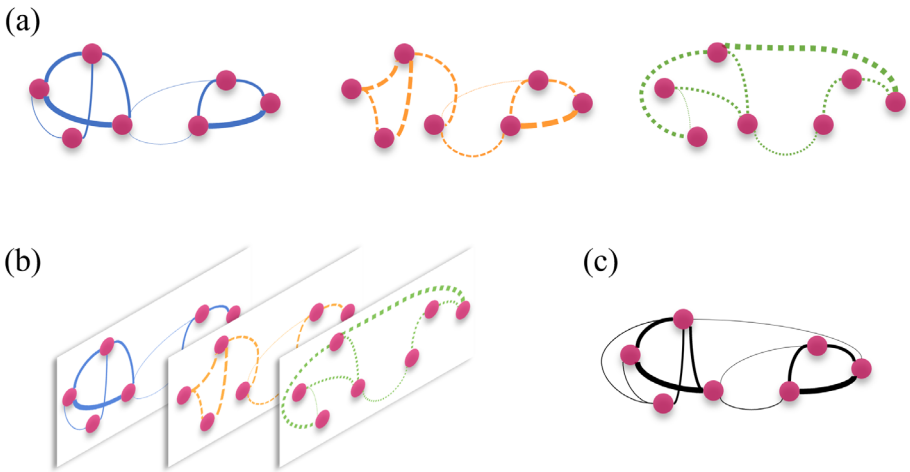

FIGURE 3 Overview of the multi-layer network construction and layers integration. (a) unimodal PSNs where nodes represent the patients, edge color codes for the type of data used to evaluate the patient similarity and edge thickness codes for the similarity strength. (b) Multi-layer PSN obtained by stacking the unimodal PSNs. (c) Result of networks fusion.

(i.e., clinical and/or molecular features). Network alignment is an alternative approach to extract similar patterns from multi-layer networks. This procedure aims to find a good node mapping between two or more layers that identifies common interaction patterns between the compared networks. It allows the extraction of "similar" subnetworks in two

or several input networks. Typical applications are associated with different organisms to uncover complex mechanisms at the basis of evolutionary conservations or to infer the biological meanings of groups of interacting cellular components belonging to organisms not yet well characterized (Faisal et al., 2015; Guzzi & Milenković, 2018). The application of the global network alignment to multi-layer PSNs may reveal patients subtypes also in the case of "node-colored graphs," that is, when the layers (unimodal PSN) are composed of different or not completely overlapping sets of nodes (patients' cohorts).

However, there is not a large availability of procedures allowing to exploit the multidimensionality of multi-layer PSNs. Rather, recently developed methods are based on layers integration by stacking or merging the edges. An overview of these approaches is provided below.

4.2.1 | Heterogeneous network

Heterogeneous networks consider *k* different kinds of nodes, each type corresponding to a separate layer of biological information. In this framework, intra-layer and inter-layer connections are formally treated in the same way, even if they can be weighed differently. The multi-layered information is therefore squeezed into just one dimension. An example of the application of heterogeneous networks for modeling gene-phenotype networks was presented by Li and Patra (2010).

4.2.2 | Network fusion

Network fusion can be performed creating a multigraph or merging the parallel edges in a single edge. Multigraphs are graphs with multiple edges (associated with different connectivity/similarity measures) connecting the same pair of nodes: in a multigraph, two or more edges are incident to the same two vertices. Such edges are also called parallel edges or multi-edges. They have been used for the modeling of metabolic pathways where the same substances can be transformed by different reactions, but also with the aim to integrate different kinds of relationships obtained from different data sources (Li & Li, 2012). Instead, edge fusion has been developed to process a set of unimodal PSNs producing an integrated PSN. The Similarity Network Fusion method was proposed by Wang and colleagues (Wang et al., 2014): it can combine many different types of measurements (such as mRNA expression data, DNA methylation, miRNA expression and more—clinical data, questionnaires, image data, etc.) for a given set of samples. Briefly, SNF consists of two main steps: constructing a sample-similarity network for each data type and integrating these networks into a single similarity network using a nonlinear combination method (more details are provided in Section 5).

