

Exploring Network Fusion methods for Multi-omic Data Integration

Ramón Reszat, Rachel Cavill^A

^A Department of Advanced Computing Sciences, Maastricht University, Maastricht, The Netherlands

Objectives

- Clustering cell lines from the Cancer Cell Line Encyclopedia (CCLE) based on their multi-omics profiles
- Comparing Similarity Network Fusion (SNF) and Affinity Network Fusion (ANF) on the CCLE
- Examine how the hyperparameters in SNF and ANF influence the topology of the fused networks

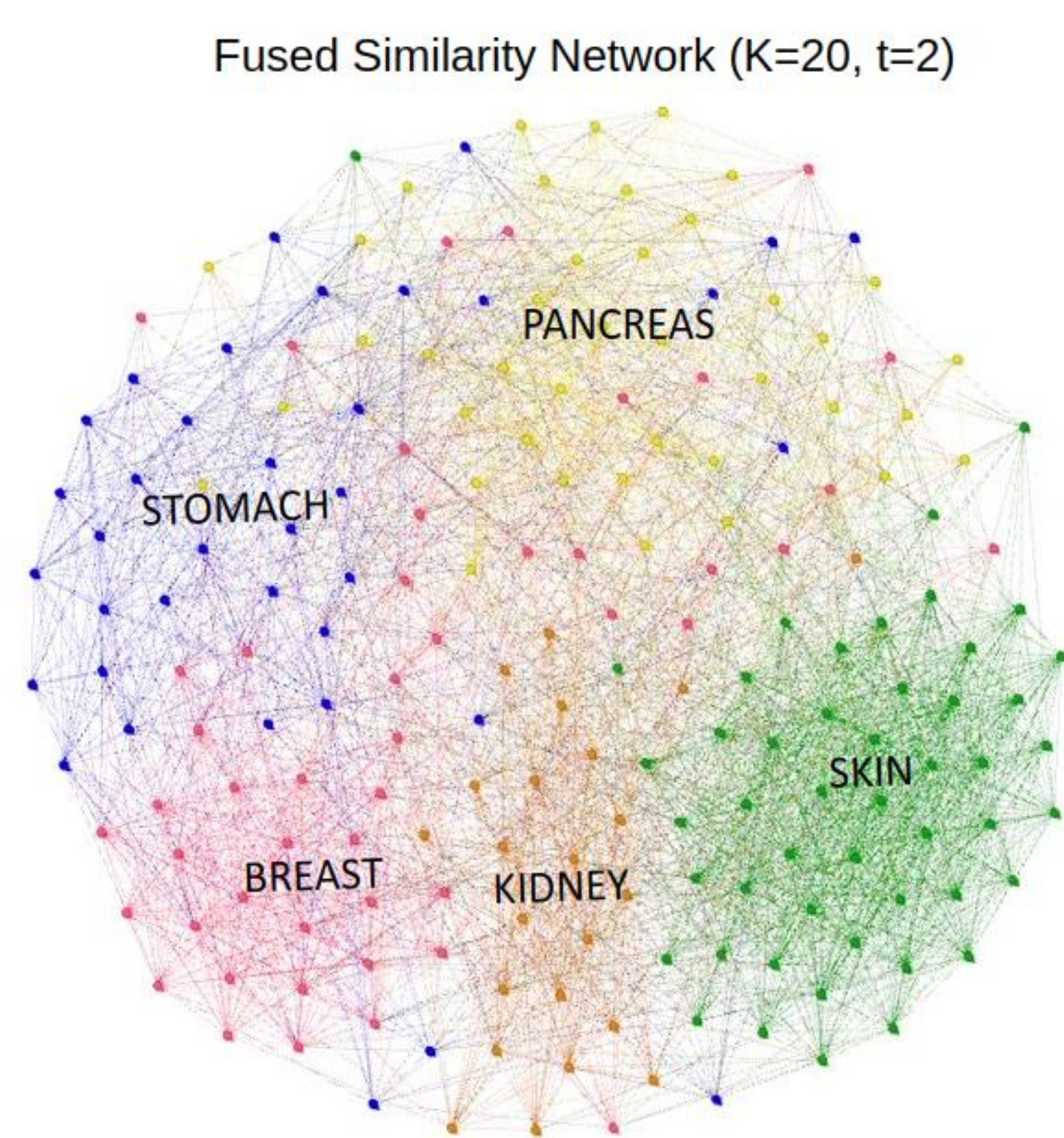


Figure 1: Fused network from SNF after $t=2$ diffusion steps with $k=20$ nearest-neighbors (gene + metabolites + micrnas)

