Exploring selective edge pruning to integrate domain knowledge in biological networks

Vlad V. Ungureanu[1,2,3], David M. Halliday[1,3], Jennifer Southgate[2,3], Stephen L. Smith[1,3], Andrew S. Mason[2,3]

1 - School of Physics, Engineering and Technology, University of York
2 - Jack Birch Unit, Department of Biology, University of York
3 - York Biomedical Research Institute, University of York

vlad.ungureanu@york.ac.uk
@vladUngu

Motivation – Bladder Cancer

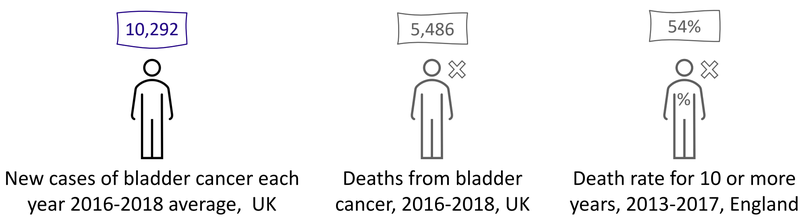


Fig 1 Bladder cancer statistics from Cancer Research UK

- For the first time, information from a normal tissue dataset is used to inform the stratification of the muscle-invasive bladder cancer (MIBC) cohort from The Cancer Genome Atlas (TCGA)[1]
- Transcription Factors (TFs) are genes encoding regulators of gene expression and are prioritised in the edge pruning strategy

Methods

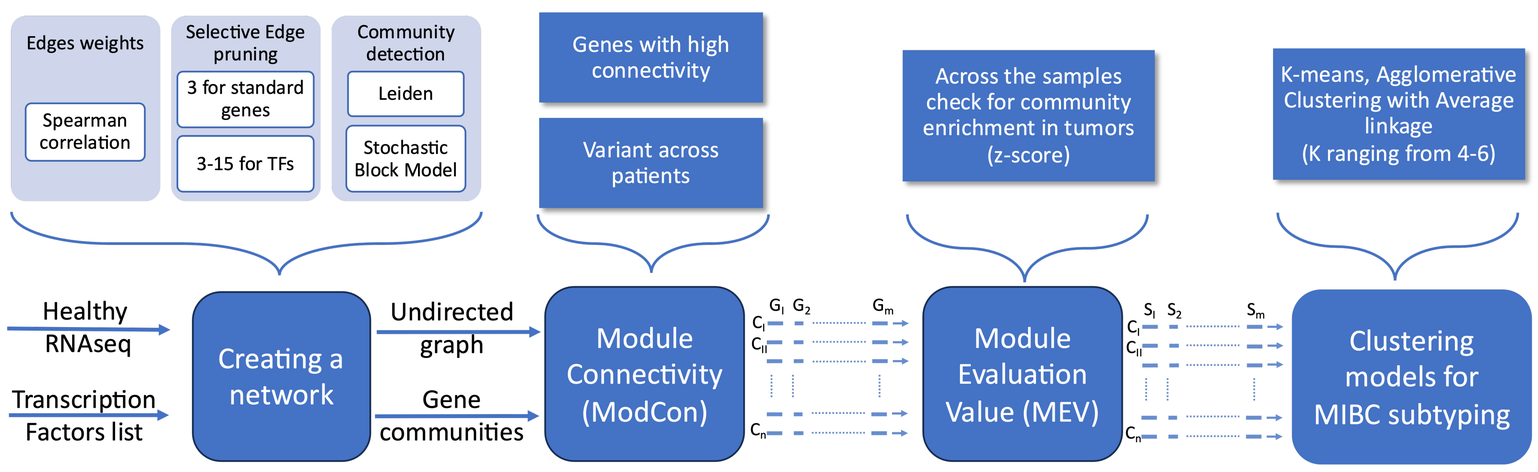


Fig 2 Network pipeline based on PGCNA[2]. The network is constructed from the 5,000 most variable genes, which include 324 TFs, taken from 88 non-cancerous samples. The tumour dataset comprises 408 samples.

Leiden vs Stochastic Block Model (SBM)

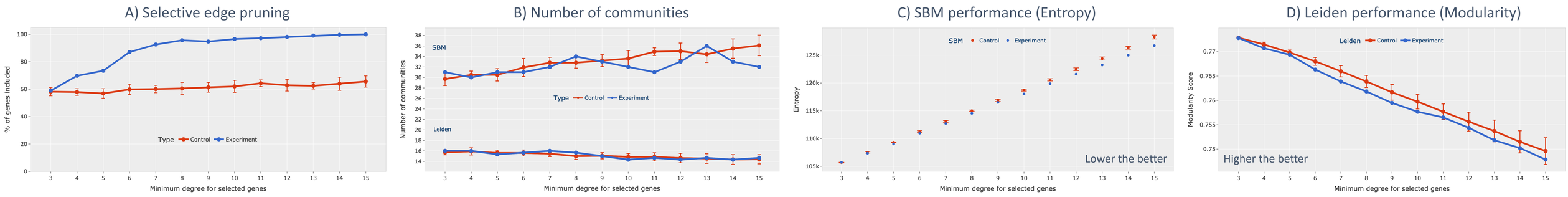
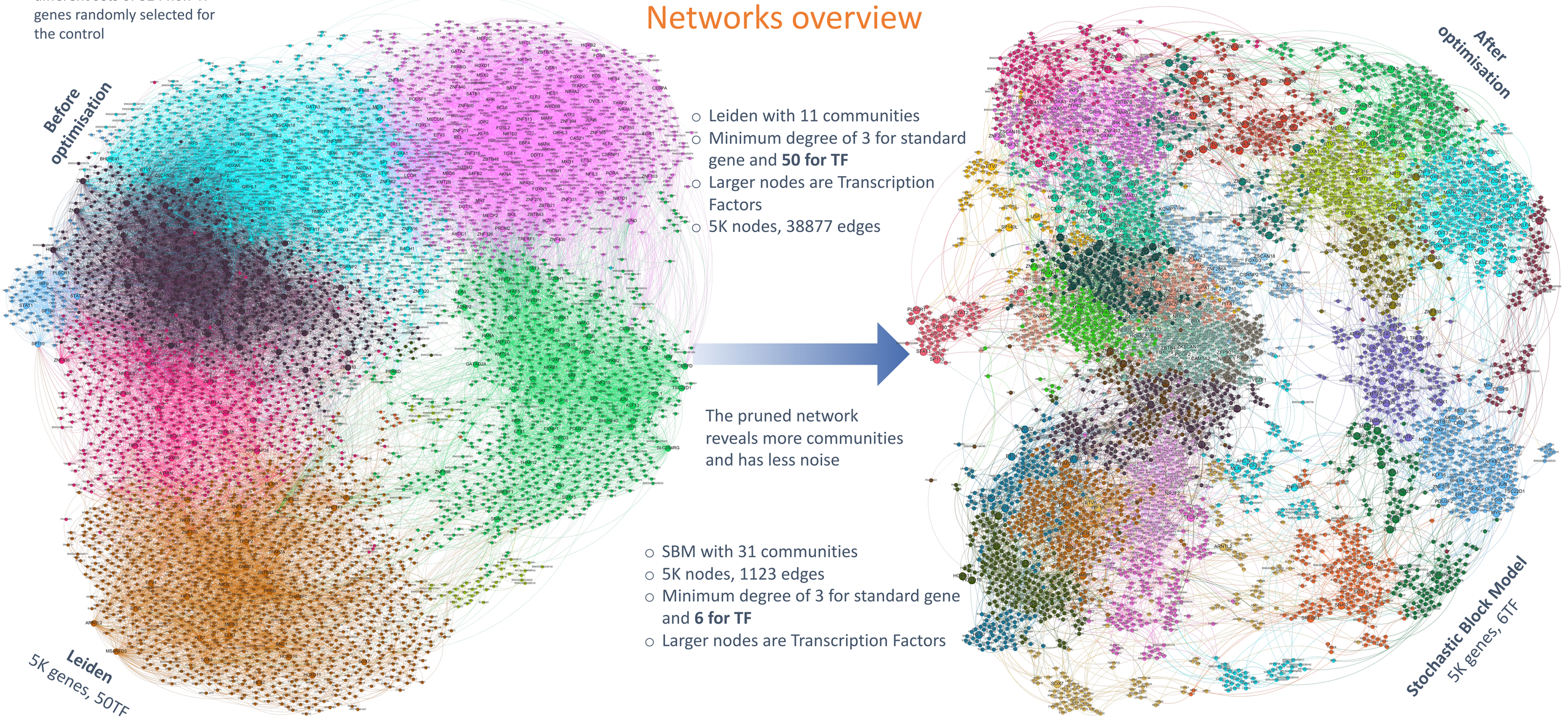


Fig 3 A series of networks were generated where selected genes (TF and control) have a minimum degree from 3-15. For each experiment there are 10 different sets of 324 non-TF genes randomly selected for the control

- In Fig 3 A) the selective pruning accentuates the role of TF in the networks with little benefit of allowing more than 6 edges for the TFs.
- TF does not seem to help community detection more than random genes

