Extending Treelets for 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]

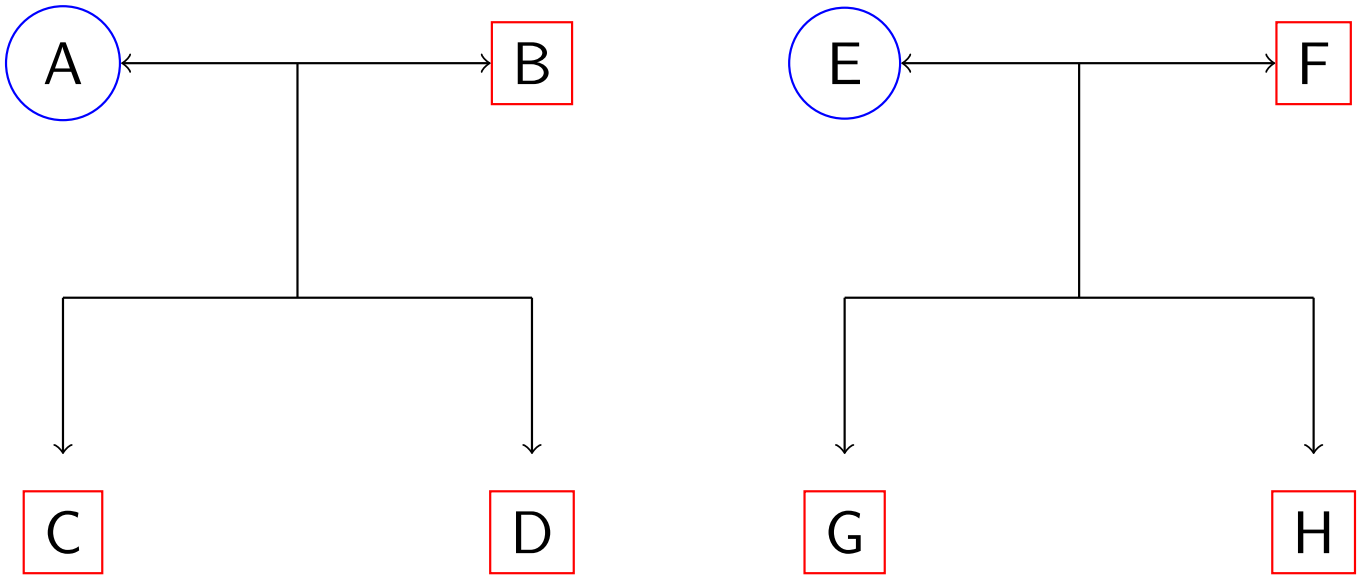
What does *that* mean?

[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.