5 | SUCCESSFUL PSN-BASED METHODS FOR PATIENT STRATIFICATION AND THEIR APPLICATIONS

5.1 | Network-based stratification

The Network-based stratification (NBS) method was proposed to integrate somatic tumor genomes with gene networks (Hofree et al., 2013). This is one of the first approaches developed to focus on the orthogonal problem of using network knowledge to stratify a cohort into meaningful subsets. The method requires that the somatic mutations for each patient are represented as a profile of binary states on genes, in which state *1* indicates a gene for which mutation has occurred in the tumor relative to germline. It consists of the following steps.

- For each patient, projection of the mutation profiles onto a human gene interaction network.
- Application of the technique of network propagation (Vanunu et al., 2010) to spread the influence of each mutation profile over its network neighborhood. This step provides a "network-smoothed" profile in which the state of each gene is no longer binary but reflects its network proximity to the mutated genes in that patient along a continuous range [0, 1].

- The patient profiles are clustered into a predefined number of subtypes ($k = 2, 3, ...12$) using an unsupervised technique based on nonnegative matrix factorization.
- Application of consensus clustering (Monti et al., 2003), in which the above procedure is repeated for 1000 different random subsamples.
- Aggregation of the 1000 runs into a (patient × patient) co-occurrence matrix, which summarizes the frequency with which each pair of patients has co-segregated into the same cluster.
- This co-occurrence matrix is then clustered a second time to recover a final stratification of the patients into clusters/subtypes.

The authors of NBS applied the proposed approach to stratify the somatic mutation profiles of three major cancers cataloged in TCGA (ovarian, uterine and lung adenocarcinoma), proving that NBS identifies patient subtypes that are predictive of clinical outcomes such as patient survival, response to therapy or tumor histology. In another study, NBS was applied to evaluate the effectiveness of various gene panels in classifying tumors into clinically meaningful subtypes (Zhong et al., 2015). NBS was applied to 13 major cancer types with exome-level mutation data and compared the classification based on the full exome data with those focusing only on small sets of genes. The results showed that small panels effectively cluster tumors and furthermore, often outperform full exome data for most cancer types.

Despite the success of the NBS applications in detecting patient subgroups, some studies investigated its limitation related to the use of a nonspecific network (He et al., 2017). In fact, with this approach, the mutations are mapped onto a network that does not describe the pattern of gene interactions in the tumor under investigation. To overcome this limitation, in (He et al., 2017), the authors propose to use a cancer-type-specific co-expression network instead of a human gene interaction network. This modification represents an additional level of data integration: the updated version of NBS has been shown to outperform the original method in identifying cancer subtypes significantly associated with clinical outcomes in most cancer types studied (He et al., 2017).

5.2 | Similarity network fusion (SNF)

The similarity network fusion (SNF) method has been developed by Wang and colleagues (Wang et al., 2014) to integrate a set of unimodal PSNs. The NBS limitation discussed above is completely overcome in SNF as it exploits cancer-specific networks obtained from data such as gene expression and DNA methylation. SNF first constructs a sample similarity network for each data type (similarity based on Euclidean distance followed by exponential scaling). This set of networks is then fused by iteratively increasing the weights of concordant edges among different layers, and decreasing the weights of those only present in some but not all layers. Spectral clustering is then applied to reveal the network community structure. Integrating data in a nonlinear fashion allows SNF to take advantage of the common as well as complementary information in different data types. Furthermore, it does not require a priori feature selection and was shown to outperform methods based on single data types as well as other multi-omics approaches.

SFN has been applied with success in many different applications. For example in (Bhalla et al., 2021), the author investigate the heterogeneity of multiple myeloma using Whole-Exome Seq (WES), Whole-Genome Seq (WGS) and RNA-Seq data from 655 tumor samples from Multiple Myeloma (MM). In this application, unimodal PSNs are based on 5 data types: (1) gene expression and (2) gene fusion data from RNA-Seq; (3) somatic single nucleotide variations (SNVs) from WES, (4) Copy Number Alterations (CNAs; focal and broad) and (5) translocations from WGS. MM-SNF identified patient subgroups not previously described defined by specific patterns of alterations, enriched for specific gene vulnerabilities, and associated with potential therapeutic options (Bhalla et al., 2021).