- The number of communities increases SBM[2, 7] tends to find more communities while Leiden[3] less (Fig 3 B)
- Performance of both algorithms declines proportional to the number of edges permitted for TF (see Fig 3 C, D)

Networks overview



- Leiden with 11 communities
- Minimum degree of 3 for standard gene and **50 for TF**
- Larger nodes are Transcription Factors
- 5K nodes, 38877 edges

The pruned network reveals more communities and has less noise

- SBM with 31 communities
- 5K nodes, 1123 edges
- Minimum degree of 3 for standard gene and **6 for TF**
- Larger nodes are Transcription Factors

Muscle invasive bladder cancer subtyping

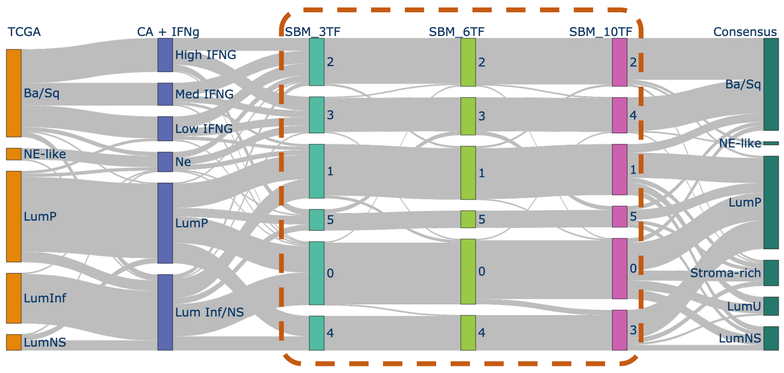


Fig 5 MIBC subtypes comparison between the network approach, the TCGA[2], consensus[**4**], and our previous work with K-means and insitu[5]

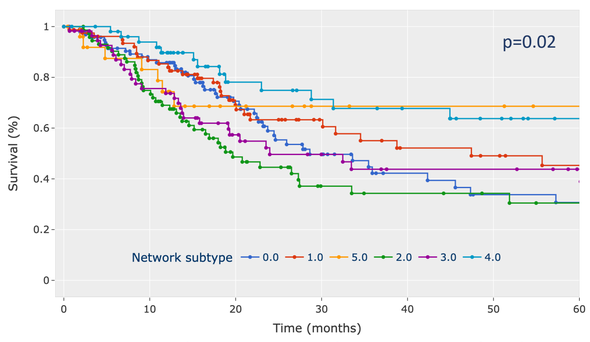


Fig 6 MIBC Subtypes survival comparison for the Network with TF=6

- While the network approach reveals subtypes that differ from those identified by standard methods, the variation in the number of TFs does not directly influence the subtypes of MIBC.
- Significant survival difference of the MIBC subtypes derived from the network with TF=6 and SBM

Discussion

- Our results show the potential of using a network approach to stratify muscle invasive bladder cancer
- Selective edge pruning emphasizes the role of the TFs within the network, but its impact diminishes when the number of edges for TFs exceeds six.
- Both Leiden and SBM perform worse when more edges are allowed, but SBM tends to find more communities

Future work

- Integrate the impact of gene mutations into the network pipeline
- Apply Hierarchical Stochastic Block Model to the network and improve the integration between healthy and tumour datasets

[1] A. G. Robertson et al., "Comprehensive Molecular Characterization of Muscle-Invasive Bladder Cancer", *Cell*, vol. 171, no. 3, pp. 540–556.e25**. [2]** Care, Matthew, et al. 2019. 'Parsimonious Gene Correlation Network Analysis (PGCNA): A Tool to Define Modular Gene Co-Expression for Refined Molecular Stratification in Cancer'. *NPJ Systems Biology and Applications* 5 (April): 13. **[3]** Peixoto, Tiago P. 2019. 'Bayesian Stochastic Blockmodeling'. Advances in Network Clustering and Blockmodeling. Wiley.**[4]** Traag et al. 2019. 'From Louvain to Leiden: Guaranteeing Well-Connected Communities'. *Scientific Reports* 9 (1): 5233 **[5]** Kamoun et al. 2020. 'A Consensus Molecular Classification of Muscle-Invasive Bladder Cancer'. *European Urology* 77 (4): 420–33. **[6]** Baker, Simon C., et al. Gamma Predicts T1 Recurrence-Free and Basal/Squamous Muscle-Invasive Bladder Cancer Survival and Better Targeted Strategies for Immune Checkpoint Blocking." **[7]** Peixoto, Tiago P. 2014. 'The Graph-Tool Python Library'.