

# Semantic Decomposition on Folk Tale Tradition Networks:

Toward a Method for Understanding Tradition  
Groups at Scale

Timothy R. Tangherlini  
*tango@berkeley.edu*

# Collaborative work

- James Abello (DIMACS, CS, Rutgers Univ)
- Peter M. Broadwell (Stanford Univ)



# Indices, Classifiers and Metadata

- The ATU tale type index (Uther 2004), and the Motif Index of Folk-Literature (Thompson 1955-58) have formed the basis for a large number of comparative folktale studies focused on individual tale types or complexes of related tales
  - Folklore Fellows Communications the main outlet for these studies
- They are structured as network “trees” (connected, planar and acyclic), with several levels of categories (max 4 for the ATU index, and 7 for the MI index)
- They are largely “attached” as meta data to tales in printed collections or archives
  - Serve largely as a finding aid



# Some basic aspects

- The ATU tale type index (Uther 2004), and the Motif Index of Folk-Literature (Thompson 1955-58) have formed the basis for a large number of comparative folktale studies focused on individual tale types or complexes of related tales.
- They are structured as network “trees” (connected, planar and acyclic), with several levels of categories (max 4 for the ATU index, and 7 for the MI index)



# Criticisms

- When Aarne designed his original *Verzeichnis* (1910), he had two goals in mind. First, he intended to [develop] a system to arrange manuscript collections of stories recorded from oral tradition in the Finnish archives; and second, he wanted to enable scholars of the geographical-historical school to quickly locate the materials they needed in order to compile tale type monographs... When Thompson added to the title of his translation of Aarne's catalogue the term 'classification', confusion resulted, as a claim was forced on Aarne's catalogue for which it was not intended.

– (Jason 2006)



# Criticisms

- Challenges to
  - applicability of one or both of these indices to a particular collection
  - or to a linguistically or culturally defined tradition.
- Criticism of the construction of the categories, their labels, and the accompanying summary descriptions.
- Concern related to the integration of the two indices with each other, despite internal references that can be used to create links between tale types, between motifs, and between tale types and motifs
- Should we abandon these labels?



# Not all is lost...

- “It has been found that a specific list of types may represent the oral-literary repertoire of a culture-society” (Jason 2006: 185)
- A catalog of the tales recorded from a particular tradition group should reflect, on some level, the aggregate characteristics of that tradition.
- A simple list of the specific types of tales and their frequency in a particular culture and society likely provides only the vaguest outline of the contours of that tradition
- To derive a fuller and, to borrow a term from Clifford Geertz (1973), “thicker” description of that folktale tradition, one should work and think macroscopically



# The Folklore Macroscope

- Macroscopes provide a 'vision of the whole,' helping us 'synthesize' the related elements and detect patterns, trends, and outliers while granting access to myriad details. Rather than make things larger or smaller, macroscopes let us observe what is at once too great, slow, or complex for the human eye and mind to notice and comprehend

-Katy Börner (2011)



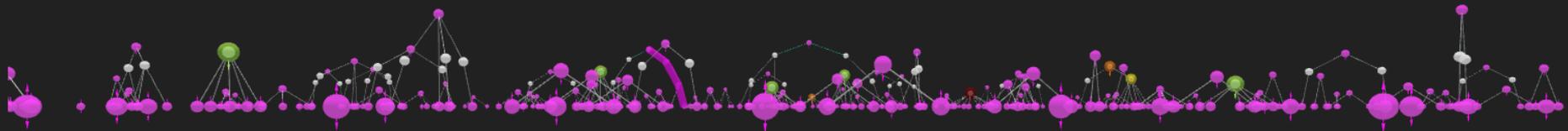
# Thicker analysis

- Consider the tale tellers
- Where they lived
- The historical context in which they lived
- The connections they drew between tales, both within a single tale (captured through multiple indexing of a tale) and within their broader repertoire
- Holbek's recognition of both the frequent combination of types and the creativity of the individual story "craftsman" in shaping the tradition (1987).



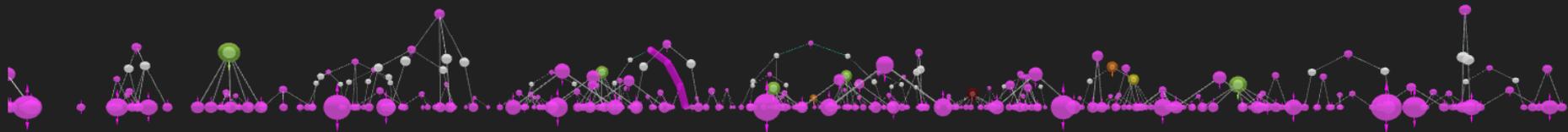
# Existing challenges

- Many large collections of folktales are at such a large scale that it is difficult to consider the corpus as a whole
  - Part of the motivation for so many studies to focus on an individual teller, a small community, or a single or small number of tale types
- How can I work at very large scale (thousands of tales) with minimal information about the contents of stories beyond their classification or indexing?
- This work focuses on addressing that common challenge in folklore archives and collections



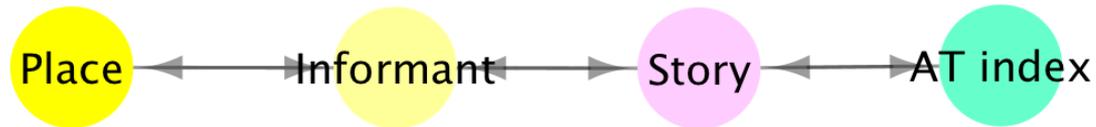
# Creating a network of networks

- Danish fairy tale corpus collected by ETK consists of:
  - 2606 tales
  - collected from 738 individuals
  - 491 unique places
- Each one indexed to the ATU index, with multiple tale type numbers and MI numbers
- ATU index includes references to:
  - Other ATU numbers
  - MI numbers
- Allowing the creation of a network that is not simply a tree

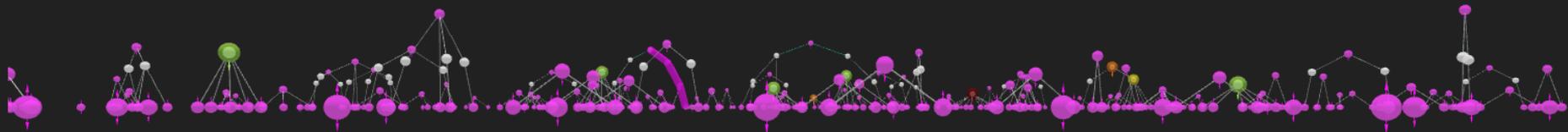


# Creating a network of networks

- We embed the collection into the ATU-MI graph

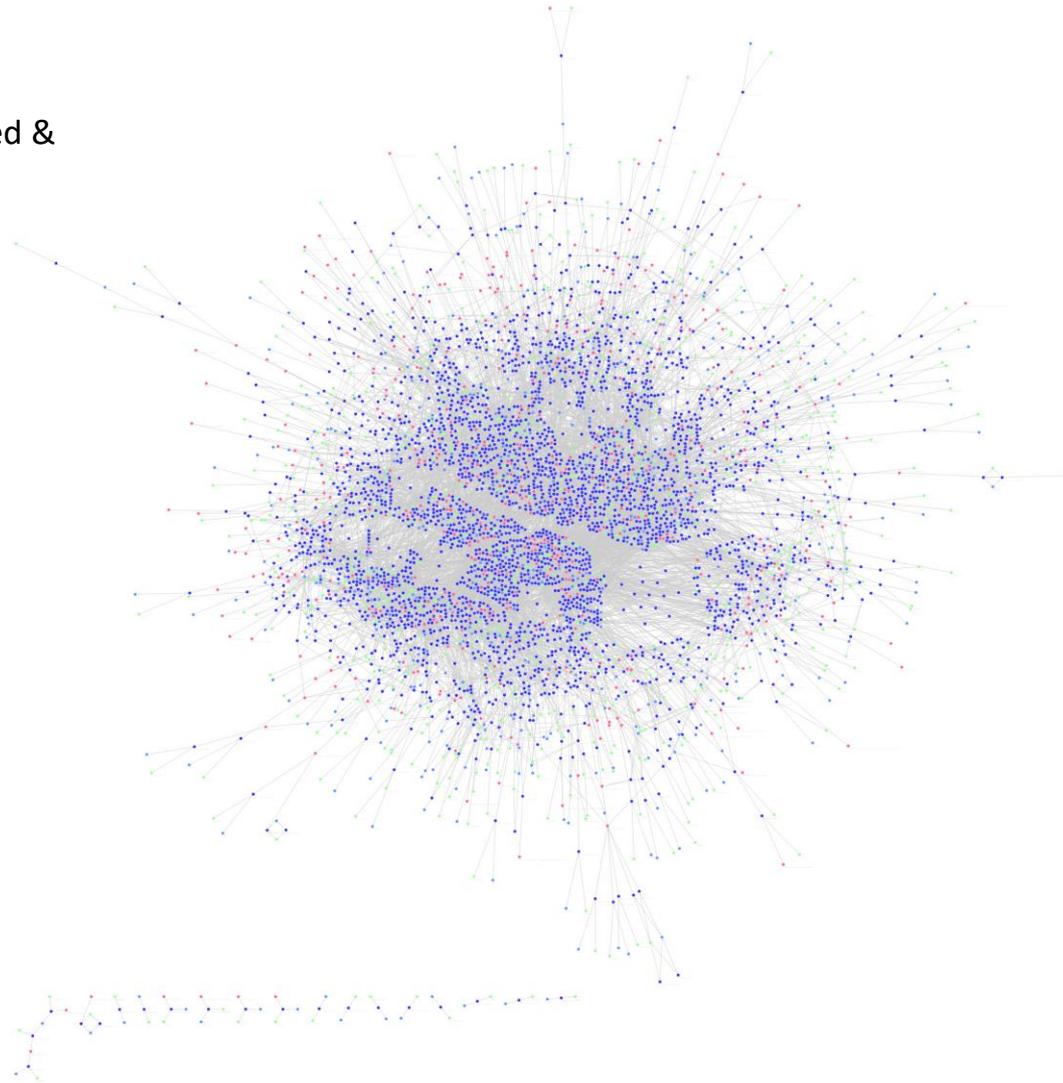


- And then prune the graph so that the unused parts of the ATU-MI trees do not crowd the network



# Creating a network of networks

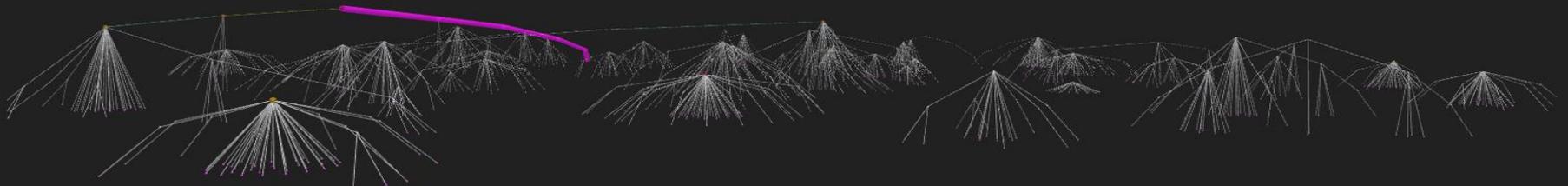
Embedded &  
Pruned



# Is this useful?

(or, “why would anyone do this?”)

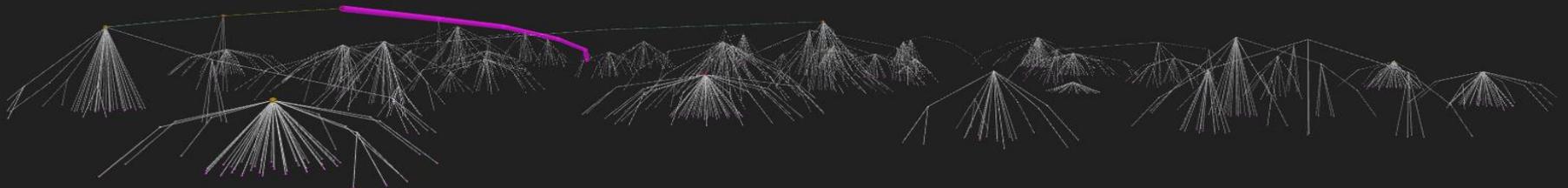
- Will an embedding of the Danish folktale corpus in the ATU-MI network:
  - allow for a characterization of themes and topics in the overall corpus not based solely on frequency?
  - help in the discovery of important tale tellers?
  - help in the discovery of places of significant influence or innovation?
- The results should not be trivial or obvious
  - ie not easily discoverable by other methods
  - stand up to scrutiny by domain experts)



# Is this useful? (part 2)

(or, “how could these discoveries help me?”)

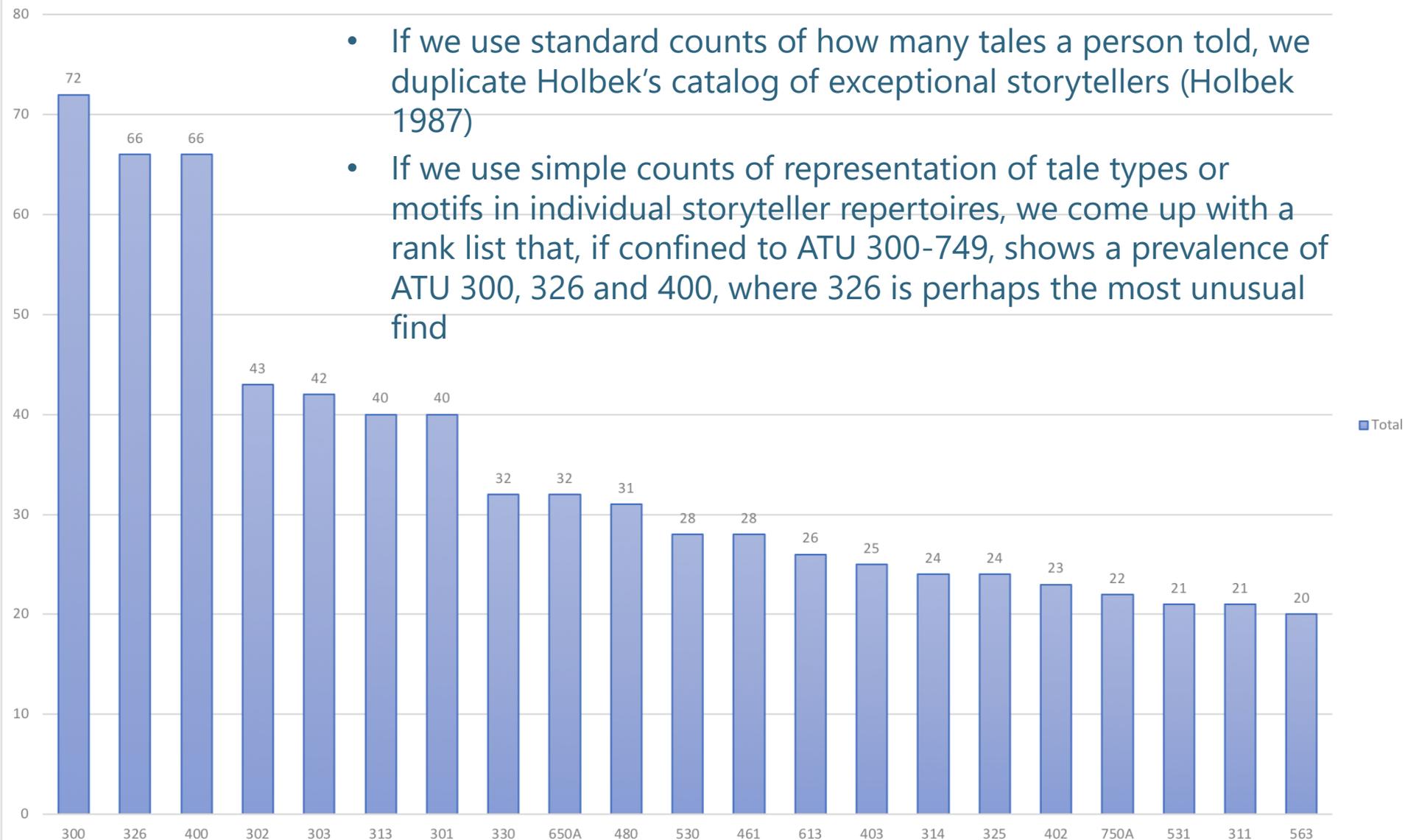
- In the context of existing scholarship on Danish fairy tale and folk tale tradition
  - Such as Bengt Holbek, *Interpretation of Fairy Tales*
- In the context of historically situated phenomena
  - Change in agricultural economic organization
  - Shifting ideas about the family
  - Ongoing changes in cultural ideology (norms, beliefs, values)



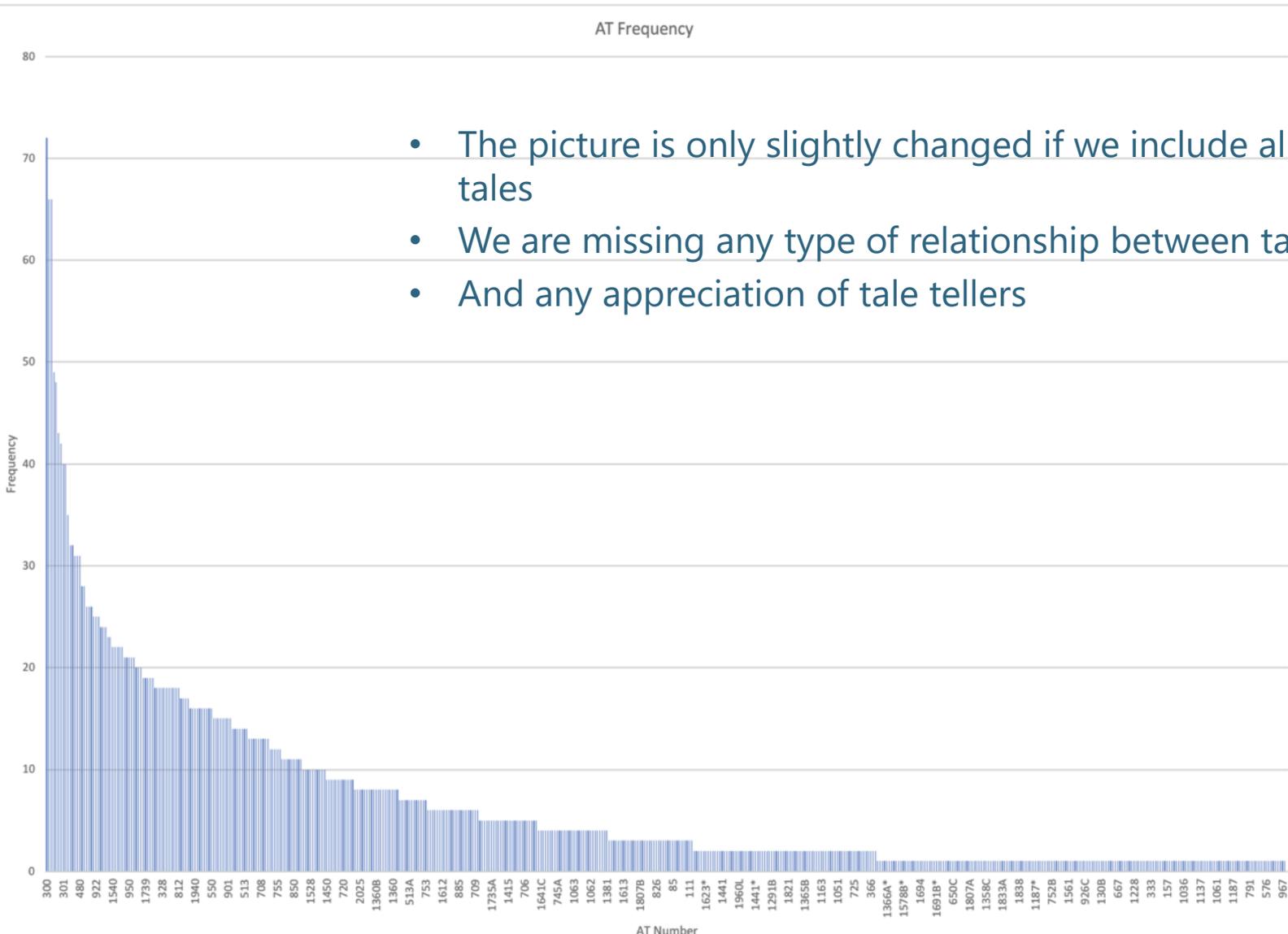
# Standard methods

ATU 300-749

- If we use standard counts of how many tales a person told, we duplicate Holbek's catalog of exceptional storytellers (Holbek 1987)
- If we use simple counts of representation of tale types or motifs in individual storyteller repertoires, we come up with a rank list that, if confined to ATU 300-749, shows a prevalence of ATU 300, 326 and 400, where 326 is perhaps the most unusual find



