Treelet Covariance Smoothers

Better Estimation of Heritability

B. Draves[1]

[1]Department of Mathematics
Lafayette College

Advisor: T. Gaugler

Lafayette College, 2016

What is Treelets?

An adaptive method for multi-scale representation and eigenanalysis of data where the variables can occur in any given order.[1]

[1] Crossett, A., Lee, A. B., Klei, L., Devlin, B., and Roeder, K., Refining Genetically Inferred Relationships Using Treelet Covariance Smoothing, Annals of Applied Statistics, 7(2):669 690, 2013.

What is Treelets?

An adaptive method for multi-scale representation and eigenanalysis of data where the variables can occur in any given order.[1]

What does *that* mean?

[1] Crossett, A., Lee, A. B., Klei, L., Devlin, B., and Roeder, K., Refining Genetically Inferred Relationships Using Treelet Covariance Smoothing, Annals of Applied Statistics, 7(2):669 690, 2013.

What is Treelets? - My Take

Treelets is a method to systematically cluster correlated data in a way that yields information of the underlying structure while smoothing noisy samples.

SNP Samples

- Each person has a specific genetic composition which can be described by a sample of single nucleotide polymorphisms (SNPs)

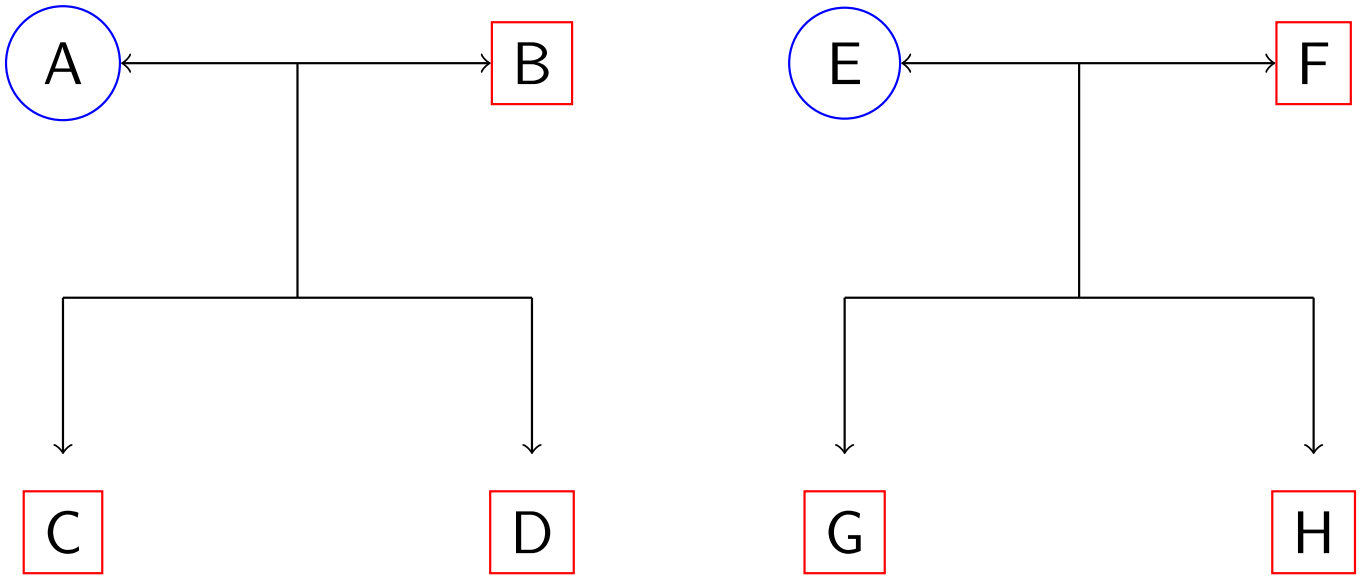
SNP Samples

- Each person has a specific genetic composition which can be described by a sample of single nucleotide polymorphisms (SNPs)
- Idea: Use Treelets to refine an estimate of relatedness in a sample of individuals