In another work (Cavalli et al., 2017), SFN was applied to genome-wide DNA methylation and gene expression data across 763 primary samples identifies very homogeneous clusters of patients, supporting the presence of medulloblastoma subtypes. The authors identified 12 different subtypes of medulloblastoma, some of which not easily distinguishable through the analysis of individual types of data.

Even in the case of this method, several studies proposed some changes to improve and expand SNF. The main aspects that these studies have successfully addressed are: weighted version of similarity network fusion to give

different importance at the different layers (Mac Aogáin et al., 2021; Xu et al., 2016); extension for data types containing discrete values (e.g., mutation profile) (Yang et al., 2017) and similarity network construction to obtain robust dense similarity matrices (Liu et al., 2021).

5.3 | netDx: Patient classification by similarity networks

Lastly, we introduce netDx, an approach combining a supervised machine learning method for patient classification and the PSN paradigm (Pai et al., 2019). The goal of netDx is to identify the input features predictive of high and low risk, and to accurately assign new patients to the correct class. The workflow requires defining a classification problem and a cohort of patients containing positive and negative examples (i.e., cases and controls). Each available data feature is converted into a PSN using a similarity measure defined on the basis of data type. netDx uses Lasso regression within the cross-validation loop to prefilter only variables carrying predictive information. Starting from the unimodal PSNs, the integrated PSN is obtained by averaging all edge weights between patients from all selected networks (features). netDx has been applied to classify patients as belonging to one of two ependymoma subtypes identified by Witt et al. (Witt et al., 2011). The two subgroups differ in terms of demographic, clinical, and molecular (gene expression) profiles. Cross-validation showed that netDx obtained an accuracy of 81% (Pai & Bader, 2018).

As for any supervised machine-learning method, netDx requires a large bank of patient samples to learn from. This limitation will gradually be mitigated as rapidly new technologies enable the acquisition and storage of data from ever-increasing patient cohorts. However, in the case of small populations other approaches are needed.

The above-described methods (with related updated versions) are the most used in cancer research. Their applications have demonstrated their potential in patient stratification and data integration providing relevant results. The common feature of these approaches is that they merge the layers obtaining a unique network. PSN-based methods able to explore the multidimensional space of health-related data are still missing. Some network tools for multi-layer network analysis are already available, but their use in patient stratification tasks has not yet been explored. Moreover, further studies are needed to identify guidelines in their application depending on the disease of interest, type of data and sample size. The application and comparison of different approaches could benefit from harmonizing different PSN techniques. Lastly, it is worth noting that to date, these approaches are only reserved for the research field. The main limitation to their use in clinical practice is to define procedures for secure sharing of private data.

6 | CONCLUSION

In this review, we aimed to provide an overview of the network-based methods for patients' stratification and biomedical data integration. In particular, we focused on the Patient Similarity Network paradigm, a recently developed approach where patients are represented as nodes connected by edges coding for the pairwise similarity calculated from clinical or omics data. This paradigm successfully addresses both challenges (patients' stratification and biomedical data integration), as proven by the recent applications in different diseases. We described approaches already used to represent and analyze patient similarity, but we highlighted the need for new algorithms and network measures to handle the multidimensionality of available health-related data. We believe that leveraging network and systemic approaches can bring about the expected shift toward precision medicine.

AUTHOR CONTRIBUTIONS

Manuela Petti: Conceptualization (lead); data curation (lead); funding acquisition (lead); visualization (lead); writing – original draft (lead); writing – review and editing (equal). **Lorenzo Farina:** Conceptualization (supporting); supervision (lead); writing – review and editing (equal).

FUNDING INFORMATION

This work was supported by Sapienza University of Rome, "Progetto di Ateneo," grant number: RM12117A8B3C3033.

CONFLICT OF INTEREST STATEMENT

The authors have declared no conflicts of interest for this article.

DATA AVAILABILITY STATEMENT

Data sharing is not applicable to this article as no new data were created or analyzed in this study.

ORCID

Manuela Petti https://orcid.org/0000-0002-0149-161X
Lorenzo Farina https://orcid.org/0000-0001-8314-6029

RELATED WIREs ARTICLES

Promise of personalized omics to precision medicine
Molecular networks in Network Medicine: Development and applications
Evolution of precision oncology-guided treatment paradigms

REFERENCES

Alfano, C., Farina, L., & Petti, M. (2023). Networks as biomarkers: Uses and purposes. *Genes*, *14*, 429. https://doi.org/10.3390/genes14020429

Ali, M., & Aittokallio, T. (2019). Machine learning and feature selection for drug response prediction in precision oncology applications. *Biophysical Reviews*, *11*, 31–39. https://doi.org/10.1007/s12551-018-0446-z

Bao, Z., Yang, Z., Huang, Z., Zhou, Y., Cui, Q., & Dong, D. (2019). LncRNADisease 2.0: An updated database of long non-coding RNA-associated diseases. *Nucleic Acids Research*, *47*, D1034–D1037. https://doi.org/10.1093/nar/gky905

Barabási, A.-L., Gulbahce, N., & Loscalzo, J. (2011). Network medicine: A network-based approach to human disease. *Nature Reviews. Genetics*, *12*, 56–68. https://doi.org/10.1038/nrg2918

Bhalla, S., Melnekoff, D. T., Aleman, A., Leshchenko, V., Restrepo, P., Keats, J., Onel, K., Sawyer, J. R., Madduri, D., Richter, J., Richard, S., Chari, A., Cho, H. J., Dudley, J. T., Jagannath, S., Laganà, A., & Parekh, S. (2021). Patient similarity network of newly diagnosed multiple myeloma identifies patient subgroups with distinct genetic features and clinical implications. *Science Advances*, *7*, eabg9551. https://doi.org/10.1126/sciadv.abg9551

Cavalli, F. M. G., Remke, M., Rampasek, L., Peacock, J., Shih, D. J. H., Luu, B., Garzia, L., Torchia, J., Nor, C., Morrissy, A. S., Agnihotri, S., Thompson, Y. Y., Kuzan-Fischer, C. M., Farooq, H., Isaev, K., Daniels, C., Cho, B.-K., Kim, S.-K., Wang, K.-C., … Taylor, M. D. (2017). Intertumoral heterogeneity within medulloblastoma subgroups. *Cancer Cell*, *31*, 737–754.e6. https://doi.org/10.1016/j.ccell.2017.05.005

Chen, J., Sun, L., Guo, C., Wei, W., & Xie, Y. (2018). A data-driven framework of typical treatment process extraction and evaluation. *Journal of Biomedical Informatics*, *83*, 178–195. https://doi.org/10.1016/j.jbi.2018.06.004

Cheng, F., Desai, R. J., Handy, D. E., Wang, R., Schneeweiss, S., Barabási, A.-L., & Loscalzo, J. (2018). Network-based approach to prediction and population-based validation of in silico drug repurposing. *Nature Communications*, *9*, 2691. https://doi.org/10.1038/s41467-018-05116-5

Cheng, F., Kovács, I. A., & Barabási, A.-L. (2019). Network-based prediction of drug combinations. *Nature Communications*, *10*, 1197. https://doi.org/10.1038/s41467-019-09186-x

Colonnese, S., Petti, M., Farina, L., Scarano, G., & Cuomo, F. (2021). Protein-protein interaction prediction via graph signal processing. *IEEE Access*, *9*, 142681–142692. https://doi.org/10.1109/ACCESS.2021.3119569

Craig, J. (2008) Complex diseases: Research and applications. *Nature Education*, *1*, 184.