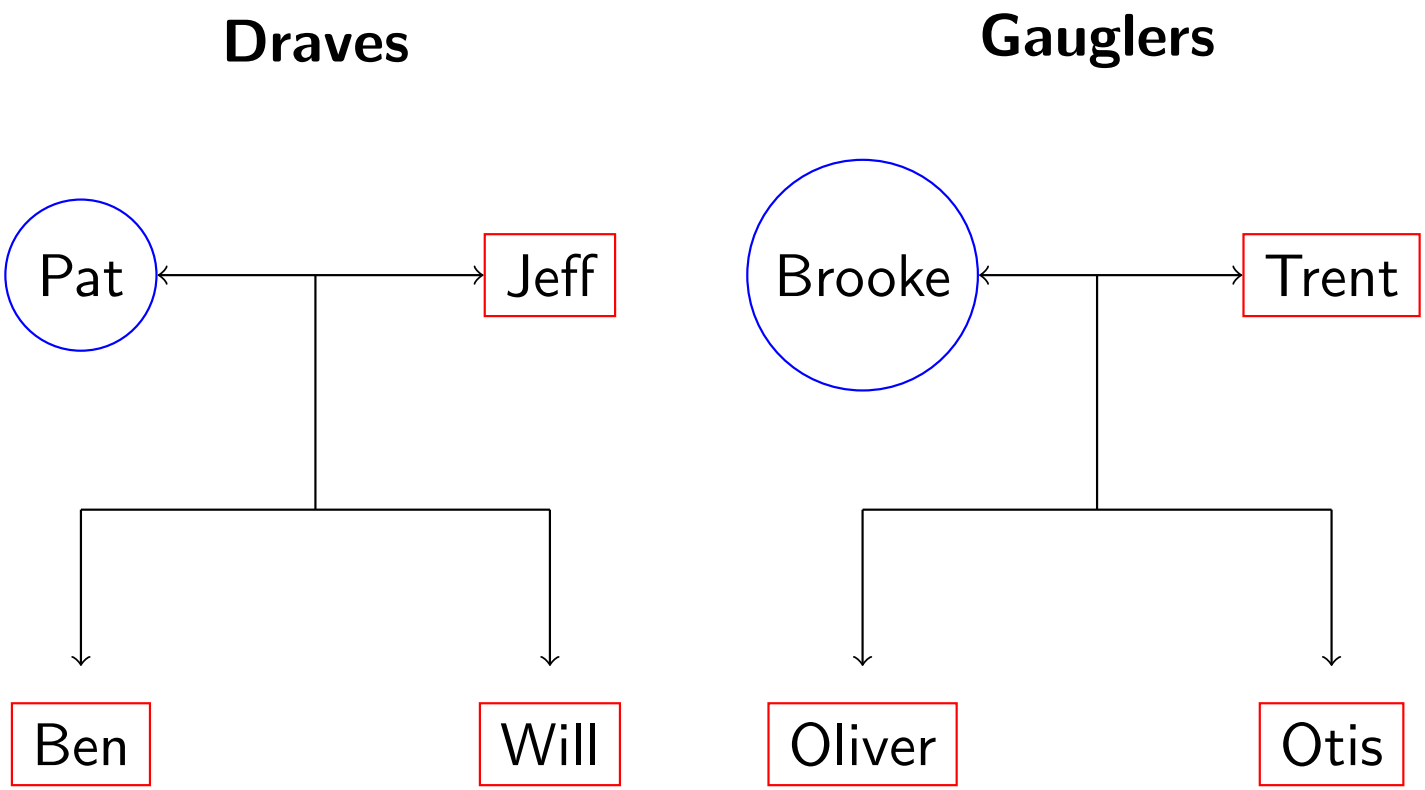
What is Treelets? - My Take

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

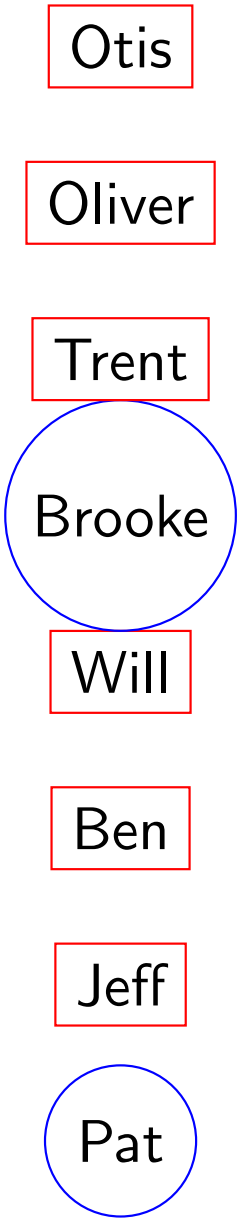
Pedigree Example