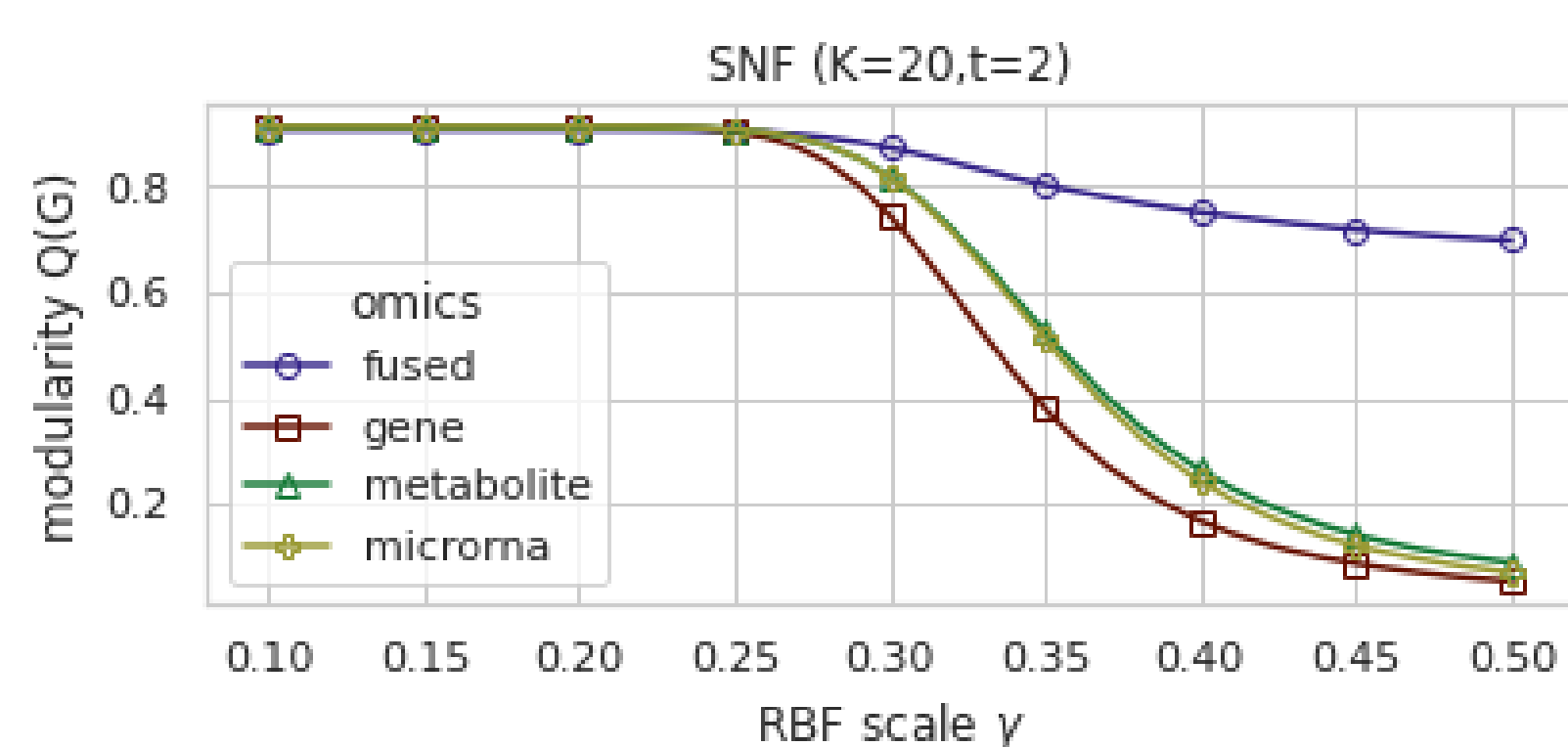


Figure 3: Modularity of the affinity graph constructed by SNF for the scale γ of the radial basis function kernel (gene + metabolites + micrnas)