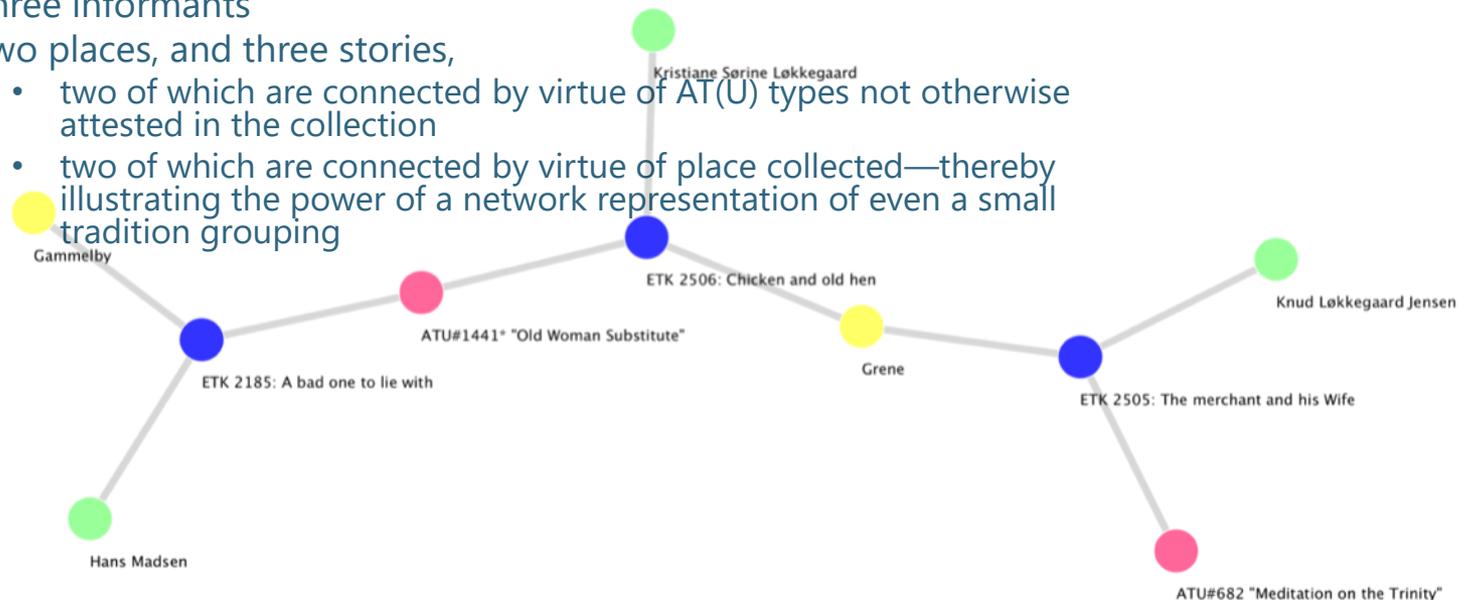
# Standard methods



- The picture is only slightly changed if we include all indexed tales
- We are missing any type of relationship between tales
- And any appreciation of tale tellers

# Standard (network) methods

- Standard network methods aim to solve that problem by generating a network that incorporates people, places, stories, as well as the indexing meta-data
- In the graph, we find:
  - a small number of disconnected components (16) that reveal the following consistent features:
    - (i) five consist of an informant, a story, a tale type, and a place of collection—in other words, storytellers who told stories indexed to tales not otherwise attested in the Danish tradition (see below);
    - (ii) nine represent tales that have neither AT nor MI indexing and are the only tale told by the informant;
    - (iii) two represent small groups of storytellers and their stories.
- In one of these small clusters, we find
  - three informants
  - two places, and three stories,
    - two of which are connected by virtue of AT(U) types not otherwise attested in the collection
    - two of which are connected by virtue of place collected—thereby illustrating the power of a network representation of even a small tradition grouping