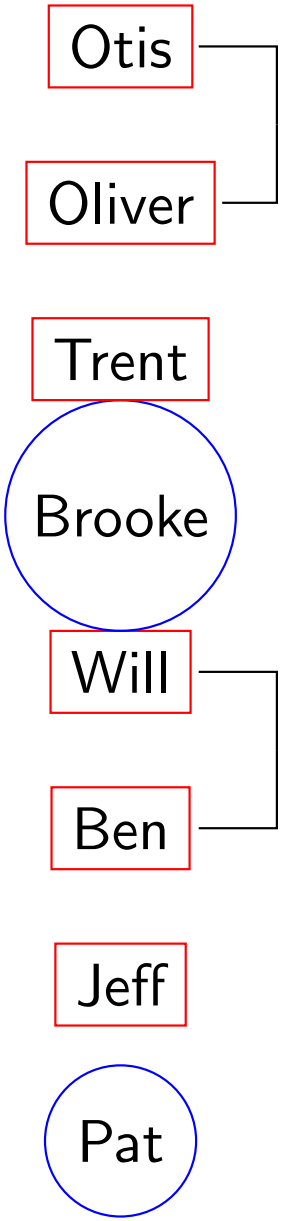
Pedigree Example



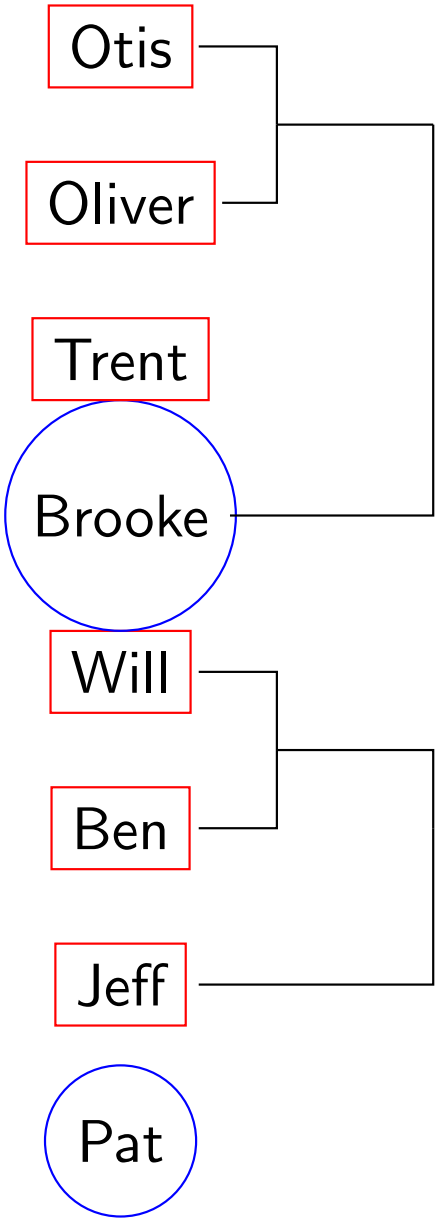
Pedigree Example



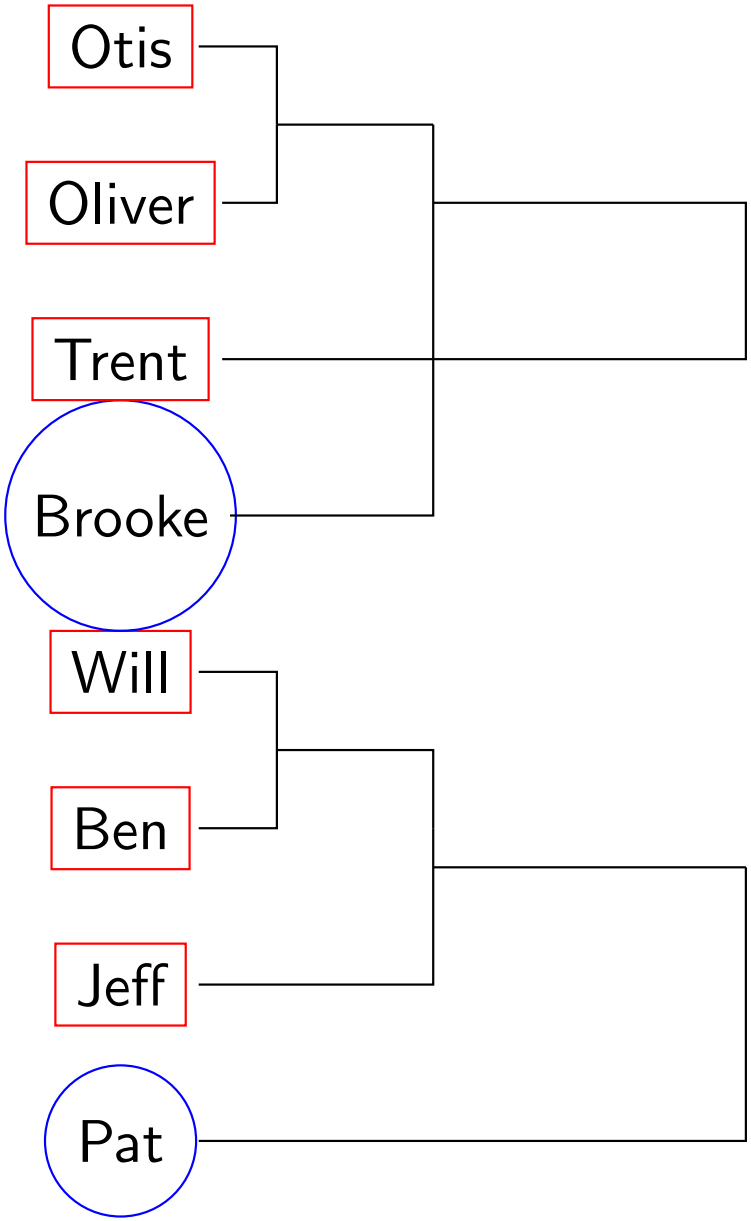
Pedigree Example



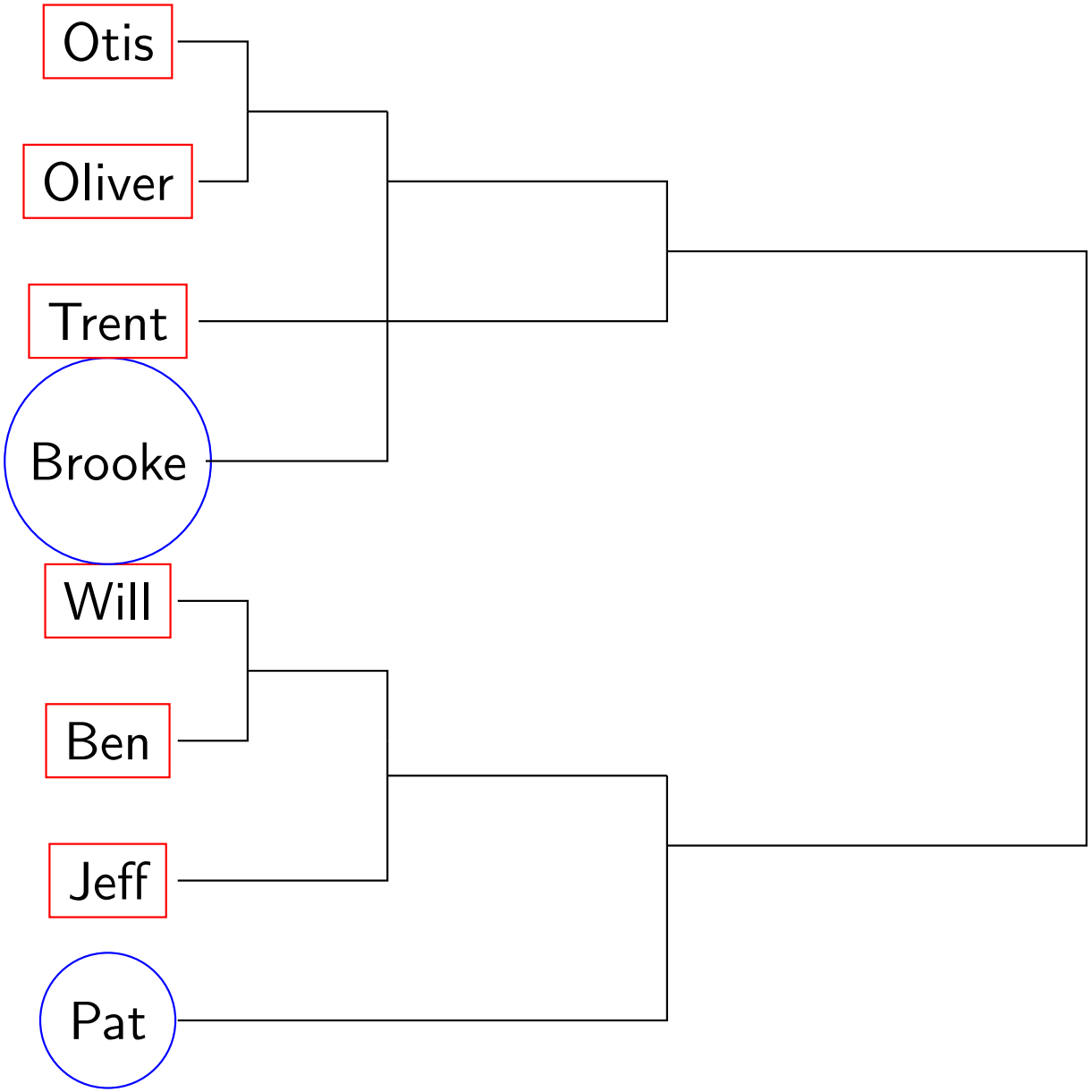