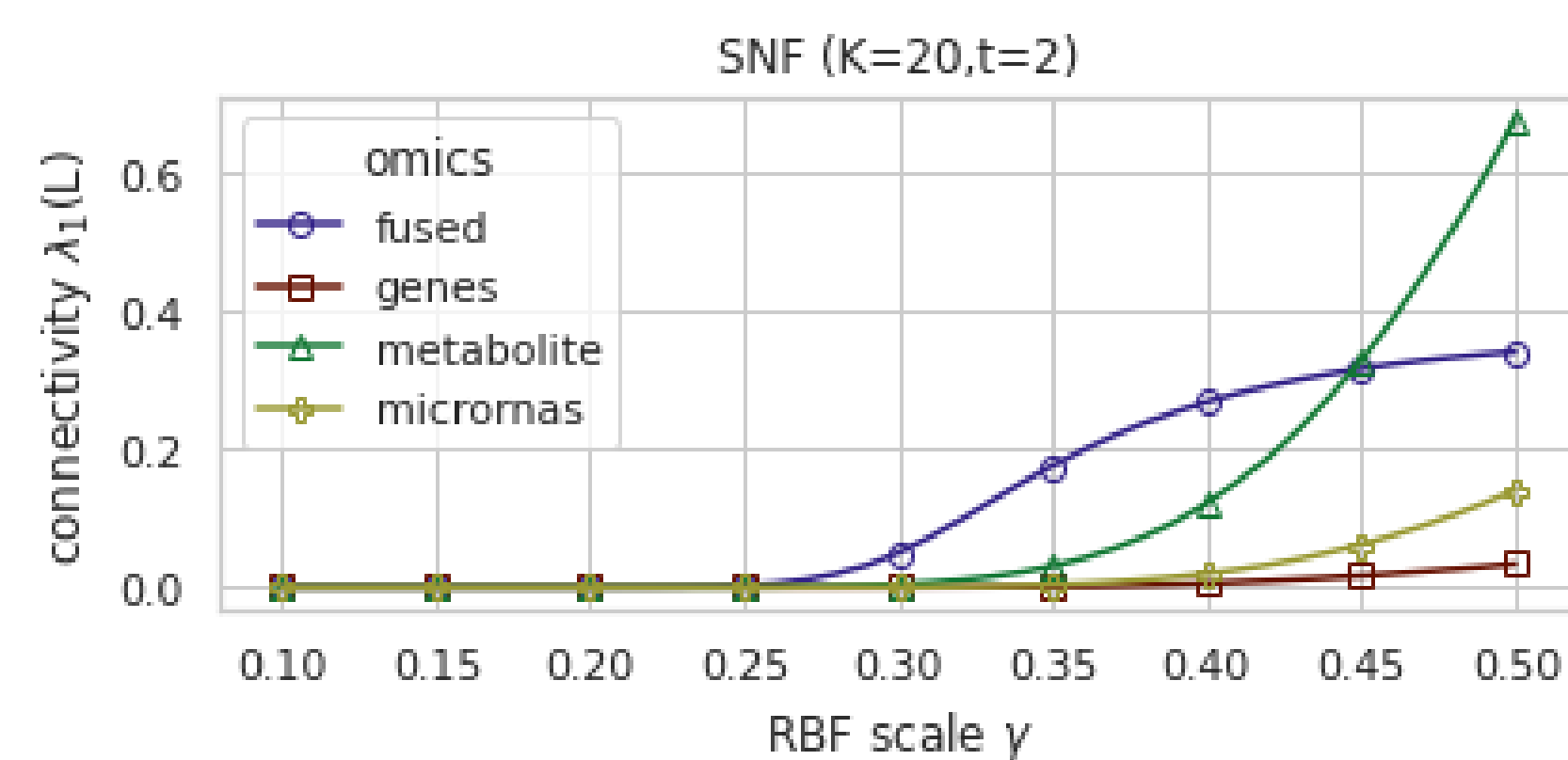


Figure 4: Smallest non-zero eigenvalue of the graph Laplacian for the scale γ of the radial basis function kernel (gene + metabolites + micrnas)

Fused Affinity Network ($K=20, t=2$)