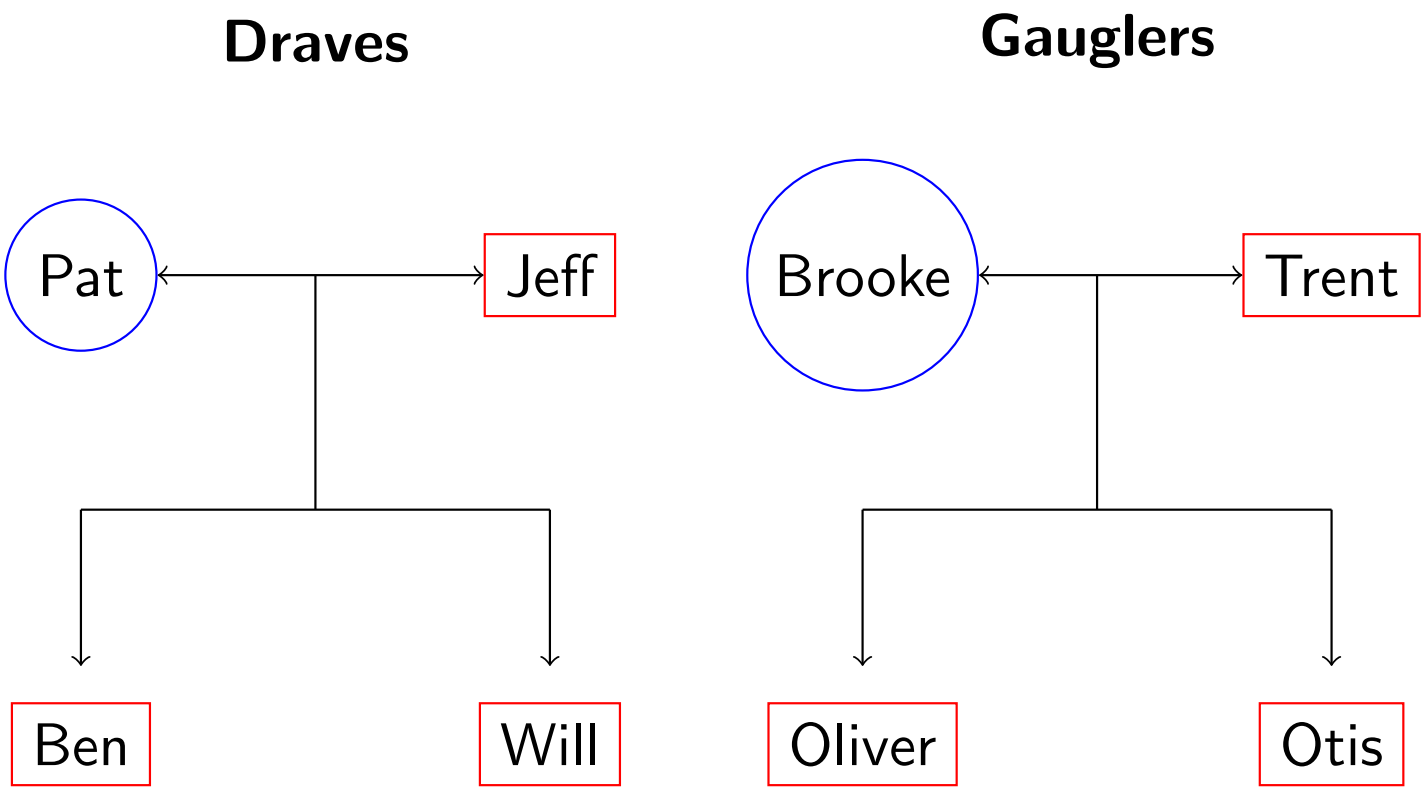
SNP Samples

- Each person has a specific genetic composition which can be described by a sample of single nucleotide polymorphisms (SNPs)
- Idea: Use Treelets to refine an estimate of relatedness in a sample of individuals
- Use this information to estimate heritability of a phenotype in the population

Pedigree Example



Pedigree Example



Estimating Relatedness - Theoretic

	Otis	Will	Trent	Pat	Jeff	Ben	Oliver	Brooke
Otis	1	0	1/2	0	0	0	1/2	1/2
Will	0	1	0	1/2	1/2	1/2	0	0
Trent	1/2	0	1	0	0	0	1/2	0
Pat	0	1/2	0	1	0	1/2	0	0
Jeff	0	1/2	0	0	1	1/2	0	0
Ben	0	1/2	0	1/2	1/2	1	0	0
Oliver	1/2	0	1/2	0	0	0	1	1/2
Brooke	1/2	0	0	0	0	0	1/2	1

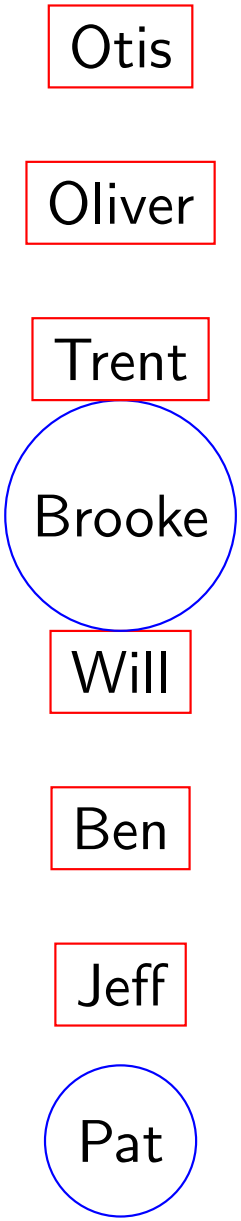
Estimating Relatedness - Theoretic

	Otis	Oliver	Trent	Brooke	Jeff	Ben	Will	Pat
Otis	1	1/2	1/2	1/2	0	0	0	0
Oliver	1/2	1	1/2	1/2	0	0	0	0
Trent	1/2	1/2	1	0	0	0	0	0
Brooke	1/2	1/2	0	1	0	0	0	0
Jeff	0	0	0	0	1	1/2	1/2	0
Ben	0	0	0	0	1/2	1	1/2	1/2
Will	0	0	0	0	1/2	1/2	1	1/2
Pat	0	0	0	0	0	1/2	1/2	1

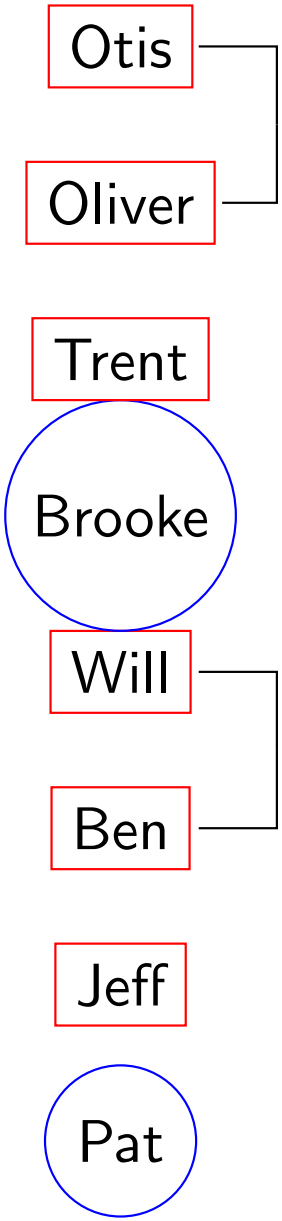
Estimating Relatedness - Sample

	Otis	Will	Trent	Pat	Jeff	Ben	Oliver	Brooke
Otis	0.82	0.31	0.52	0.04	0.39	0.15	0.51	0.44
Will	0.31	0.96	0.10	0.48	0.41	0.43	0.29	0.01
Trent	0.52	0.10	0.89	0.17	0.02	0.09	0.58	0.16
Pat	0.04	0.48	0.17	0.95	0.02	0.45	0.01	0.07
Jeff	0.39	0.41	0.02	0.02	0.83	0.54	0.05	0.13
Ben	0.41	0.43	0.09	0.45	0.54	0.96	0.03	0.04
Oliver	0.51	0.29	0.58	0.01	0.05	0.03	0.85	0.46
Brooke	0.44	0.01	0.16	0.07	0.13	0.04	0.46	0.79

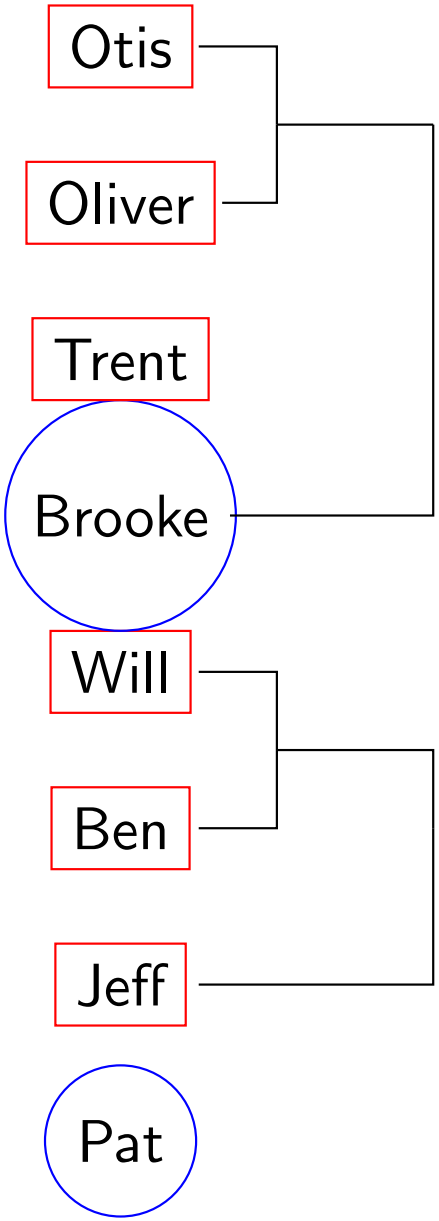
Treelets in Action



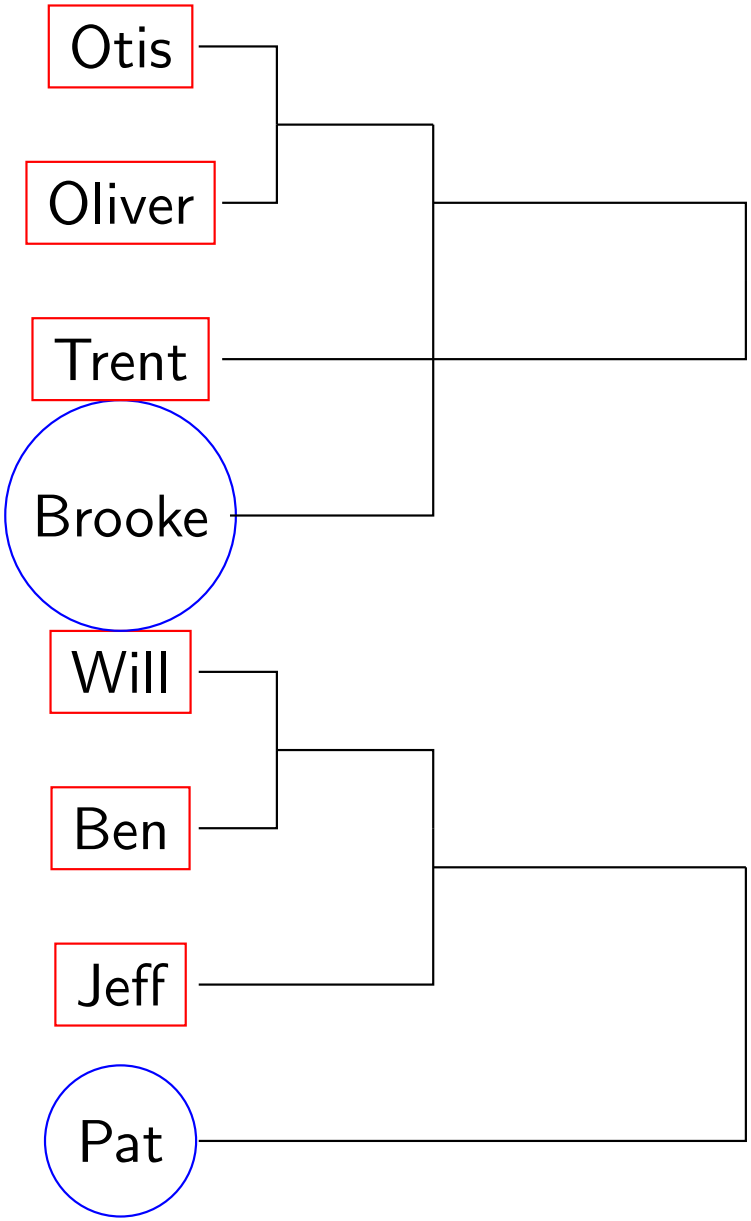
Treelets in Action



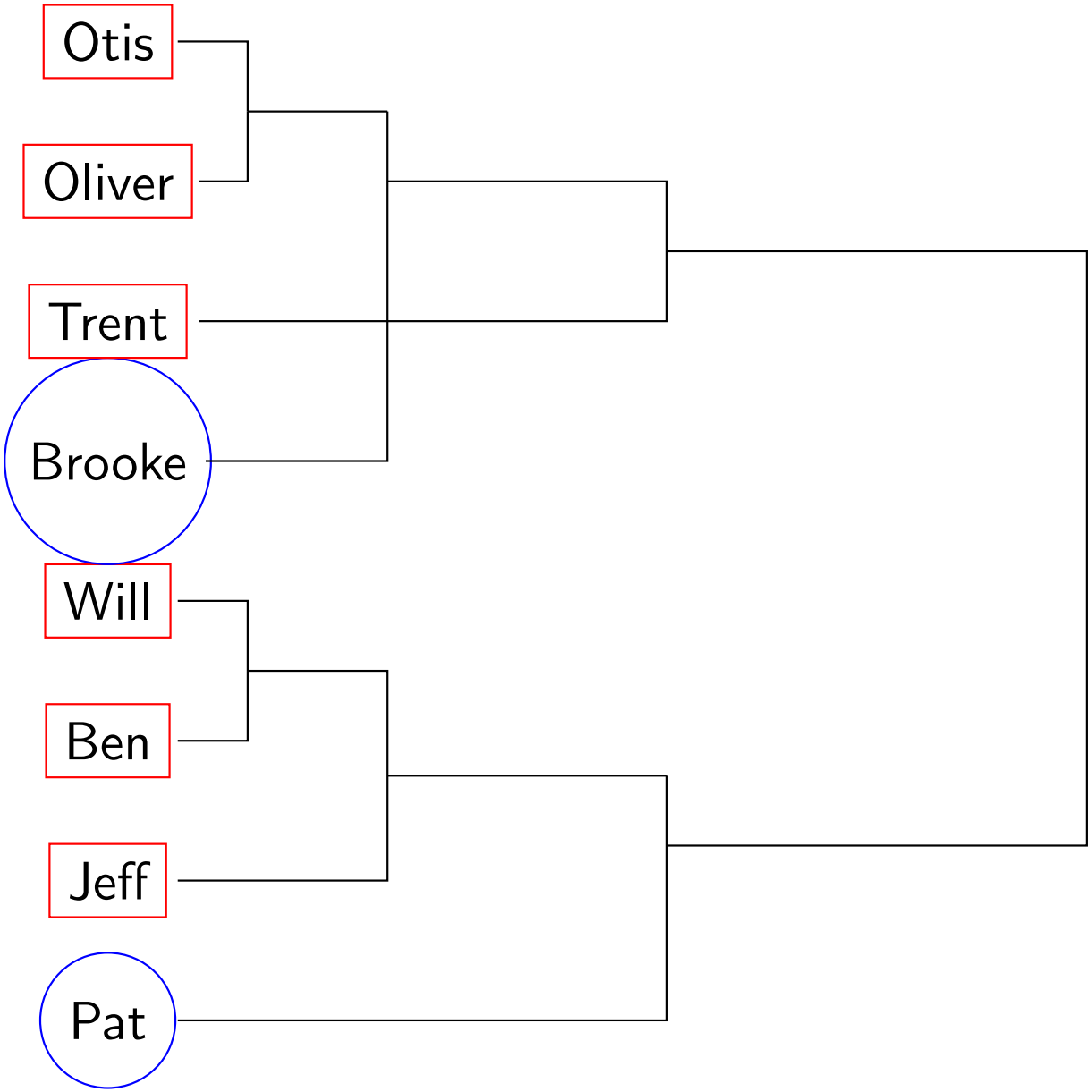
Treelets in Action



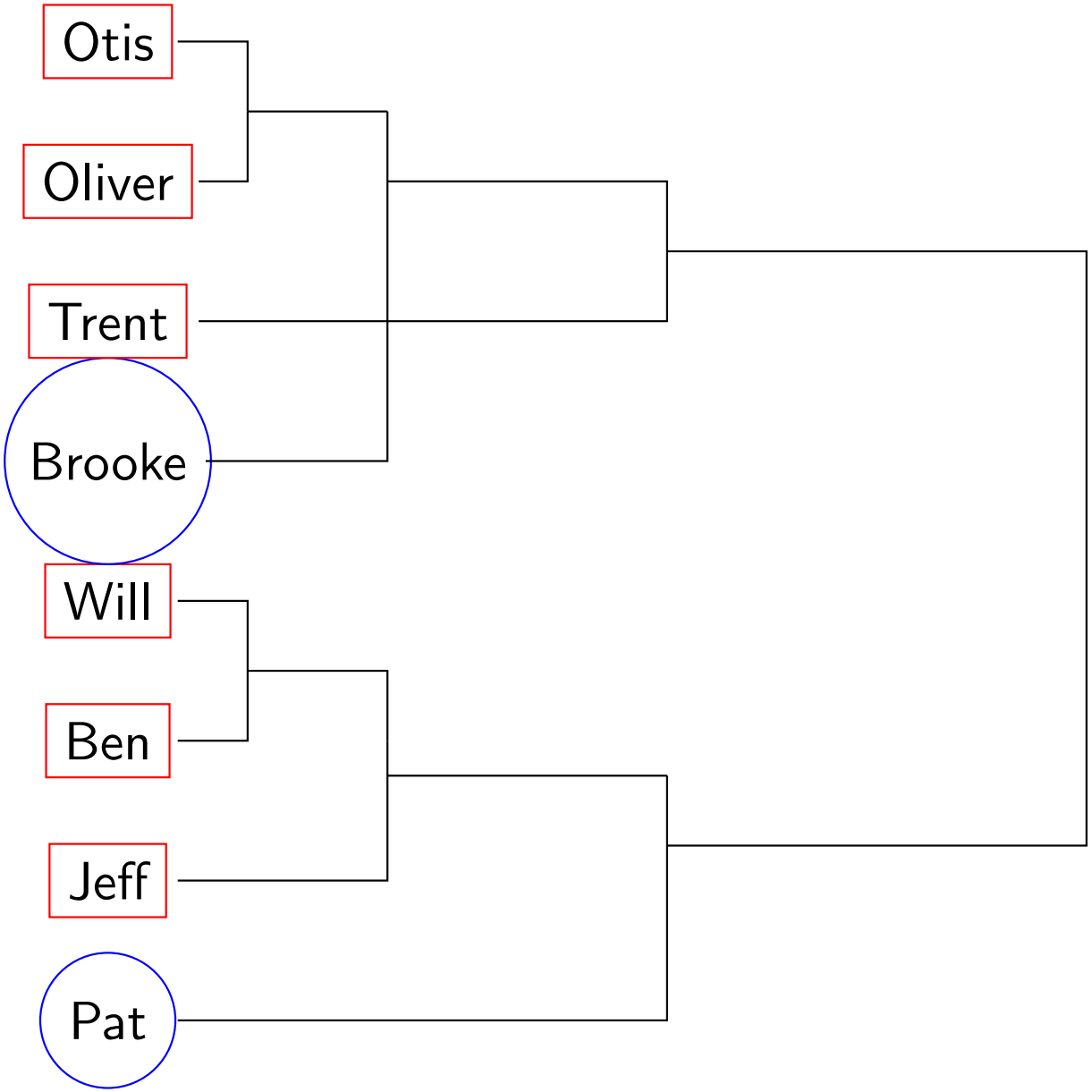
Treelets in Action



Treelets in Action

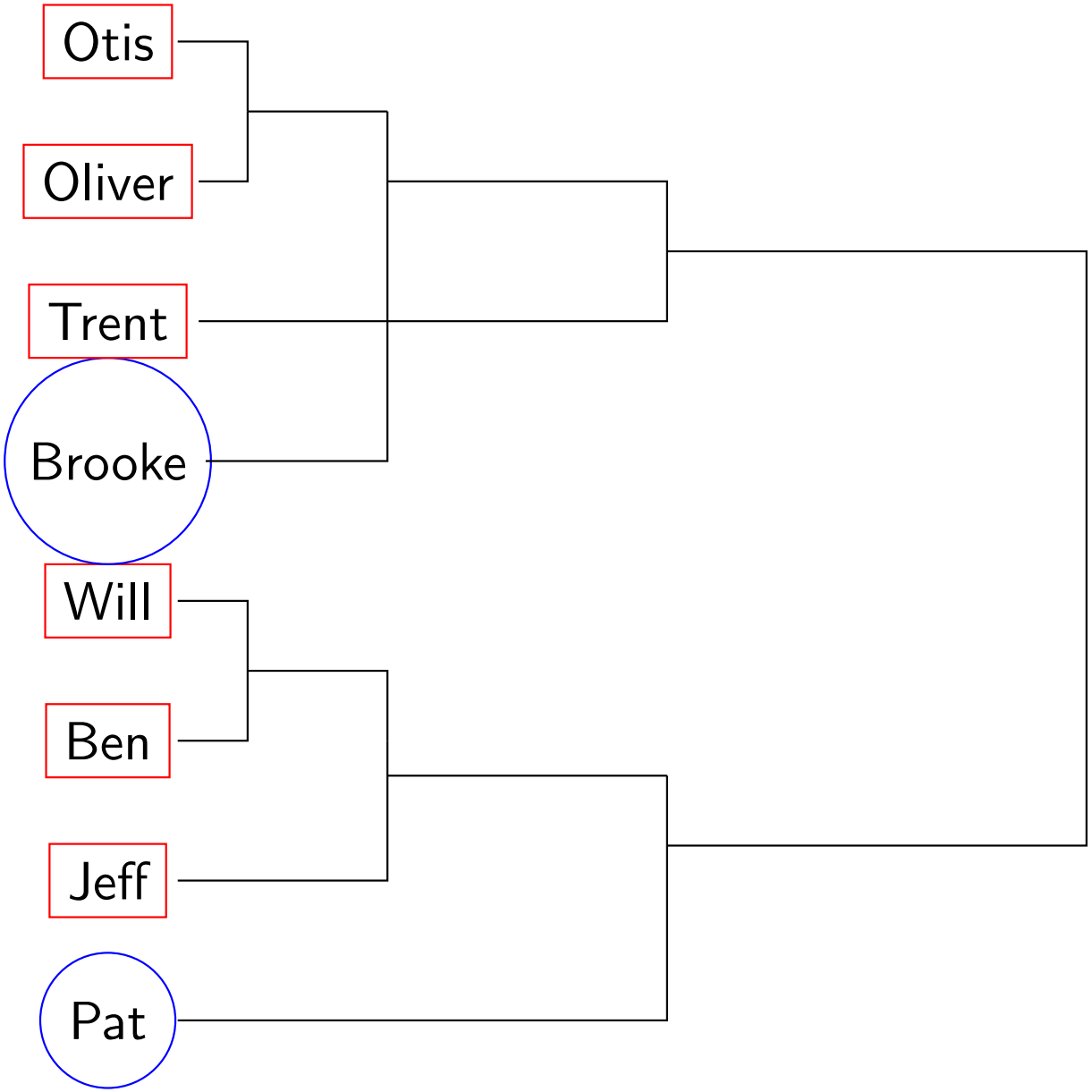


Treelets in Action



1. Associated relationship matrix

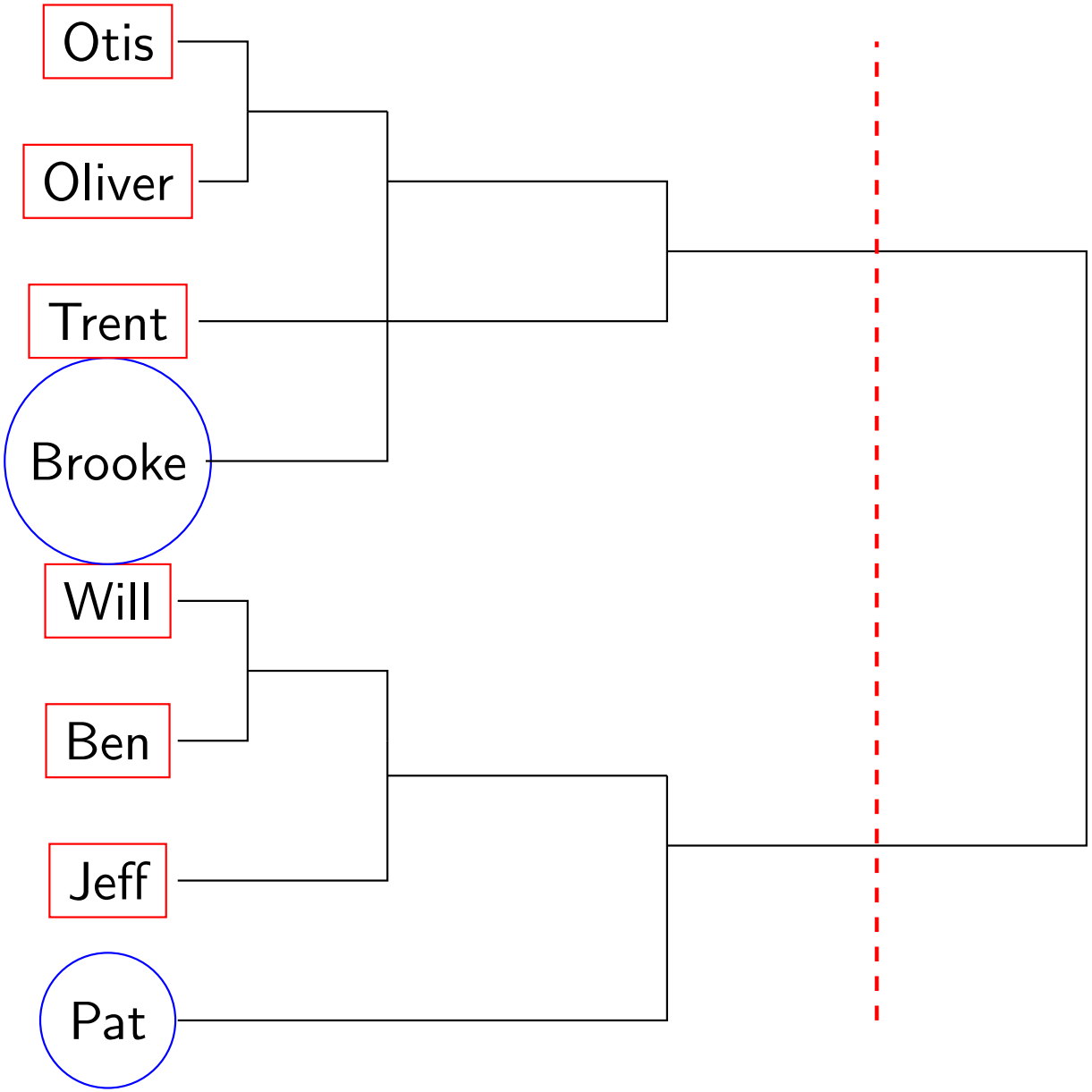
Treelets in Action



1. Associated relationship matrix

2. Set any "insignificant" ($< \lambda$) relationships to zero

Treelets in Action



1. Associated relationship matrix

2. Set any "insignificant" ($< \lambda$) relationships to zero

Alternative Smoothing Methods

- How can we avoid clustering variables we expect are unrelated?

Alternative Smoothing Methods

- How can we avoid clustering variables we expect are unrelated?
- Don't cluster variables we expect are unrelated