Pedigree Example



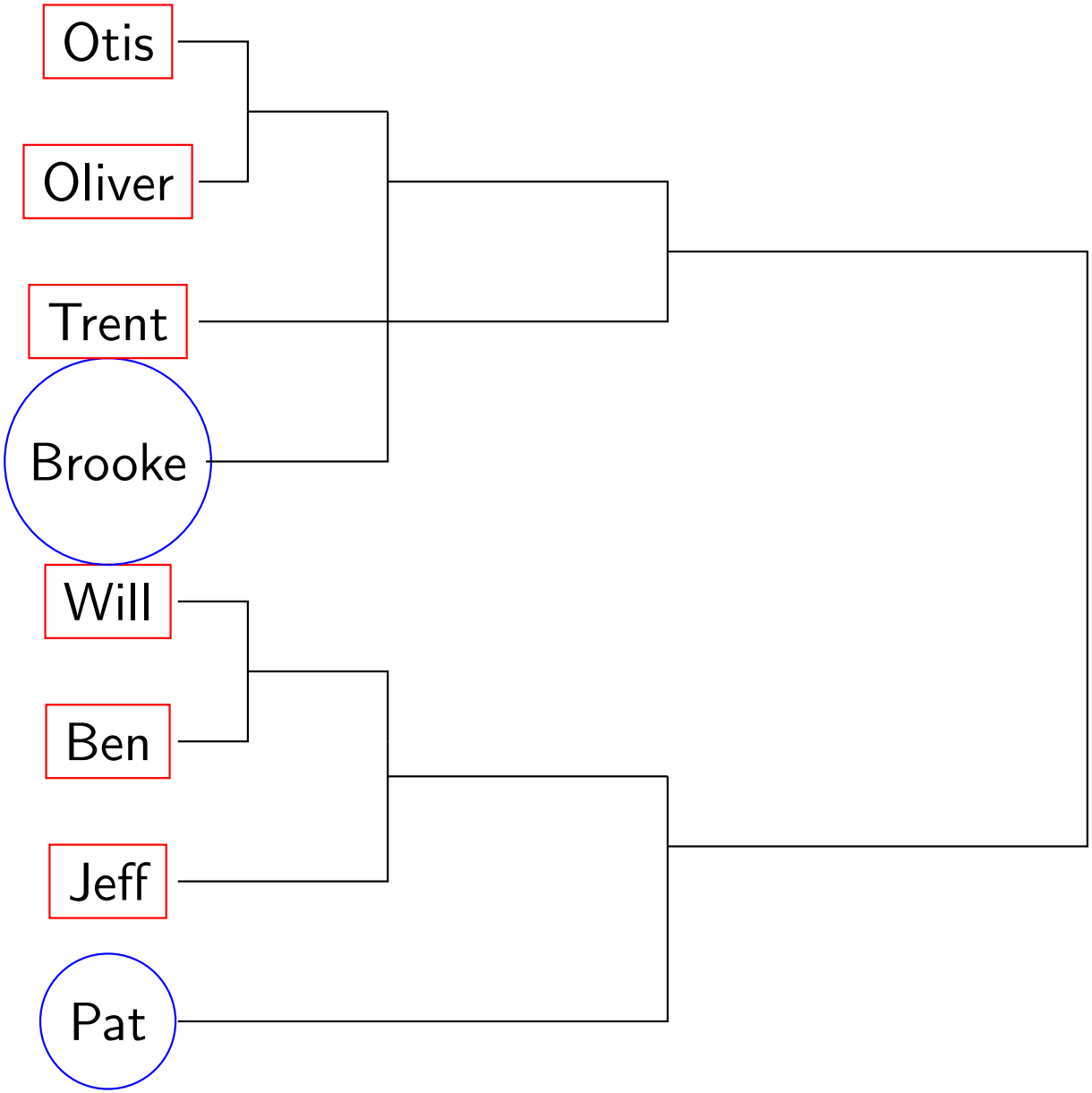
Pedigree Example



Pedigree Example



Pedigree Example



Treelets

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