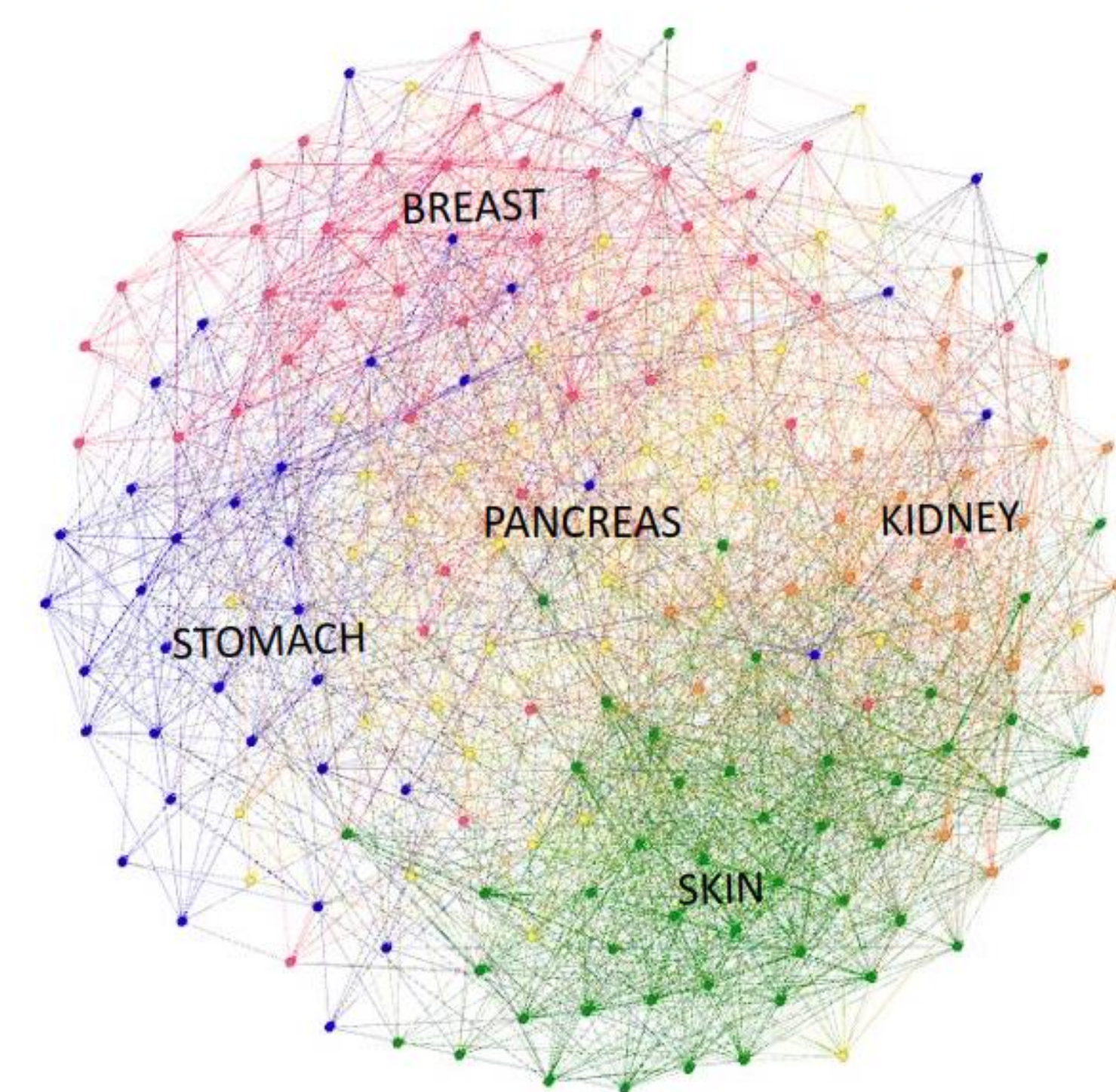
