1. **Patient similarity for precision medicine: A systematic review. Parimbelli. 2018** https://www.sciencedirect.com/science/article/pii/S1532046418301072

- **Types of Data**:
  - Molecular (genomic), Data Integration, Clinical, Imaging, Lab (Fig 2)
- **Analysis**:
  - Clustering (census)
  - Dimensionality Reduction
  - Similarity Measures (info theory, disease state index)
  - Clustering + Supervised Learning

2. Patient Similarity Networks for Precision Medicine. Shraddha. 2018 https://www.sciencedirect.com/science/article/pii/S0022283618305321?via%3Dihub

"**Patient similarity networks** are an emerging paradigm for precision medicine, in which patients are clustered or classified based on their similarities in various features, including genomic profiles."

- Table 1. Pros/Cons of Methods used in clinical risk models
- Patient data are encoded as "input features" (e.g., age, gender, genotypes at individual SNPs, metabolite quantities, gene expression levels)
- Interpretability - Random Forest (decision tree)
- PSN - each input patient data feature (e.g., age, sex, mutation status) is represented as a network of pairwise patient similarities

\+ Can handle heterogeneous data

3. L. Li, *et al.* Identification of type 2 diabetes subgroups through topological analysis of patient similarity https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4780757/

- “We developed a novel TDA-based (topology data analysis) approach to perform unsupervised clustering of patients using various clinical features to produce a patient-patient network organized according to the high-dimensional clinical phenotype similarity among patients.”
- cosine distance metric

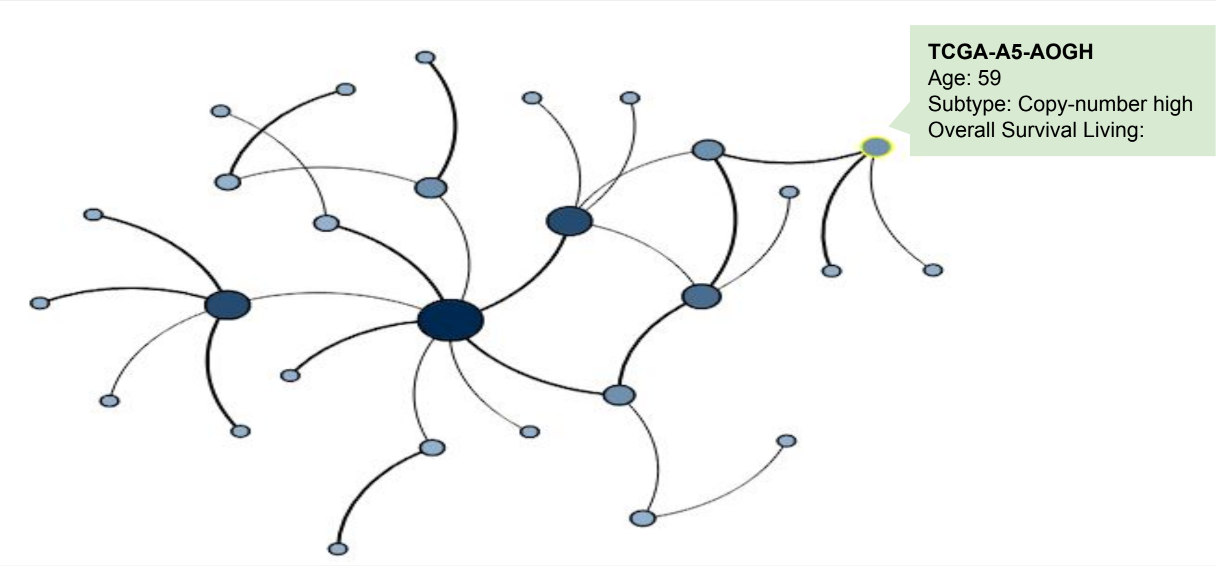
4. B. Wang, *et al.* Similarity network fusion for aggregating data types on a genomic scale http://compbio.cs.toronto.edu/SNF/SNF/Software.html

- R/Matlab code.
- “SNF first constructs a sample similarity network for each of the data types and then iteratively integrates these networks using a novel network fusion method.“

5. Visualizing omics and clinical data: Which challenges for dealing with their variety?

https://www.researchgate.net/publication/319574674_Visualizing_omics_and_clinical_data_Which_challenges_for_dealing_with_their_variety