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

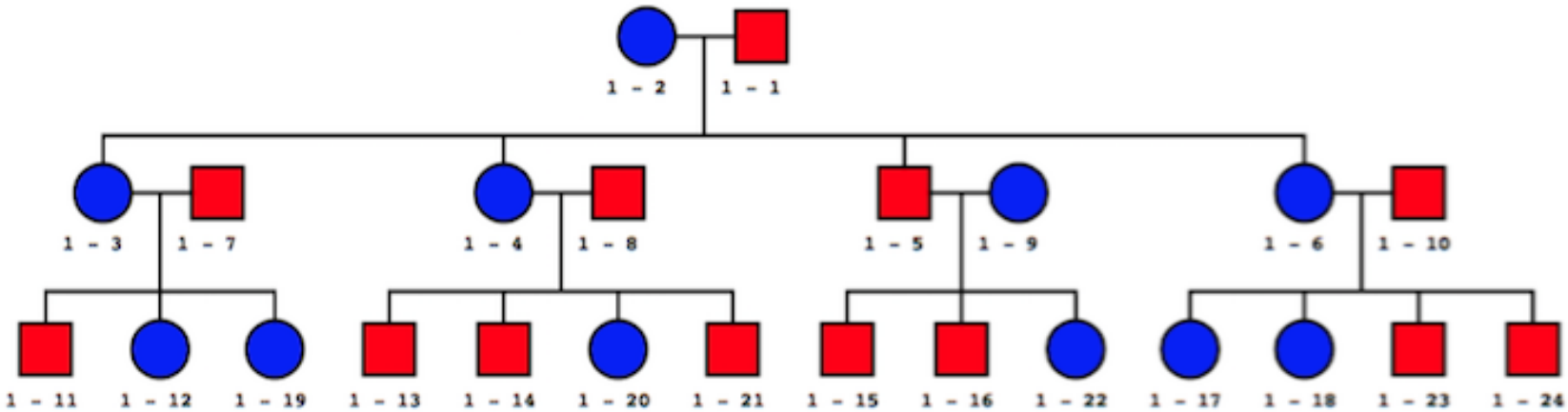
Estimating Relatedness- Theoretic

	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