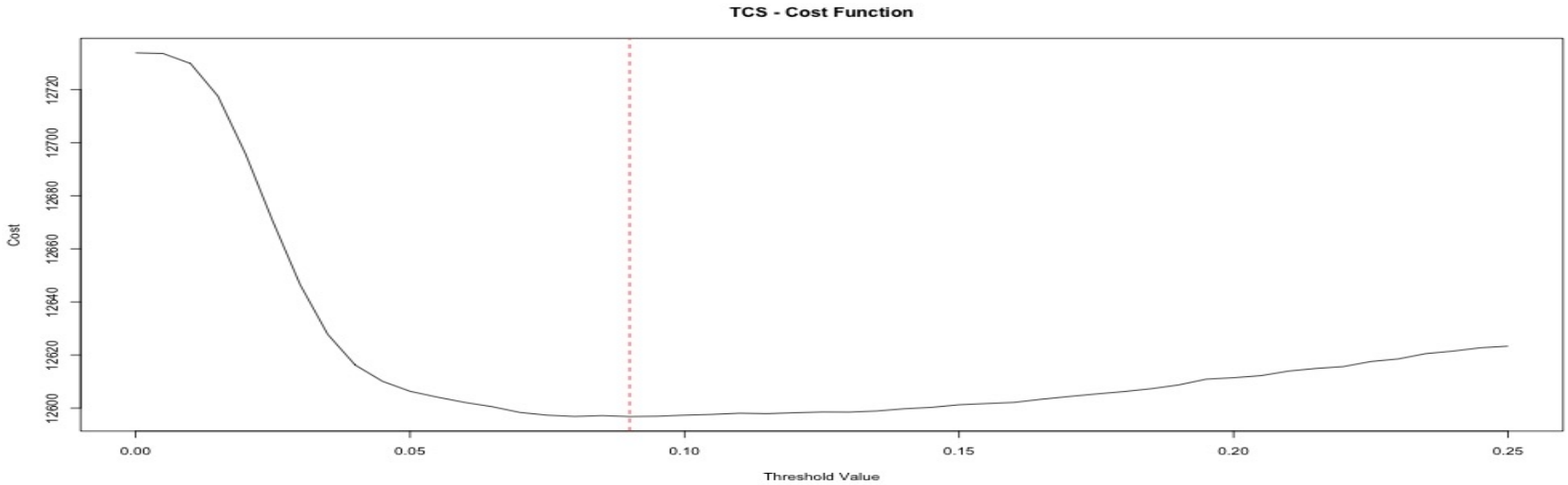
Alternative Smoothing Methods

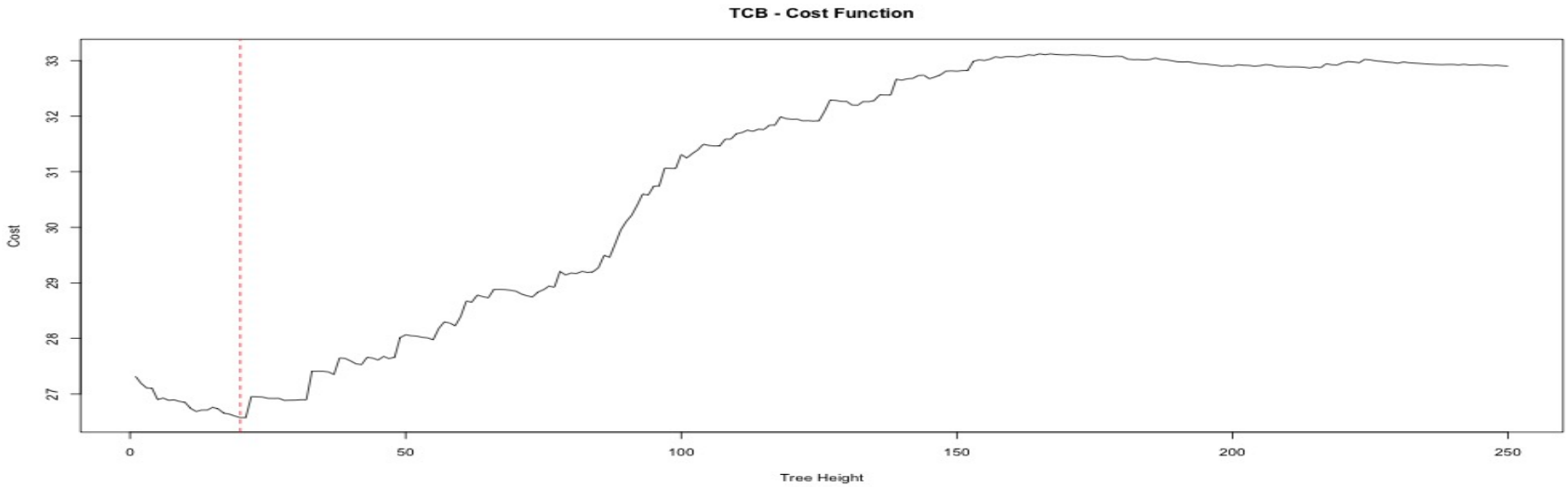
- How can we avoid clustering variables we expect are unrelated?
- Don't cluster variables we expect are unrelated
- Don't require Treelets to merge every variable