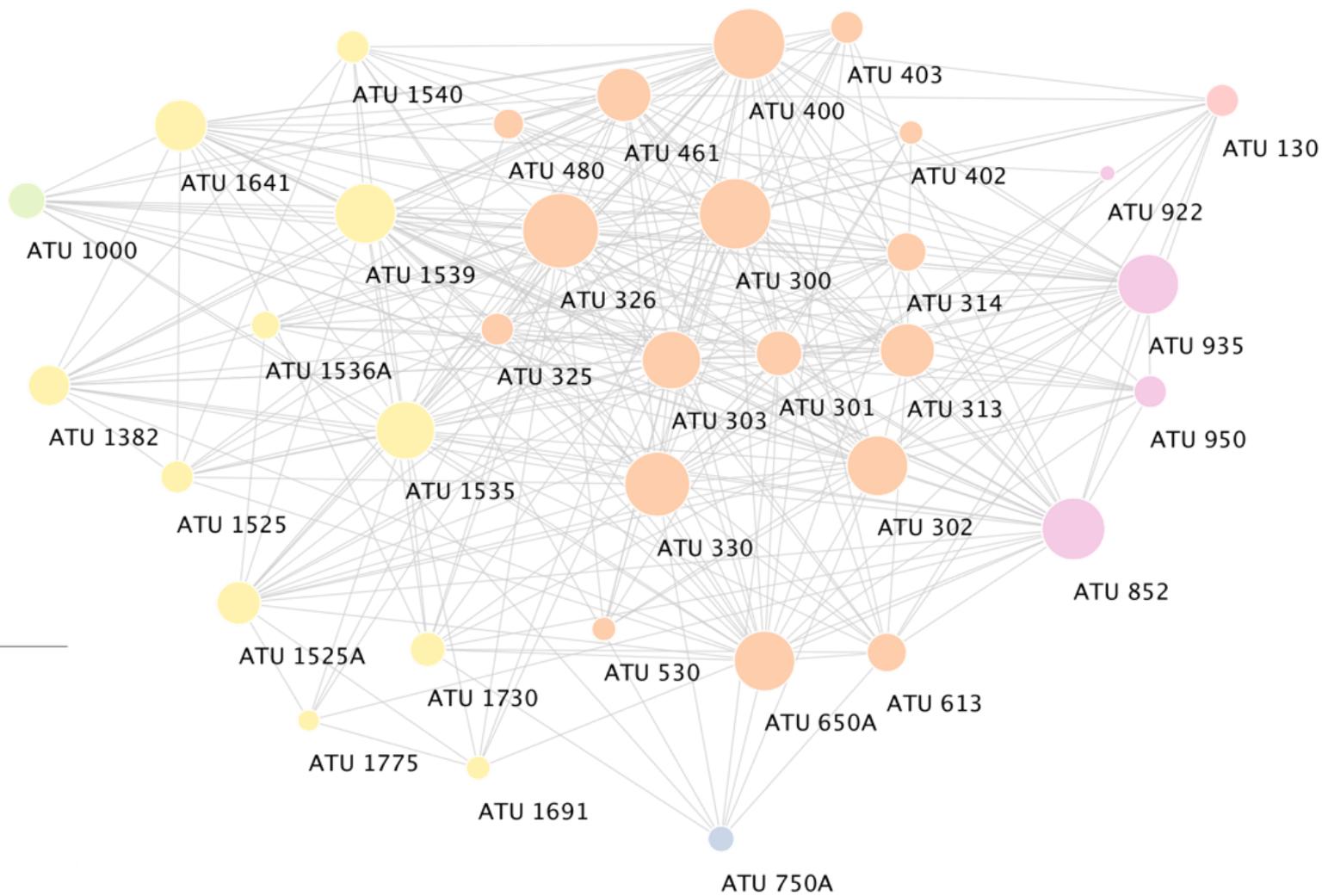
# Standard network methods

- The two attestations of ATU 1441 and the single attestation of ATU 682 in the entire Danish fairy tale corpus were collected in the 20<sup>th</sup> century (1901 for ETK 2185 and 1905 for ETK 2505 and 2506),
- The two tales collected in Grene were collected from a farmer and his aged mother who told no other fairy tales. Indeed, the remainder of their repertoires consisted of a pair of ghost stories and some riddles.
- The only story Hans Madsen told was the ATU 1441 fairy tale, and it is the only recording in the broader Tang Kristensen collection attributed to him.



# Projections of multimodal graphs

- For an ATU-ATU projection, where the edges are places (which in effect aggregates informants, as informants live in places) and the nodes are ATU numbers (which in effect aggregates individual story variants), we find