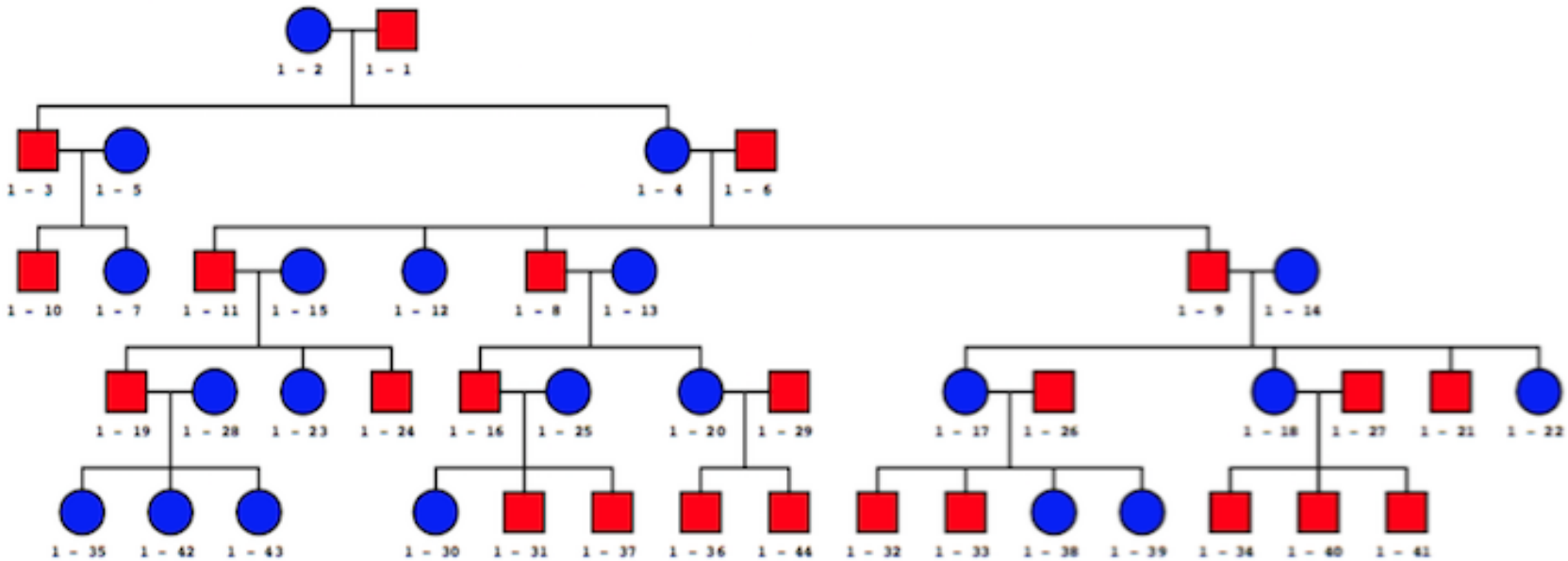
Estimating Relatedness - Sample

	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

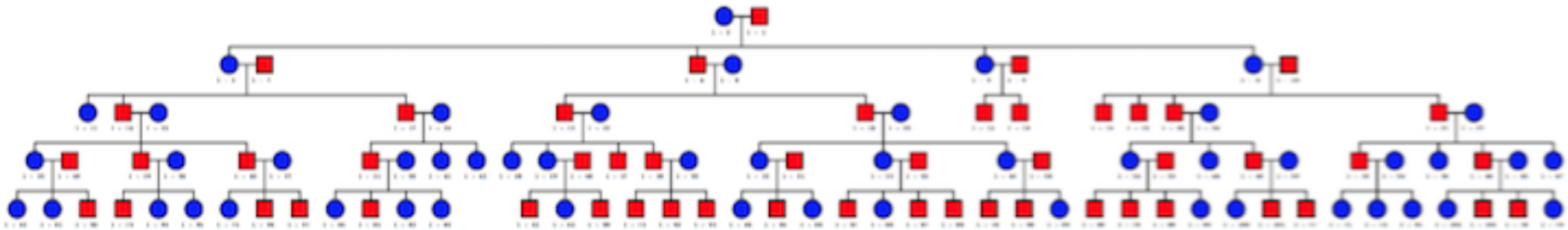
But what about...