Alternative Smoothing Methods

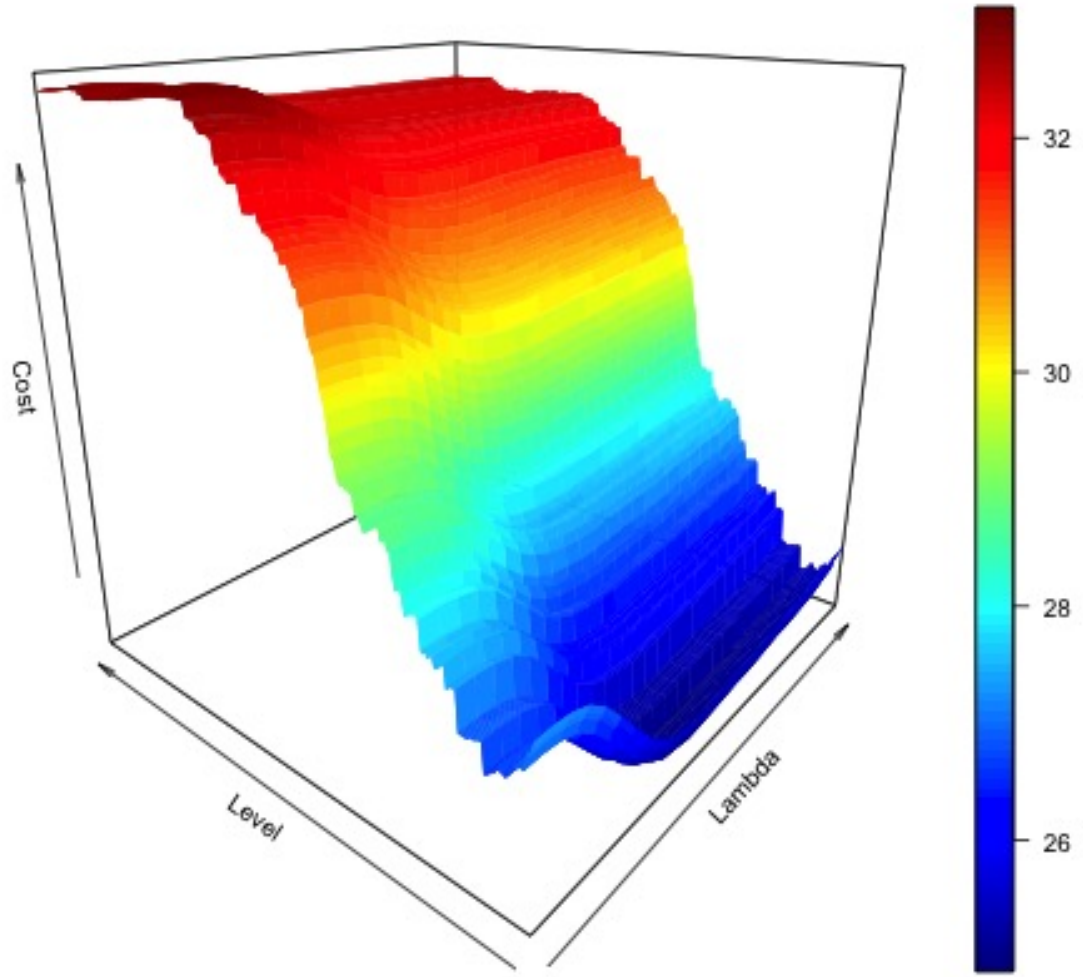
Method	Tree Height (ℓ)	Threshold (λ)
Treelet Covariance Smoothing	Top	Variable
Treelet Covariance Blocking	Variable	Fixed
Treelet Covariance Blocked Smoothing	Variable	Variable