Erikainen, S., & Chan, S. (2019). Contested futures: Envisioning "Personalized," "Stratified," and "Precision" medicine. *New Genetics and Society*, *38*, 308–330. https://doi.org/10.1080/14636778.2019.1637720

Faisal, F. E., Meng, L., Crawford, J., & Milenković, T. (2015). The post-genomic era of biological network alignment. *EURASIP Journal on Bioinformatics and Systems Biology*, *2015*, 3. https://doi.org/10.1186/s13637-015-0022-9

Farina, L. (2021). Network as a language for precision medicine. *Annali dell'Istituto Superiore di Sanità*, *57*, 330–342. https://doi.org/10.4415/ANN_21_04_08

Fortunato, S. (2010). Community detection in graphs. *Physics Reports*, *486*, 75–174. https://doi.org/10.1016/j.physrep.2009.11.002

Ghiassian, S. D., Menche, J., & Barabási, A.-L. (2015). A DIseAse MOdule detection (DIAMOnD) algorithm derived from a systematic analysis of connectivity patterns of disease proteins in the human Interactome. *PLoS Computational Biology*, *11*, e1004120. https://doi.org/10.1371/journal.pcbi.1004120

Gliozzo, J., Perlasca, P., Mesiti, M., Casiraghi, E., Vallacchi, V., Vergani, E., Frasca, M., Grossi, G., Petrini, A., Re, M., Paccanaro, A., & Valentini, G. (2020). Network modeling of patients' biomolecular profiles for clinical phenotype/outcome prediction. *Scientific Reports*, *10*, 1–15. https://doi.org/10.1038/s41598-020-60235-8

Guzzi, P. H., & Milenković, T. (2018). Survey of local and global biological network alignment: The need to reconcile the two sides of the same coin. *Briefings in Bioinformatics*, *19*, 472–481. https://doi.org/10.1093/bib/bbw132

Hammoud, Z., & Kramer, F. (2020). Multilayer networks: Aspects, implementations, and application in biomedicine. *Big Data Analytics*, *5*, 2. https://doi.org/10.1186/s41044-020-00046-0

He, Z., Zhang, J., Yuan, X., Liu, Z., Liu, B., Tuo, S., & Liu, Y. (2017). Network based stratification of major cancers by integrating somatic mutation and gene expression data. *PLoS One*, *12*, e0177662. https://doi.org/10.1371/journal.pone.0177662

Hofree, M., Shen, J. P., Carter, H., Gross, A., & Ideker, T. (2013). Network-based stratification of tumor mutations. *Nature Methods*, *10*, 1108–1115. https://doi.org/10.1038/nmeth.2651

Huang, H.-Y., Lin, Y.-C.-D., Cui, S., Huang, Y., Tang, Y., Xu, J., Bao, J., Li, Y., Wen, J., Zuo, H., Wang, W., Li, J., Ni, J., Ruan, Y., Li, L., Chen, Y., Xie, Y., Zhu, Z., Cai, X., … Huang, H.-D. (2022). miRTarBase update 2022: An informative resource for experimentally validated miRNA–target interactions. *Nucleic Acids Research*, *50*, D222–D230. https://doi.org/10.1093/nar/gkab1079

Jalali-najafabadi, F., Stadler, M., Dand, N., Jadon, D., Soomro, M., Ho, P., Marzo-Ortega, H., Helliwell, P., Korendowych, E., Simpson, M. A., Packham, J., Smith, C. H., Barker, J. N., McHugh, N., Warren, R. B., Barton, A., & Bowes, J. (2021). Application of information theoretic feature selection and machine learning methods for the development of genetic risk prediction models. *Scientific Reports*, *11*, 1–14. https://doi.org/10.1038/s41598-021-00854-x

König, I. R., Fuchs, O., Hansen, G., von Mutius, E., & Kopp, M. V. (2017). What is precision medicine? *The European Respiratory Journal*, *50*, 1700391. https://doi.org/10.1183/13993003.00391-2017

Labellapansa, A., Efendi, A., Yulianti, A., & Kadir, E. A. (2016). Lambda value analysis on weighted Minkowski distance model in CBR of schizophrenia type diagnosis, in: 2016 4th international conference on information and communication technology (ICoICT). In *Presented at the 2016 4th international conference on information and communication technology (ICoICT)* (pp. 1–4). IEEE. https://doi.org/10.1109/ICoICT.2016.7571898

Langfelder, P., & Horvath, S. (2008). WGCNA: An R package for weighted correlation network analysis. *BMC Bioinformatics*, *9*, 559. https://doi.org/10.1186/1471-2105-9-559