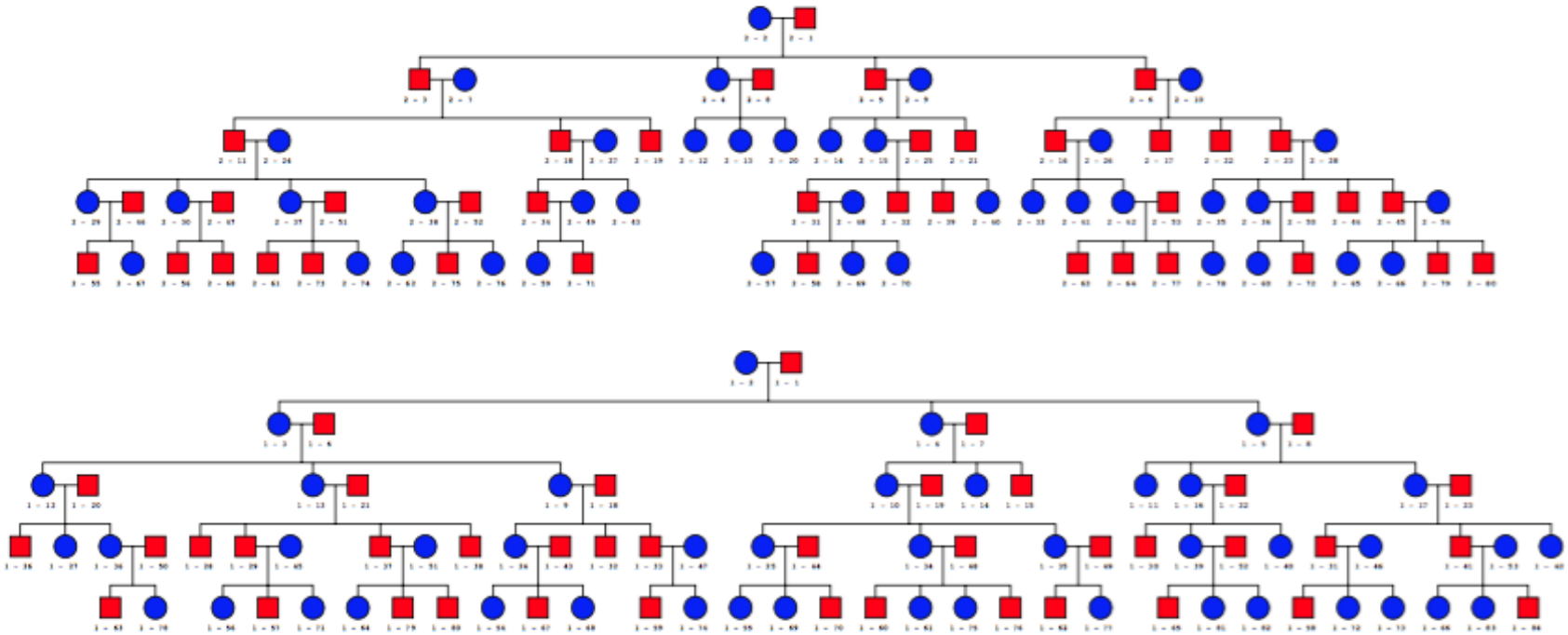
But what about...



But what about...



But what about...



Outline of Our Work

- Treelets works well with distant relatedness - what about closely related?

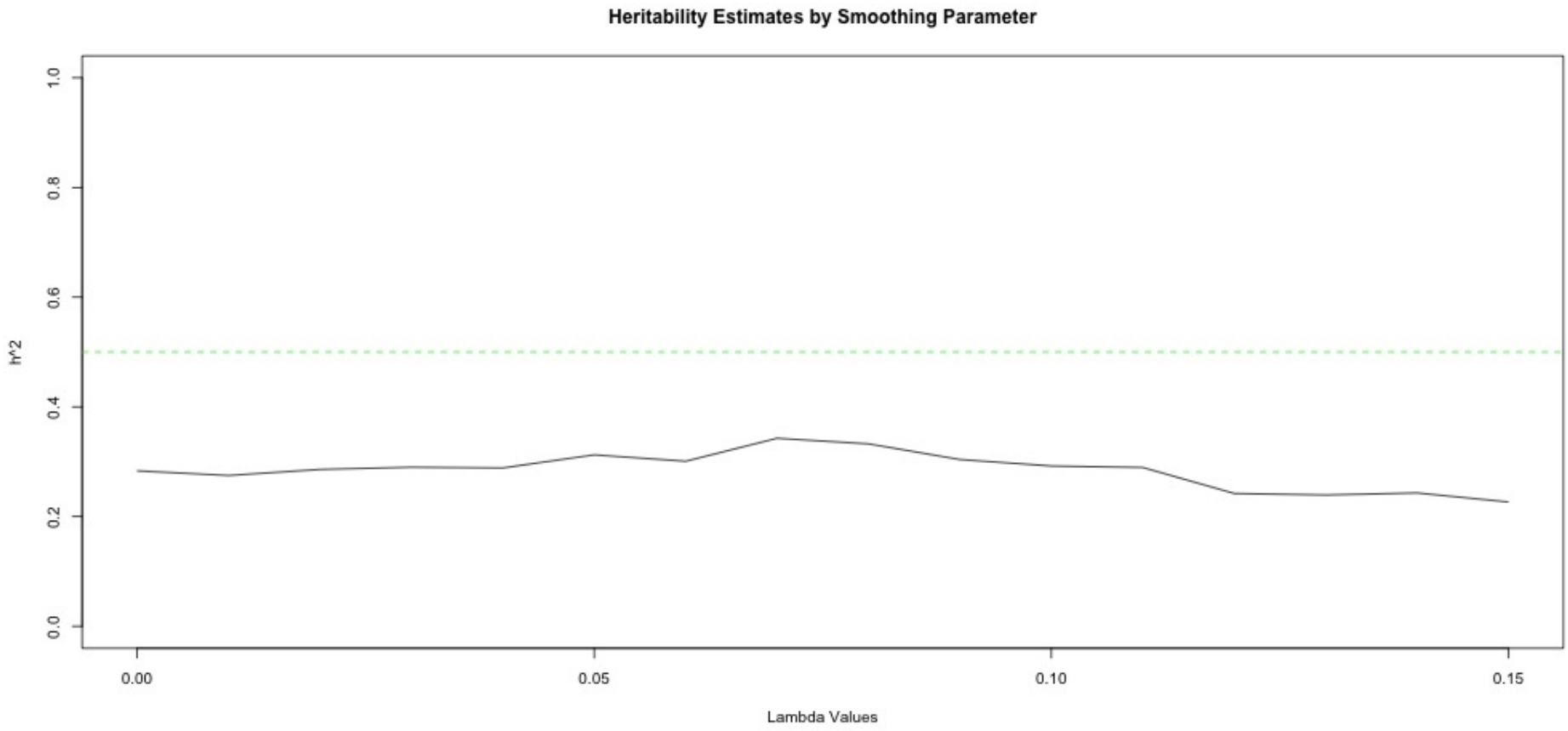
Outline of Our Work

- Treelets works well with distant relatedness - what about closely related?
- Generate several different pedigrees and phenotypes to estimate a known heritability

Outline of Our Work

- Treelets works well with distant relatedness - what about closely related?
- Generate several different pedigrees and phenotypes to estimate a known heritability
- Modify Treelets to better predict heritability in a sample given some measure of relatedness

Estimating Heritability - TCS



Conclusions

- Treelets is an exciting and powerful algorithm

Conclusions

- Treelets is an exciting and powerful algorithm
- ... that is in no way perfect

Conclusions

- Treelets is an exciting and powerful algorithm
- ... that is in no way perfect
- Our work looks to offer an alternative - perhaps more realistic - algorithm that handles highly correlated variables well

Thanks for listening

Questions? Comments? Jokes?