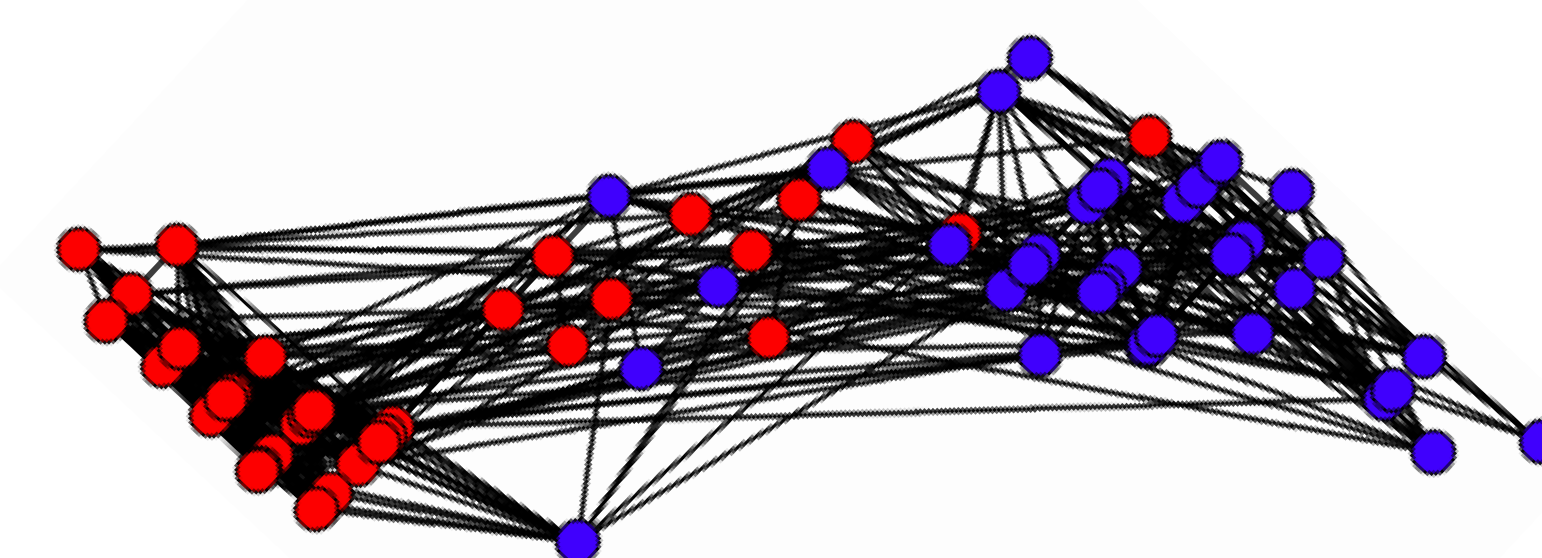