Lee, J., Maslove, D. M., & Dubin, J. A. (2015). Personalized mortality prediction driven by electronic medical data and a patient similarity metric. *PLoS One*, *10*, e0127428. https://doi.org/10.1371/journal.pone.0127428

Lee, J., Sun, J., Wang, F., Wang, S., Jun, C.-H., & Jiang, X. (2018). Privacy-preserving patient similarity learning in a federated environment: Development and analysis. *JMIR Medical Informatics*, *6*, e7744. https://doi.org/10.2196/medinform.7744

Li, S., Jiang, L., Tang, J., Gao, N., & Guo, F. (2020). Kernel fusion method for detecting cancer subtypes via selecting relevant expression data. *Frontiers in Genetics*, *11*, 570679. https://doi.org/10.3389/fgene.2020.00979

Li, Y., & Li, J. (2012). Disease gene identification by random walk on multigraphs merging heterogeneous genomic and phenotype data. *BMC Genomics*, *13*(suppl 7), S27. https://doi.org/10.1186/1471-2164-13-S7-S27

Li, Y., & Patra, J. C. (2010). Genome-wide inferring gene–phenotype relationship by walking on the heterogeneous network. *Bioinformatics*, *26*, 1219–1224. https://doi.org/10.1093/bioinformatics/btq108

Liu, J., Liu, W., Cheng, Y., Ge, S., & Wang, X. (2021). Similarity network fusion based on random walk and relative entropy for cancer subtype prediction of multigenomic data [WWW document]. *Scientific Programming*, *2021*, 2292703. https://doi.org/10.1155/2021/2292703

Mac Aogáin, M., Narayana, J. K., Tiew, P. Y., Ali, N. A. B. M., Yong, V. F. L., Jaggi, T. K., Lim, A. Y. H., Keir, H. R., Dicker, A. J., Thng, K. X., Tsang, A., Ivan, F. X., Poh, M. E., Oriano, M., Aliberti, S., Blasi, F., Low, T. B., Ong, T. H., Oliver, B., … Chotirmall, S. H. (2021). Integrative microbiomics in bronchiectasis exacerbations. *Nature Medicine*, *27*, 688–699. https://doi.org/10.1038/s41591-021-01289-7

Monti, S., Tamayo, P., Mesirov, J., & Golub, T. (2003). Consensus clustering: A resampling-based method for class discovery and visualization of gene expression microarray data. *Machine Learning*, *52*, 91–118. https://doi.org/10.1023/A:1023949509487

Mucha, P. J., Richardson, T., Macon, K., Porter, M. A., & Onnela, J.-P. (2010). Community structure in time-dependent, multiscale, and multiplex networks. *Science*, *328*, 876–878. https://doi.org/10.1126/science.1184819

Newman, M. E. J. (2006). Modularity and community structure in networks. *Proceedings of the National Academy of Sciences*, *103*, 8577–8582. https://doi.org/10.1073/pnas.0601602103

Oughtred, R., Rust, J., Chang, C., Breitkreutz, B.-J., Stark, C., Willems, A., Boucher, L., Leung, G., Kolas, N., Zhang, F., Dolma, S., Coulombe-Huntington, J., Chatr-Aryamontri, A., Dolinski, K., & Tyers, M. (2021). The BioGRID database: A comprehensive biomedical resource of curated protein, genetic, and chemical interactions. *Protein Science*, *30*, 187–200. https://doi.org/10.1002/pro.3978

Paci, P., Colombo, T., Fiscon, G., Gurtner, A., Pavesi, G., & Farina, L. (2017). SWIM: A computational tool to unveiling crucial nodes in complex biological networks. *Scientific Reports*, *7*, 1–16. https://doi.org/10.1038/srep44797

Paci, P., Fiscon, G., Conte, F., Wang, R.-S., Handy, D. E., Farina, L., & Loscalzo, J. (2022). Comprehensive network medicine-based drug repositioning via integration of therapeutic efficacy and side effects. *npj Systems Biology and Applications*, *8*, 1–12. https://doi.org/10.1038/s41540-022-00221-0