- Big Data/High Dimensions = # patients * # clinical and genomic data.
- Scalability: abstract visualizations: grouping entities to meta-entities. zoom.
- Relational data visualized: Node-link diagram (Cytoscape, Gephi, Tulip), Matrix-based diagram
- "Only few softwares provide **extended interactive explorations**"
- State of the art software
    - Java TreeView: heatmaps and scatter plots
    - Cytoscape: omics data only
    - IGV: omics data and clinical data as metadata
- Goal: 1) Overview of clinical and omics data 2) Interactive



Similarity Measure: jaccard

- ◉ Jaccard
- ○ Cosine
- ○ Cluster

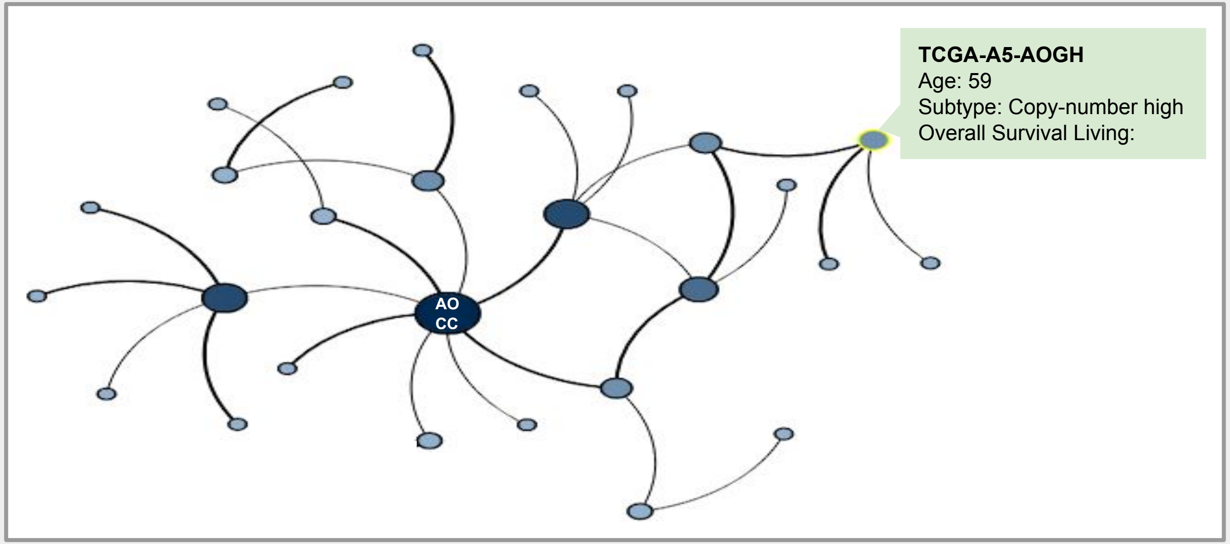
Features: Age

- ☑ Age
- ☑ Genomic alterations
- ☐ Mutation count

Patients: 100

Patient ID	Age	Overall Survival	Subtype	Historical Subtype	...	Similarity Metric
TCGA-BK-A0CC	69	LIVING	Copy-number high	Serious	...	1.0
TCGA-A5-A0GH	57	LIVING	MSI (Hyper-mutated)	Endometrioid	...	0.9
TCGA-AX-A062	53	LIVING	Copy-number low	Endometrioid	...	0.8
TCGA-D1-A17L	81	LIVING	Copy-number low	Endometrioid	...	0.7

Patient Similarity Network



Similarity Metric:

Jaccard

- (•) Jaccard
- ( ) Cosine
- ( ) Cluster

Patient Features:

Tumor Type

- [■] Tumor Type
- [■] Common Genomes
- [ ] Alterations

 # Patients: 100

Genomic Data

Patient ID	Age	Overall Survival	Subtype	...	Similarity
TCGA-BK-A0CC	69	LIVING	Copy-number high	...	1.0
TCGA-A5-A0GH	57	LIVING	MSI (Hyper-mutated)	...	0.9
TCGA-AX-A062	53	LIVING	Copy-number low	...	0.8
TCGA-D1-A17L	81	LIVING	Copy-number low	...	0.7
TCGA-D5-A0GL	46	LIVING	MSI (Hyper-mutated)	...	0.7

Cancer Type