Choosing Smoothing Parameters - TCS & TCB

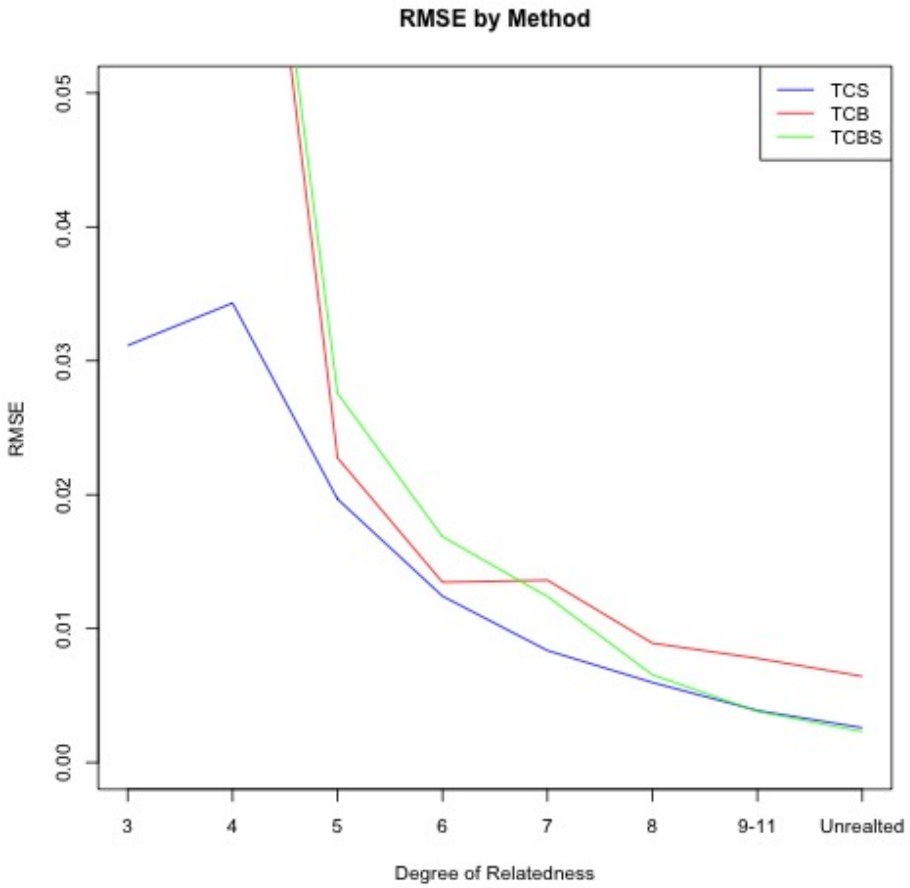


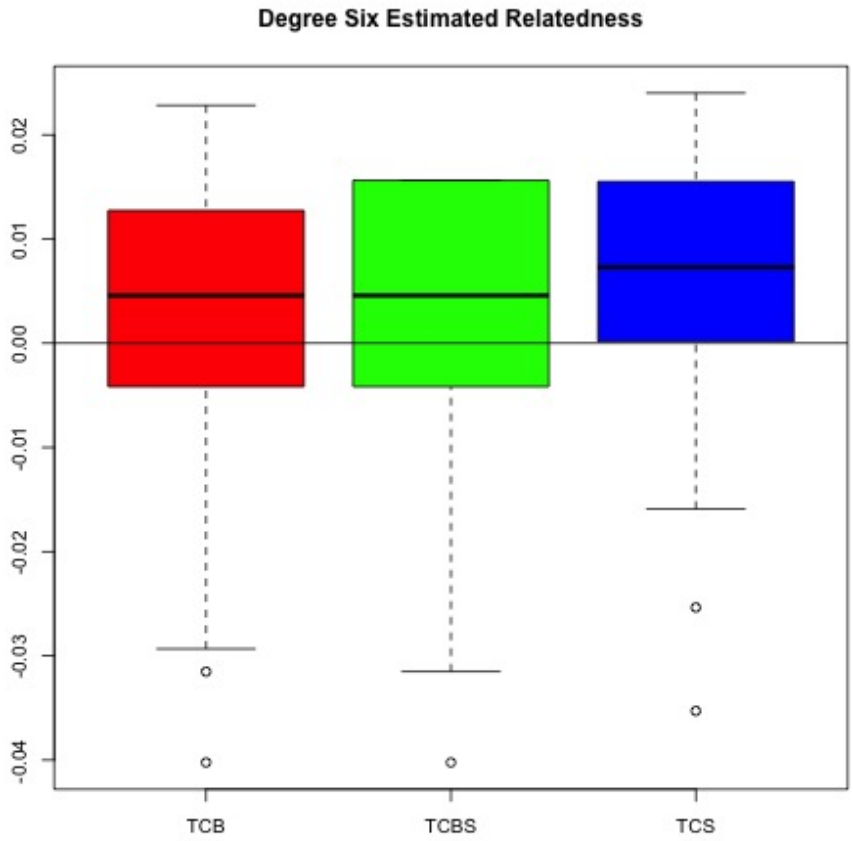


Choosing Smoothing Parameters - TCBS



Estimating Relatedness - Results





Current Work

- Using these refined estimations of relatedness, we estimate Body Mass Index (BMI) and Abdomen Visceral Fat Density (AVFD) in the Health Aging and Body Composition Study

Current Work

- Using these refined estimations of relatedness, we estimate Body Mass Index (BMI) and Abdomen Visceral Fat Density (AVFD) in the Health Aging and Body Composition Study
- TCS has estimated the heritability of BMI close to the literature's accepted range and fails to do so for AVFD

Current Work

- Using these refined estimations of relatedness, we estimate Body Mass Index (BMI) and Abdomen Visceral Fat Density (AVFD) in the Health Aging and Body Composition Study
- TCS has estimated the heritability of BMI close to the literature's accepted range and fails to do so for AVFD
- We look to show TCB or TCBS better estimates heritability of these traits

Conclusions

- Treelets offers a new and useful methodology to the field of high - dimensional statistics

Conclusions

- Treelets offers a new and useful methodology to the field of high - dimensional statistics
- TCB and TCBS look to improve this method by better utilizing the tree structure inherent in the algorithm

Conclusions

- Treelets offers a new and useful methodology to the field of high - dimensional statistics
- TCB and TCBS look to improve this method by better utilizing the tree structure inherent in the algorithm
- TCB and TCBS can be applied to a variety of problems - specifically the estimation of relatedness and heritability

Thanks for listening

Questions? Comments?