Pai, S., & Bader, G. D. (2018). Patient similarity networks for precision medicine. *Journal of Molecular Biology*, *430*, 2924–2938. https://doi.org/10.1016/j.jmb.2018.05.037

Pai, S., Hui, S., Isserlin, R., Shah, M. A., Kaka, H., & Bader, G. D. (2019). netDx: Interpretable patient classification using integrated patient similarity networks. *Molecular Systems Biology*, *15*, e8497. https://doi.org/10.15252/msb.20188497

Pavlopoulos, G. A., Kontou, P. I., Pavlopoulou, A., Bouyioukos, C., Markou, E., & Bagos, P. G. (2018). Bipartite graphs in systems biology and medicine: A survey of methods and applications. *GigaScience*, *7*, giy014. https://doi.org/10.1093/gigascience/giy014

Petti, M., Alfano, C., & Farina, L. (2023). Molecular network analysis of hormonal contraceptives side effects via database integration. *Informatics in Medicine Unlocked*, *36*, 101163. https://doi.org/10.1016/j.imu.2023.101163

Petti, M., Bizzarri, D., Verrienti, A., Falcone, R., & Farina, L. (2020). Connectivity significance for disease gene prioritization in an expanding universe. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, *17*, 2155–2161. https://doi.org/10.1109/TCBB.2019.2938512

Petti, M., Farina, L., Francone, F., Lucidi, S., Macali, A., Palagi, L., & De Santis, M. (2021). MOSES: A new approach to integrate Interactome topology and functional features for disease gene prediction. *Genes*, *12*, 1713. https://doi.org/10.3390/genes12111713

Pian, C., Zhang, G., Tu, T., Ma, X., & Li, F. (2018). LncCeRBase: A database of experimentally validated human competing endogenous long non-coding RNAs. *Database*, *2018*, bay061. https://doi.org/10.1093/database/bay061

Russnes, H. G., Lingjærde, O. C., Børresen-Dale, A.-L., & Caldas, C. (2017). Breast cancer molecular stratification: From intrinsic subtypes to integrative clusters. *The American Journal of Pathology*, *187*, 2152–2162. https://doi.org/10.1016/j.ajpath.2017.04.022

Shu, Z., Liu, W., Wu, H., Xiao, M., Wu, D., Cao, T., Ren, M., Tao, J., Zhang, C., He, T., Li, X., Zhang, R., & Zhou, X. (2019). Symptom-based network classification identifies distinct clinical subgroups of liver diseases with common molecular pathways. *Computer Methods and Programs in Biomedicine*, *174*, 41–50. https://doi.org/10.1016/j.cmpb.2018.02.014

Tashkandi, A., Wiese, I., & Wiese, L. (2018). Efficient In-database patient similarity analysis for personalized medical decision support systems. big data res. *Big Medical/Healthcare Data Analytics*, *13*, 52–64. https://doi.org/10.1016/j.bdr.2018.05.001

Vanunu, O., Magger, O., Ruppin, E., Shlomi, T., & Sharan, R. (2010). Associating genes and protein complexes with disease via network propagation. *PLoS Computational Biology*, *6*, e1000641. https://doi.org/10.1371/journal.pcbi.1000641

Vitali, F., Marini, S., Pala, D., Demartini, A., Montoli, S., Zambelli, A., & Bellazzi, R. (2018). Patient similarity by joint matrix trifactorization to identify subgroups in acute myeloid leukemia. *JAMIA Open*, *1*, 75–86. https://doi.org/10.1093/jamiaopen/ooy008

von Luxburg, U. (2007). A tutorial on spectral clustering. *Statistics and Computing*, *17*, 395–416. https://doi.org/10.1007/s11222-007-9033-z

Wang, B., Mezlini, A. M., Demir, F., Fiume, M., Tu, Z., Brudno, M., Haibe-Kains, B., & Goldenberg, A. (2014). Similarity network fusion for aggregating data types on a genomic scale. *Nature Methods*, *11*, 333–337. https://doi.org/10.1038/nmeth.2810

Wang, F., & Sun, J. (2015). PSF: A unified patient similarity evaluation framework through metric learning with weak supervision. *IEEE Journal of Biomedical and Health Informatics*, *19*, 1053–1060. https://doi.org/10.1109/JBHI.2015.2425365