Figure 2: Fused network from ANF after $t=2$ diffusion steps with $k=20$ nearest-neighbors (gene + metabolites + micrnas)



Purity: 87%
ARI: 0.54
NMI: 0.46

	0	1
adenocarcinoma	37	2
small_cell_carcinoma	8	29

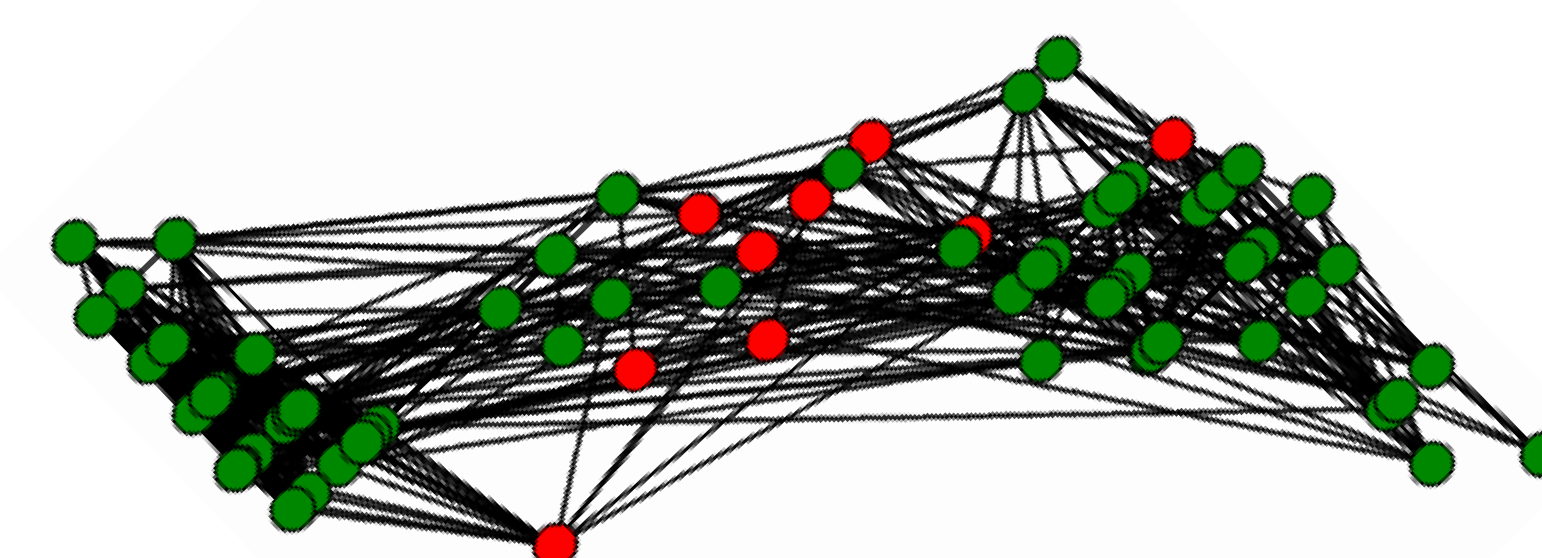


Figure 5: KMeans clustering on a GSVD embedding of the graph for histological lung cancer subtypes (red: small-cell lung carcinoma & blue: adenocarcinoma)

Results

- Subtype clustering on the fused ANF network achieves a 87% purity on the CCLE lung cancer types and 73% on the kidney cancer type.
- The scaling factors in SNF and ANF change the connectivity and modularity of the networks. Both stay high in the fused network.
- Fusing multi-omics data results in better clustering ($p=6.7e-06$) of dasatinib resistant small-cell lung carcinoma cell lines when compared to single-omics networks (genes: $p=1.5e-02$, metabolites: $p=3.9e-02$, miRNAs: $p=1.2e-03$).

Link

Browse this project on GitHub <https://github.com/ramonreszat/affinity-graphs>:

