The Computer Says "Maybe"

Embracing Uncertainty in Computer-Assisted Textual Scholarship

Studia Stemmatologica IX, Bergen, Norway
Teemu Roos, University of Helsinki, 30.6.2022

91,4%

of statistics you'll ever see are made up

38%

published results* with uncertainty quantification

*) 5 out n=13 papers, including the phrase "computer-assisted" in the title that present quantitative results, published in DSH between 1/2010–11/2021

Uncertainty Quantification (UQ)

What Does that even Mean?

Uncertainty Quantification (UQ)

What Does that even Mean?

- Bald people sitting in front of me on the plane to Bergen: 6 out of 24 => 25%

Uncertainty Quantification (UQ)

What Does that even Mean?

- Bald people sitting in front of me on the plane to Bergen: 6 out of 24 => 25%

Uncertainty Quantification (UQ)

What Does that even Mean?

- Bald people sitting in front of me on the plane to Bergen: 6 out of 24 => 25%

Uncertainty Quantification (UQ)

What Does that even Mean?

- Bald people sitting in front of me on the plane to Bergen: 6 out of 24 => 25%
- Is this a reliable statistic?

Uncertainty Quantification (UQ)

What Does that even Mean?

- Bald people sitting in front of me on the plane to Bergen: 6 out of 24 => 25%
- Is this a reliable statistic?

- Let's simulate 24 new people with 25% probability of being bald:

Uncertainty Quantification (UQ)

What Does that even Mean?

- Bald people sitting in front of me on the plane to Bergen: 6 out of 24 => 25%
- Is this a reliable statistic?

- Let's simulate 24 new people with 25% probability of being bald:

```
B ~ ~ ~ ~ B ~ B ~ ~ ~ B ~ ~ ~ ~ ~ ~ B ~ B ~ ~ ~: 6
```

Uncertainty Quantification (UQ)

What Does that even Mean?

- Bald people sitting in front of me on the plane to Bergen: 6 out of 24 => 25%
- Is this a reliable statistic?

- Let's simulate 24 new people with 25% probability of being bald:

```
B ~ ~ ~ ~ B ~ B ~ ~ ~ B ~ ~ ~ ~ ~ ~ B ~ B ~ ~ ~: 6
~ ~ ~ ~ ~ ~ ~ ~ ~ B ~ ~ B B B ~ ~ B ~ ~ ~ ~ ~ ~: 5
```

Uncertainty Quantification (UQ)

What Does that even Mean?

- Bald people sitting in front of me on the plane to Bergen: 6 out of 24 => 25%
- Is this a reliable statistic?

- Let's simulate 24 new people with 25% probability of being bald:

```
B ~ ~ ~ ~ B ~ B ~ ~ ~ B ~ ~ ~ ~ ~ ~ B ~ B ~ ~ ~: 6
~ ~ ~ ~ ~ ~ ~ ~ ~ B ~ ~ B B B ~ ~ B ~ ~ ~ ~ ~ ~: 5
B B ~ B ~ ~ ~ B ~ B ~ ~ ~ ~ ~ B ~ ~ ~ ~ ~ ~ ~ ~: 6
```

Uncertainty Quantification (UQ)

What Does that even Mean?

- Bald people sitting in front of me on the plane to Bergen: 6 out of 24 => 25%
- Is this a reliable statistic?

- Let's simulate 24 new people with 25% probability of being bald:

```
B ~ ~ ~ ~ B ~ B ~ ~ ~ B ~ ~ ~ ~ ~ ~ B ~ B ~ ~ ~: 6
~ ~ ~ ~ ~ ~ ~ ~ ~ B ~ ~ B B B ~ ~ B ~ ~ ~ ~ ~ ~: 5
B B ~ B ~ ~ ~ B ~ B ~ ~ ~ ~ ~ B ~ ~ ~ ~ ~ ~ ~ ~: 6
B B ~ ~ ~ ~ B ~ ~ ~ ~ ~ ~ ~ ~ ~ B ~ B ~ ~ ~ B ~: 6
```

Uncertainty Quantification (UQ)

What Does that even Mean?

- Bald people sitting in front of me on the plane to Bergen: 6 out of 24 => 25%
- Is this a reliable statistic?

- Let's simulate 24 new people with 25% probability of being bald:

```
B ~ ~ ~ ~ B ~ B ~ ~ ~ B ~ ~ ~ ~ ~ ~ B ~ B ~ ~ ~: 6
~ ~ ~ ~ ~ ~ ~ ~ ~ ~ B ~ ~ B B B ~ ~ B ~ ~ ~ ~ ~ ~: 5
B B ~ B ~ ~ ~ B ~ B ~ ~ ~ ~ ~ B ~ ~ ~ ~ ~ ~ ~ ~: 6
B B ~ ~ ~ ~ B ~ ~ ~ ~ ~ ~ ~ ~ ~ B ~ B ~ ~ ~ B ~: 6
B B B ~ ~ ~ ~ B B B ~ ~ B ~ B ~ ~ ~ B ~ ~ ~ ~ ~: 9
```

Uncertainty Quantification (UQ)

What Does that even Mean?

- Bald people sitting in front of me on the plane to Bergen: 6 out of 24 => 25%
- Is this a reliable statistic?

- Let's simulate 24 new people with 25% probability of being bald:

```
B ~ ~ ~ ~ B ~ B ~ ~ ~ B ~ ~ ~ ~ ~ ~ B ~ B ~ ~ ~: 6
~ ~ ~ ~ ~ ~ ~ ~ ~ ~ B ~ ~ B B B ~ ~ B ~ ~ ~ ~ ~ ~: 5
B B ~ B ~ ~ ~ B ~ B ~ ~ ~ ~ ~ B ~ ~ ~ ~ ~ ~ ~ ~: 6
B B ~ ~ ~ ~ B ~ ~ ~ ~ ~ ~ ~ ~ ~ B ~ B ~ ~ ~ B ~: 6
B B B ~ ~ ~ ~ B B B ~ ~ B ~ B ~ ~ ~ B ~ ~ ~ ~ ~: 9
~ ~ ~ ~ ~ ~ ~ B B ~ ~ ~ B ~ ~ ~ ~ B ~ ~ ~ ~ ~ ~: 4
```

Uncertainty Quantification (UQ)

What Does that even Mean?

Uncertainty Quantification (UQ)

What Does that even Mean?

- Based on the data, we have **uncertainty** about the ratio of bald people flying to Bergen

Uncertainty Quantification (UQ)

What Does that even Mean?

- Based on the data, we have **uncertainty** about the ratio of bald people flying to Bergen

- **Uncertainty quantification** means characterizing the magnitude of this uncertainty:

Uncertainty Quantification (UQ)

What Does that even Mean?

- Based on the data, we have **uncertainty** about the ratio of bald people flying to Bergen

- **Uncertainty quantification** means characterizing the magnitude of this uncertainty:
  - *confidence intervals*: baldness rate = 25% [10–47%]

Uncertainty Quantification (UQ)

What Does that even Mean?

- Based on the data, we have **uncertainty** about the ratio of bald people flying to Bergen

- **Uncertainty quantification** means characterizing the magnitude of this uncertainty:
  - *confidence intervals*: baldness rate = 25% [10–47%]
  - *p-values*: is the baldness rate different on flights to Bergen and to Helsinki?

Uncertainty Quantification (UQ)

What Does that even Mean?

- Based on the data, we have **uncertainty** about the ratio of bald people flying to Bergen

- **Uncertainty quantification** means characterizing the magnitude of this uncertainty:
  - *confidence intervals*: baldness rate = 25% [10–47%]
  - *p-values*: is the baldness rate different on flights to Bergen and to Helsinki?
  - *Bayesian posteriors*: baldness rate has a Beta(7, 19) distribution*

*) when the prior is uniform

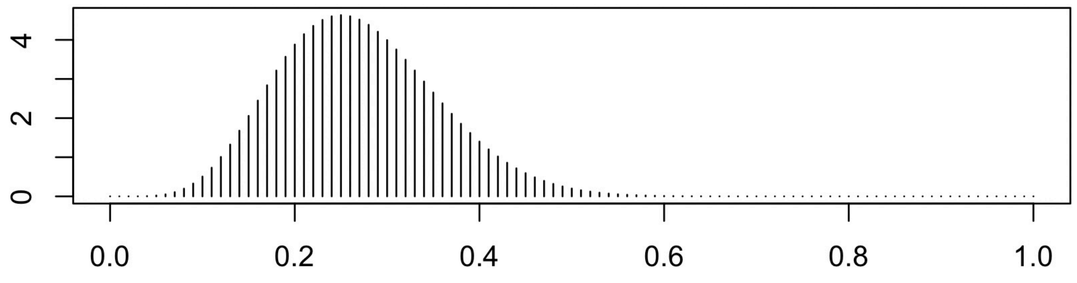
Uncertainty Quantification (UQ)

What Does that even Mean?

- Based on the data, we have **uncertainty** about the ratio of bald people flying to Bergen

- **Uncertainty quantification** means characterizing the magnitude of this uncertainty:
  - *confidence intervals*: baldness rate = 25% [10–47%]
  - *p-values*: is the baldness rate different on flights to Bergen and to Helsinki?
  - *Bayesian posteriors*: baldness rate has a Beta(7, 19) distribution*

*) when the prior is uniform



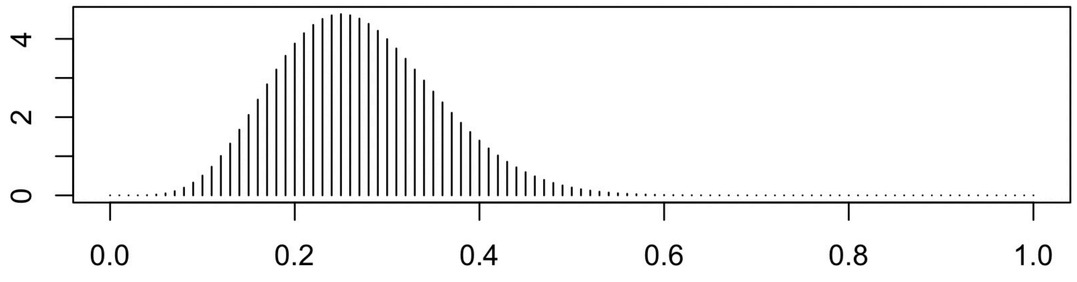
Uncertainty Quantification (UQ)

What Does that even Mean?

- Based on the data, we have **uncertainty** about the ratio of bald people flying to Bergen

- **Uncertainty quantification** means characterizing the magnitude of this uncertainty:
  - *confidence intervals*: baldness rate = 25% [10–47%]
  - *p-values*: is the baldness rate different on flights to Bergen and to Helsinki?
  - *Bayesian posteriors*: baldness rate has a Beta(7, 19) distribution*
  - *bootstrap*: bootstrap distribution of the baldness rate

*) when the prior is uniform



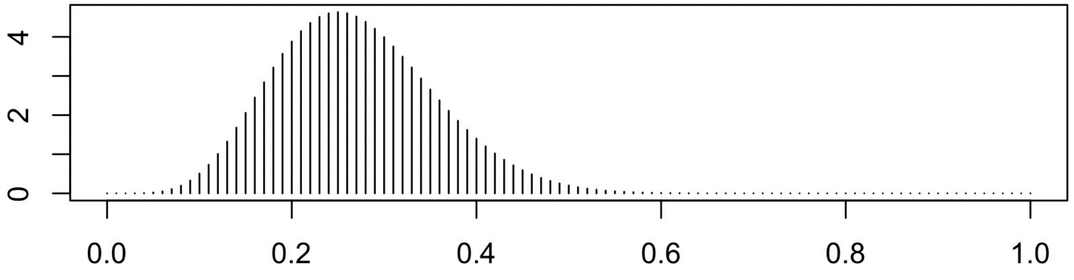
Uncertainty Quantification (UQ)

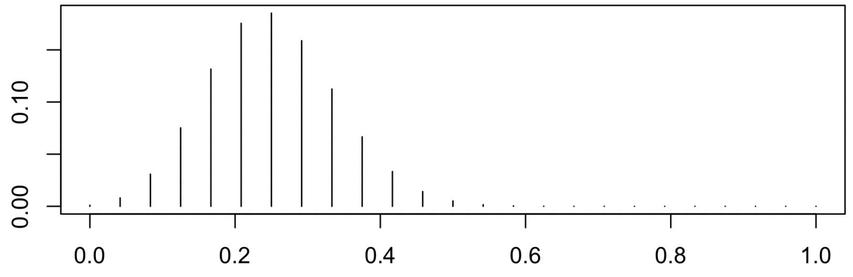
What Does that even Mean?

- Based on the data, we have **uncertainty** about the ratio of bald people flying to Bergen

- **Uncertainty quantification** means characterizing the magnitude of this uncertainty:
  - *confidence intervals*: baldness rate = 25% [10–47%]
  - *p-values*: is the baldness rate different on flights to Bergen and to Helsinki?
  - *Bayesian posteriors*: baldness rate has a Beta(7, 19) distribution*
  - *bootstrap*: bootstrap distribution of the baldness rate

*) when the prior is uniform





Why Bother?

Why is UQ Important?

- When you get a result from a fancy computer program ("Artificial Intelligence, ooh-la-la!"), and you haven't a clue how it works, you tend to just take the result at face value
  - you can always come up with a *post hoc* explanation

- Since the data and the methods always have limitations, the next time you study the same subject, you'll get a somewhat different method => you must be able to judge which result to take more seriously

Solutions

How to do UQ at Home

- Use confidence intervals for quantities

- **Bootstrap**: Randomize the data and repeat the analysis many times
  - each of the repetitions produces a slightly different outcome
  - bootstrap values represent the support for a split in the tree
  - consensus tree represents as many of the best supported splits as possible

- **Bayesian methods** offer basically the same benefits with a more solid justification (but they require that you choose priors)

Caveat: Bootstrap is not Enough

"But there are also unknown unknowns—the ones we don't know we don't know."

D. Rumsfeld, 2002

Caveat: Bootstrap is not Enough

"But there are also unknown unknowns—the ones we don't know we don't know."

D. Rumsfeld, 2002

- Even if uncertainty due to limited data is taken into account (e.g. bootstrap), we may go wrong

Caveat: Bootstrap is not Enough

"But there are also unknown unknowns—the ones we don't know we don't know."

D. Rumsfeld, 2002

- Even if uncertainty due to limited data is taken into account (e.g. bootstrap), we may go wrong
- When the assumed model is wrong, we get systematically biased results

Caveat: Bootstrap is not Enough

"But there are also unknown unknowns—the ones we don't know we don't know."

D. Rumsfeld, 2002

- Even if uncertainty due to limited data is taken into account (e.g. bootstrap), we may go wrong
- When the assumed model is wrong, we get systematically biased results
  - tree?

Caveat: Bootstrap is not Enough

"But there are also unknown unknowns—the ones we don't know we don't know."

D. Rumsfeld, 2002

- Even if uncertainty due to limited data is taken into account (e.g. bootstrap), we may go wrong

- When the assumed model is wrong, we get systematically biased results
  - tree?
  - independent characters?

Caveat: Bootstrap is not Enough

"But there are also unknown unknowns—the ones we don't know we don't know."

D. Rumsfeld, 2002

- Even if uncertainty due to limited data is taken into account (e.g. bootstrap), we may go wrong

- When the assumed model is wrong, we get systematically biased results
  - tree?
  - independent characters?
  - random errors?

Caveat: Bootstrap is not Enough

"But there are also unknown unknowns—the ones we don't know we don't know."

D. Rumsfeld, 2002

- Even if uncertainty due to limited data is taken into account (e.g. bootstrap), we may go wrong

- When the assumed model is wrong, we get systematically biased results
  - tree?
  - independent characters?
  - random errors?
  - ...

Caveat: Bootstrap is not Enough

"But there are also unknown unknowns—the ones we don't know we don't know."

D. Rumsfeld, 2002

- Even if uncertainty due to limited data is taken into account (e.g. bootstrap), we may go wrong

- When the assumed model is wrong, we get systematically biased results
  - tree?
  - independent characters?
  - random errors?
  - ...

Caveat: Bootstrap is not Enough

"But there are also unknown unknowns—the ones we don't know we don't know."

D. Rumsfeld, 2002

- Even if uncertainty due to limited data is taken into account (e.g. bootstrap), we may go wrong

- When the assumed model is wrong, we get systematically biased results
  - tree?
  - independent characters?
  - random errors?
  - ...

- For this there is no (complete and automatic) remedy => **always doubt**

Another Idea

Case: Little Red Riding Hood (Tehrani, Nguyen & Roos, 2015)

- Network analysis: PhyloDAG method produces a slightly different tree each time
- We display 12 of them to give an idea of the variability

Another Idea

Case: Little Red Riding Hood (Tehrani, Nguyen & Roos, 2015)

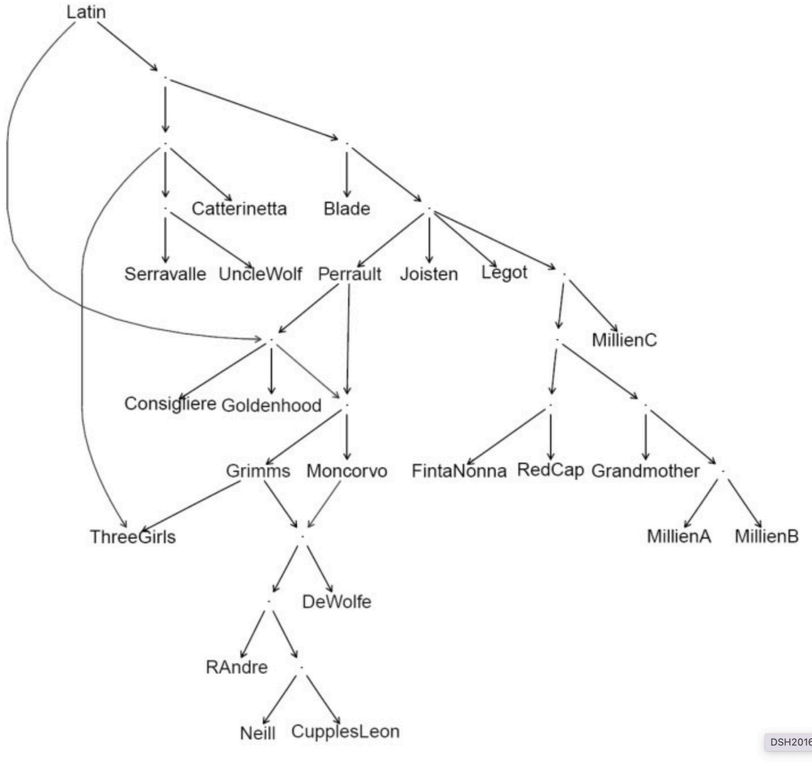
- Network analysis: PhyloDAG method produces a slightly different tree each time
- We display 12 of them to give an idea of the variability

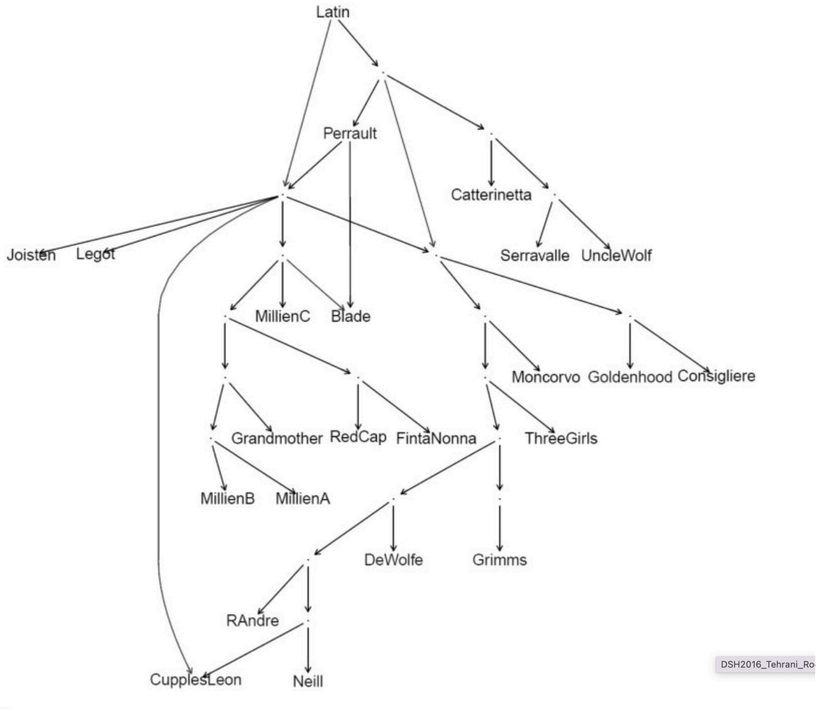
DSH2016_Tehrani_Roos

Another Idea

Case: Little Red Riding Hood (Tehrani, Nguyen & Roos, 2015)

- Network analysis: PhyloDAG method produces a slightly different tree each time
- We display 12 of them to give an idea of the variability

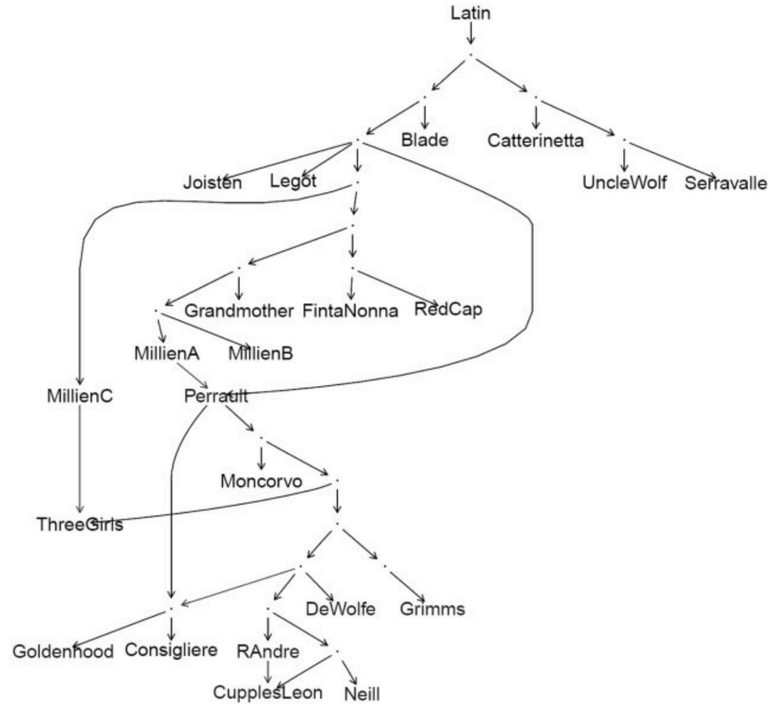




Another Idea

Case: Little Red Riding Hood (Tehrani, Nguyen & Roos, 2015)

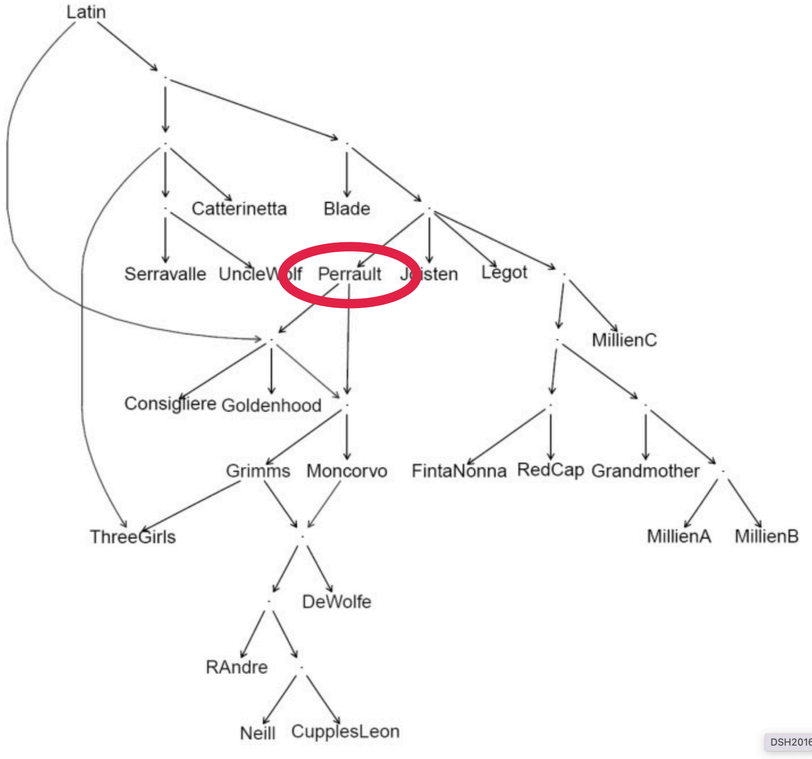
- Network analysis: PhyloDAG method produces a slightly different tree each time
- We display 12 of them to give an idea of the variability

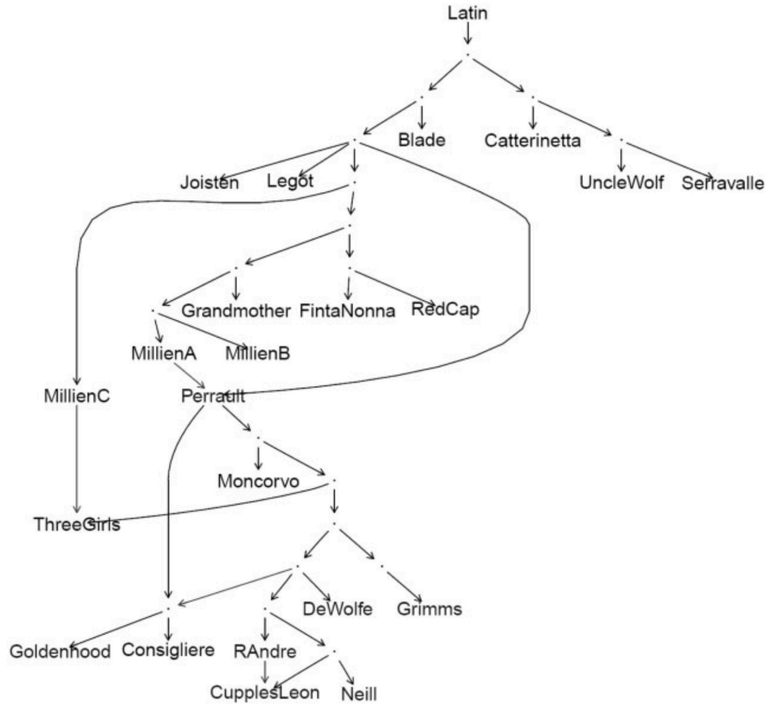


Another Idea

Case: Little Red Riding Hood (Tehrani, Nguyen & Roos, 2015)

- Network analysis: PhyloDAG method produces a slightly different tree each time
- We display 12 of them to give an idea of the variability

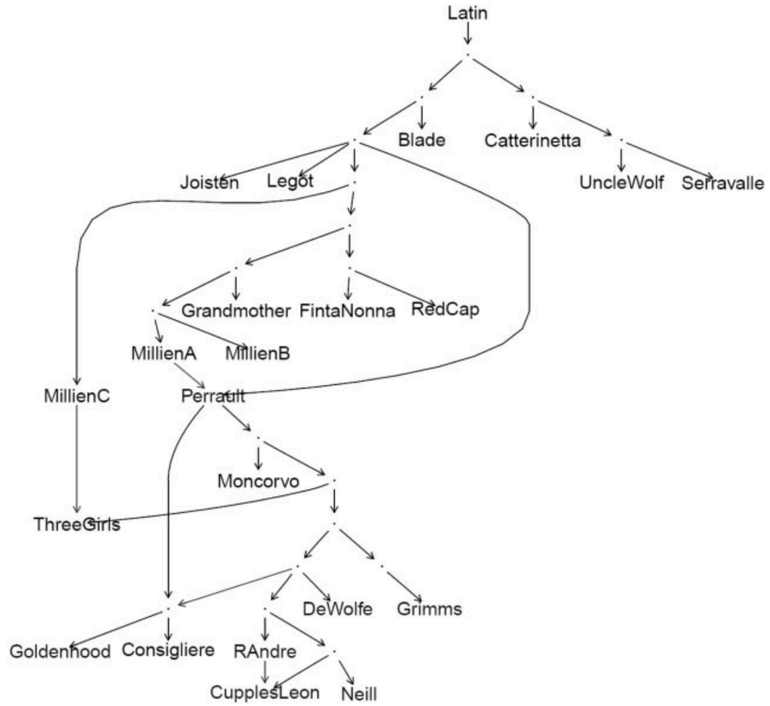




Another Idea

Case: Little Red Riding Hood (Tehrani, Nguyen & Roos, 2015)

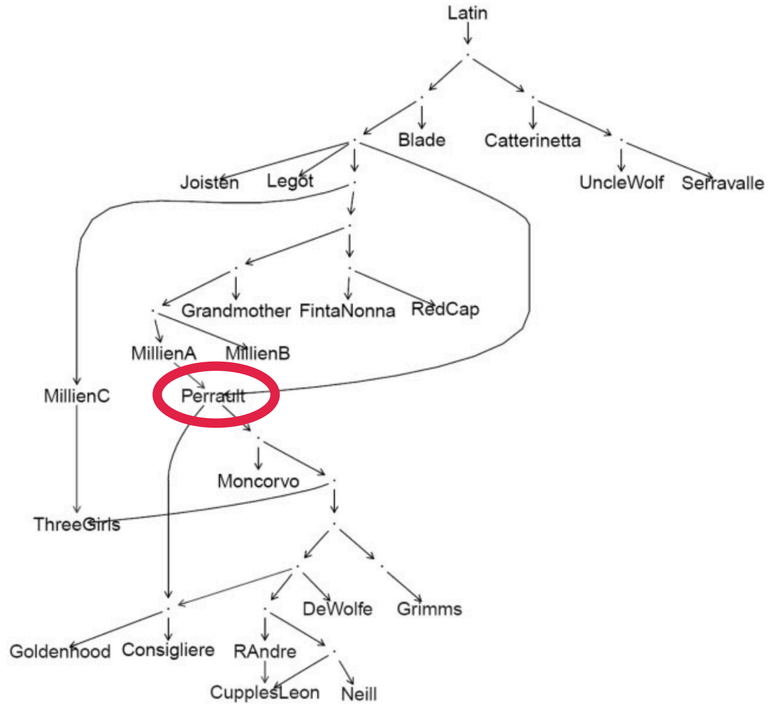
- Network analysis: PhyloDAG method produces a slightly different tree each time
- We display 12 of them to give an idea of the variability



Another Idea

Case: Little Red Riding Hood (Tehrani, Nguyen & Roos, 2015)

- Network analysis: PhyloDAG method produces a slightly different tree each time
- We display 12 of them to give an idea of the variability



Another Idea

Case: Little Red Riding Hood (Tehrani, Nguyen & Roos, 2015)

- Parametric bootstrap (see Huelsenbeck and Crandall (1997)) method used to compare all hypotheses against each other

Another Idea

Case: Little Red Riding Hood (Tehrani, Nguyen & Roos, 2015)

- Parametric bootstrap (see Huelsenbeck and Crandall (1997)) method used to compare all hypotheses against each other

Table 1 Statistical hypothesis test results (parametric bootstrap)

Null	Alternative hypothesis											
Hypothesis	**tree**	***a***	***b***	***c***	***d***	***e***	***f***	***g***	***h***	***i***	***j***	***k***
tree		*	*	*	*	+	*	*	*	.	*	.
a	.		*	*	*	*	*	*	*	.	*	*
b	.	.		+	+	+	+	+	+	.	*	+
c	.	.	+		.	.	.	+	.	.	.	.
d	.	+	*	*		+	.	*	.	.	*	+
e	+	*	*	+	*		*	*	+	.	*	*
f	+	*	*	*	.	*		*	+	.	+	.
g	+	.	+	.	*	*	*		.	.	+	.
h	*	*	*	*	*	*	*	*		*	*	*
i	*	*	*	*	*	*	*	*	*		*	*
j	*	*	*	*	*	*	*	*	.	.		*
K	*	*	*	*	*	*	*	*	.	.	*	

Rows: null hypothesis. Columns: alternative hypothesis. 'tree': parsimony tree. '·': not rejected. '+': rejected at significance level 0.05. '*': rejected at significance level 0.01.

Another Idea

Case: Little Red Riding Hood (Tehrani, Nguyen & Roos, 2015)

- Parametric bootstrap (see Huelsenbeck and Crandall (1997)) method used to compare all hypotheses against each other

Table 1 Statistical hypothesis test results (parametric bootstrap)

Null	Alternative hypothesis											
Hypothesis	**tree**	***a***	***b***	***c***	***d***	***e***	***f***	***g***	***h***	***i***	***j***	***k***
tree		*	*	*	*	+	*	*	*	.	*	.
a	.		*	*	*	*	*	*	*	.	*	*
b	.	.		+	+	+	+	+	+	.	*	+
c	.	.	+		.	.	.	+	.	.	.	.
d	.	+	*	*		+	.	*	.	.	*	+
e	+	*	*	+	*		*	*	+	.	*	*
f	+	*	*	*	.	*		*	+	.	+	.
g	+	.	+	.	*	*	*		.	.	+	.
h	*	*	*	*	*	*	*	*		*	*	*
i	*	*	*	*	*	*	*	*	*		*	*
j	*	*	*	*	*	*	*	*	.	.		*
K	*	*	*	*	*	*	*	*	.	.	*	

Rows: null hypothesis. Columns: alternative hypothesis. 'tree': parsimony tree. '·': not rejected. '+': rejected at significance level 0.05. '*': rejected at significance level 0.01.

Another Idea

Case: Little Red Riding Hood (Tehrani, Nguyen & Roos, 201

"All models are false, but some are useful"

- Parametric bootstrap (see Huelsenbeck and Crandall (1997)) method used to compare all hypotheses against each other

Table 1 Statistical hypothesis test results (parametric bootstrap)

Null	Alternative hypothesis											
Hypothesis	**tree**	***a***	***b***	***c***	***d***	***e***	***f***	***g***	***h***	***i***	***j***	***k***
tree		*	*	*	*	+	*	*	*	.	*	.
a	.		*	*	*	*	*	*	*	.	*	*
b	.	.		+	+	+	+	+	+	.	*	+
c	.	.	+		.	.	.	+	.	.	.	.
d	.	+	*	*		+	.	*	.	.	*	+
e	+	*	*	+	*		*	*	+	.	*	*
f	+	*	*	*	.	*		*	+	.	+	.
g	+	.	+	.	*	*	*		.	.	+	.
h	*	*	*	*	*	*	*	*		*	*	*
i	*	*	*	*	*	*	*	*	*		*	*
j	*	*	*	*	*	*	*	*	.	.		*
K	*	*	*	*	*	*	*	*	.	.	*	

Rows: null hypothesis. Columns: alternative hypothesis. 'tree': parsimony tree. '·': not rejected. '+': rejected at significance level 0.05. '*': rejected at significance level 0.01.

Table 1

				Computational methods used	Results presented as charts or diagrams Y/N	Is there uncertainty quantification Y/N	UQ methods used	
Rachel McCarthy, James O'Sullivan	Who wrote Wuthering Heights?	Volume 36, Issue 2, June 2021, Pages 383–391	https://doi.org/10.1093/llc/fqaa031	Stylometry, hierarchical clustering	Y	Y	statistical confidence values, multiple results from different methods	
Sangeetha Kutty, Richi Nayak, Paul Turnbull, Ron Chernich, Gavin Kennedy, Kerry Raymond	PaperMiner—a real-time spatiotemporal visualization for newspaper articles	Volume 35, Issue 1, April 2020, Pages 83–100	https://doi.org/10.1093/llc/fqy084	OCR, named entity recognition, clustering	Y	N		
Yu-Fang Ho, Jane Lugea, Dan McIntyre, Zhijie Xu, Jing Wang	Text-world annotation and visualization for crime narrative reconstruction	Volume 34, Issue 2, June 2019, Pages 310–334	https://doi.org/10.1093/llc/fqy044	Expert systems	Y	N		
Sabine Lang, Björn Ommer	Attesting similarity: Supporting the organization and study of art image collections with computer vision	Volume 33, Issue 4, December 2018, Pages 845–856	https://doi.org/10.1093/llc/fqy006	Image similarity search	N	Y	Precision-recall	
Rui Hu, Carlos Pallán Gayol, Jean-Marc Odobez, Daniel Gatica-Perez	Analyzing and visualizing ancient Maya hieroglyphics using shape: From computer vision to Digital Humanities	Volume 32, Issue suppl_2, December 2017, Pages ii179–ii194	https://doi.org/10.1093/llc/fqx028	Image similarity search	N	Y	Precision-recall	
Melissa Terras, James Baker, James Hetherington, David Beavan, Martin Zaltz Austwick, Anne Welsh, Helen O'Neill, Will Finley, Oliver Duke-Williams, Adam Farquhar	Enabling complex analysis of large-scale digital collections: humanities research, high-performance computing, and transforming access to British Library digital collections	Volume 33, Issue 2, June 2018, Pages 456–466	https://doi.org/10.1093/llc/fqx020	Database infrastructure	N/A	N/A		
Alejandro H Toselli, Luis A Leiva, Isabel Bordes-Cabrera, Celio Hernández-Tornero, Vicent	Transcribing a 17th-century botanical manuscript: Longitudinal evaluation of	Volume 33, Issue 1, April 2018, Pages 173–202	https://doi.org/10.1093/llc/fqw064	OCR, manual transcription tools	N/A	N/A		

Emiliano Giovannetti, Davide Albanesi, Andrea Bellandi, Giulia Benotto	Traduco: A collaborative web-based CAT environment for the interpretation and translation of texts	Volume 32, Issue suppl_1, April 2017, Pages i47–i62	https://doi.org/10.1093/llc/fqw054	Machine translation	N/A	N/A		
Eythan Levy, Frédéric Pluquet	Computer experiments on the Khirbet Qeiyafa ostracon	Volume 32, Issue 4, December 2017, Pages 816–836	https://doi.org/10.1093/llc/fqw028	Dictionary search	N/A	N/A		
Marina Buzzoni, Eugenio Burgio, Martina Modena, Samuela Simion	Open versus closed recensions (Pasquali): Pros and cons of some methods for computer-assisted stemmatology	Volume 31, Issue 3, September 2016, Pages 652–669	https://doi.org/10.1093/llc/fqw014	Stemmatic analysis (parsimony, neighbour joining, neighbornet, RHM, Semstem)	Y	N*	*) Except for using multiple methods for the same data	
Barbara Bordalejo	The genealogy of texts: Manuscript traditions and textual traditions	Volume 31, Issue 3, September 2016, Pages 563–577	https://doi.org/10.1093/llc/fqv038	Stemmatic analysis (parsimony, neighbornet)	Y	N*	*) Except for using multiple methods for the same data	
Jamshid Tehrani, Quan Nguyen, Teemu Roos	Oral fairy tale or literary fake? Investigating the origins of Little Red Riding Hood using phylogenetic network analysis	Volume 31, Issue 3, September 2016, Pages 611–636	https://doi.org/10.1093/llc/fqv016	Stemmatic analysis (PhyloDag)	Y	Y	statistical confidence values, multiple results from the same method	
Marko Halonen	Computer-assisted stemmatology in studying Paulus Juusten's 16th-century chronicle Catalogus et ordinaria successio Episcoporum Finlandensium	Volume 31, Issue 3, September 2016, Pages 578–593	https://doi.org/10.1093/llc/fqv004	Stemmatic analysis (parsimony, RHM, neighbornet)	Y	Y	statistical confidence values (bootstrap)	
John Lee, Ying Cheuk Hui, Yin Hei Kong	Knowledge-rich, computer-assisted composition of Chinese couplets	Volume 31, Issue 1, April 2016, Pages 152–163	https://doi.org/10.1093/llc/fqu052	Automatic text generation	N	N		
David-Antoine Williams	Method as tautology in the digital humanities	Volume 30, Issue 2, June 2015, Pages 280–293	https://doi.org/10.1093/llc/fqt068	N/A	N/A	N/A		
Jennifer von	The MayaArch3D	Volume 28, Issue 4,	https://doi.org/	N/A	N/A	N/A		

Jennifer von Schwerin, Heather Richards-Rissetto, Fabio Remondino, Giorgio Agugiaro, Gabrio Girardi	The MayaArch3D project: A 3D WebGIS for analyzing ancient architecture and landscapes	Volume 28, Issue 4, December 2013, Pages 736–753	https://doi.org/10.1093/llc/fqt059	N/A	N/A	N/A		
Alexis Antonia, Hugh Craig, Jack Elliott	Language chunking, data sparseness, and the value of a long marker list: explorations with word n-grams and authorial attribution	Volume 29, Issue 2, June 2014, Pages 147–163	https://doi.org/10.1093/llc/fqt028	N-gram analysis	N	N		
Gábor Mihály Tóth	The computer-assisted analysis of a medieval commonplace book and diary (MS Zibaldone Quaresimale by Giovanni Rucellai)	Volume 28, Issue 3, September 2013, Pages 432–443	https://doi.org/10.1093/llc/fqs055	Expert systems	Y	N		
Maxime B. Sainte-Marie, Jean-Guy Meunier, Nicolas Payette, Jean-François Chartier	The concept of evolution in the Origin of Species: a computer-assisted analysis	Volume 26, Issue 3, September 2011, Pages 329–334	https://doi.org/10.1093/llc/fqr019	Clustering	Y	N		
William A. Kretzschmar, Jr, William Gray Potter	Library collaboration with large digital humanities projects	Volume 25, Issue 4, December 2010, Pages 439–445	https://doi.org/10.1093/llc/fqq022	N/A	N/A	N/A		