Wang, R.-S., & Loscalzo, J. (2018). Network-based disease module discovery by a novel seed connector algorithm with Pathobiological implications. *Journal of Molecular Biology*, *430*, 2939–2950. https://doi.org/10.1016/j.jmb.2018.05.016

Wang, T., Bendayan, R., Msosa, Y., Pritchard, M., Roberts, A., Stewart, R., & Dobson, R. (2022). Patient-centric characterization of multimorbidity trajectories in patients with severe mental illnesses: A temporal bipartite network modeling approach. *Journal of Biomedical Informatics*, *127*, 104010. https://doi.org/10.1016/j.jbi.2022.104010

West, L., Vidwans, S. J., Campbell, N. P., Shrager, J., Simon, G. R., Bueno, R., Dennis, P. A., Otterson, G. A., & Salgia, R. (2012). A novel classification of lung cancer into molecular subtypes. *PLoS One*, *7*, e31906. https://doi.org/10.1371/journal.pone.0031906

Wishart, D. S., Feunang, Y. D., Guo, A. C., Lo, E. J., Marcu, A., Grant, J. R., Sajed, T., Johnson, D., Li, C., Sayeeda, Z., Assempour, N., Iynkkaran, I., Liu, Y., Maciejewski, A., Gale, N., Wilson, A., Chin, L., Cummings, R., Le, D., ... Wilson, M. (2018). DrugBank 5.0: A major update to the DrugBank database for 2018. *Nucleic Acids Research*, *46*, D1074–D1082. https://doi.org/10.1093/nar/gkx1037

Witt, H., Mack, S. C., Ryzhova, M., Bender, S., Sill, M., Isserlin, R., Benner, A., Hielscher, T., Milde, T., Remke, M., Jones, D. T. W., Northcott, P. A., Garzia, L., Bertrand, K. C., Wittmann, A., Yao, Y., Roberts, S. S., Massimi, L., Van Meter, T., ... Pfister, S. M. (2011). Delineation of two clinically and molecularly distinct subgroups of posterior fossa Ependymoma. *Cancer Cell*, *20*, 143–157. https://doi.org/10.1016/j.ccr.2011.07.007

Wu, T., Wang, Y., Wang, Y., Zhao, E., & Yuan, Y. (2021). Leveraging graph-based hierarchical medical entity embedding for healthcare applications. *Scientific Reports*, *11*, 5858. https://doi.org/10.1038/s41598-021-85255-w

Xu, T., Le, T. D., Liu, L., Wang, R., Sun, B., & Li, J. (2016). Identifying cancer subtypes from miRNA-TF-mRNA regulatory networks and expression data. *PLoS One*, *11*, e0152792. https://doi.org/10.1371/journal.pone.0152792

Yang, C., Ge, S.-G., & Zheng, C.-H. (2017). ndmaSNF: Cancer subtype discovery based on integrative framework assisted by network diffusion model. *Oncotarget*, *8*, 89021–89032. https://doi.org/10.18632/oncotarget.21643

Yin, T., Chen, S., Wu, X., & Tian, W. (2017). GenePANDA—A novel network-based gene prioritizing tool for complex diseases. *Scientific Reports*, *7*, 43258. https://doi.org/10.1038/srep43258

Zhong, X., Yang, H., Zhao, S., Shyr, Y., & Li, B. (2015). Network-based stratification analysis of 13 major cancer types using mutations in panels of cancer genes. *BMC Genomics*, *16*, S7. https://doi.org/10.1186/1471-2164-16-S7-S7

Zhou, Y., Hou, Y., Shen, J., Huang, Y., Martin, W., & Cheng, F. (2020). Network-based drug repurposing for novel coronavirus 2019-nCoV/SARS-CoV-2. *Cell Discovery*, *6*, 1–18. https://doi.org/10.1038/s41421-020-0153-3

How to cite this article: Petti, M., & Farina, L. (2023). Network medicine for patients' stratification: From single-layer to multi-omics. *WIREs Mechanisms of Disease*, *15*(6), e1623. https://doi.org/10.1002/wsbm.1623