  - 35 tale types (out of a total of 528) with a very high degree (or number of connections),
  - these 35 types, connected by place, cover 144 of the approximately 500 unique places in the collection
- This ATU-ATU projection on places has the benefit of aggregating all of the places represented in the collection, and thus can be used to capture those tales that are broadly represented across the tradition area.



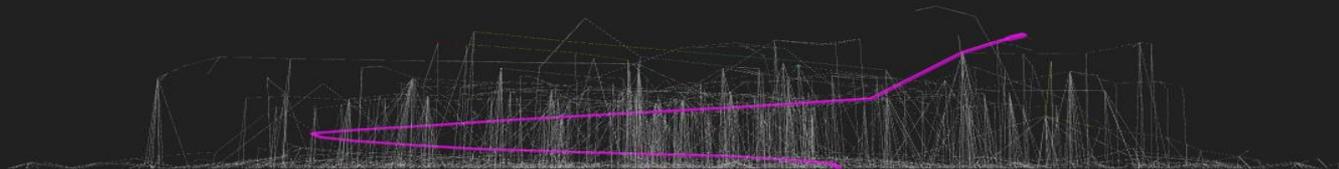
Node Fill Color	Hierarchy
<span style="color: #f08080;">■</span>	Animal Tales
<span style="color: #ff8c00;">■</span>	Tales of Magic
<span style="color: #6495ed;">■</span>	Religious Tales
<span style="color: #dda0dd;">■</span>	Realistic Tales
<span style="color: #90ee90;">■</span>	Tales of the Stupid Ogre
<span style="color: #ffff00;">■</span>	Anecdotes and Jokes

# Network topology

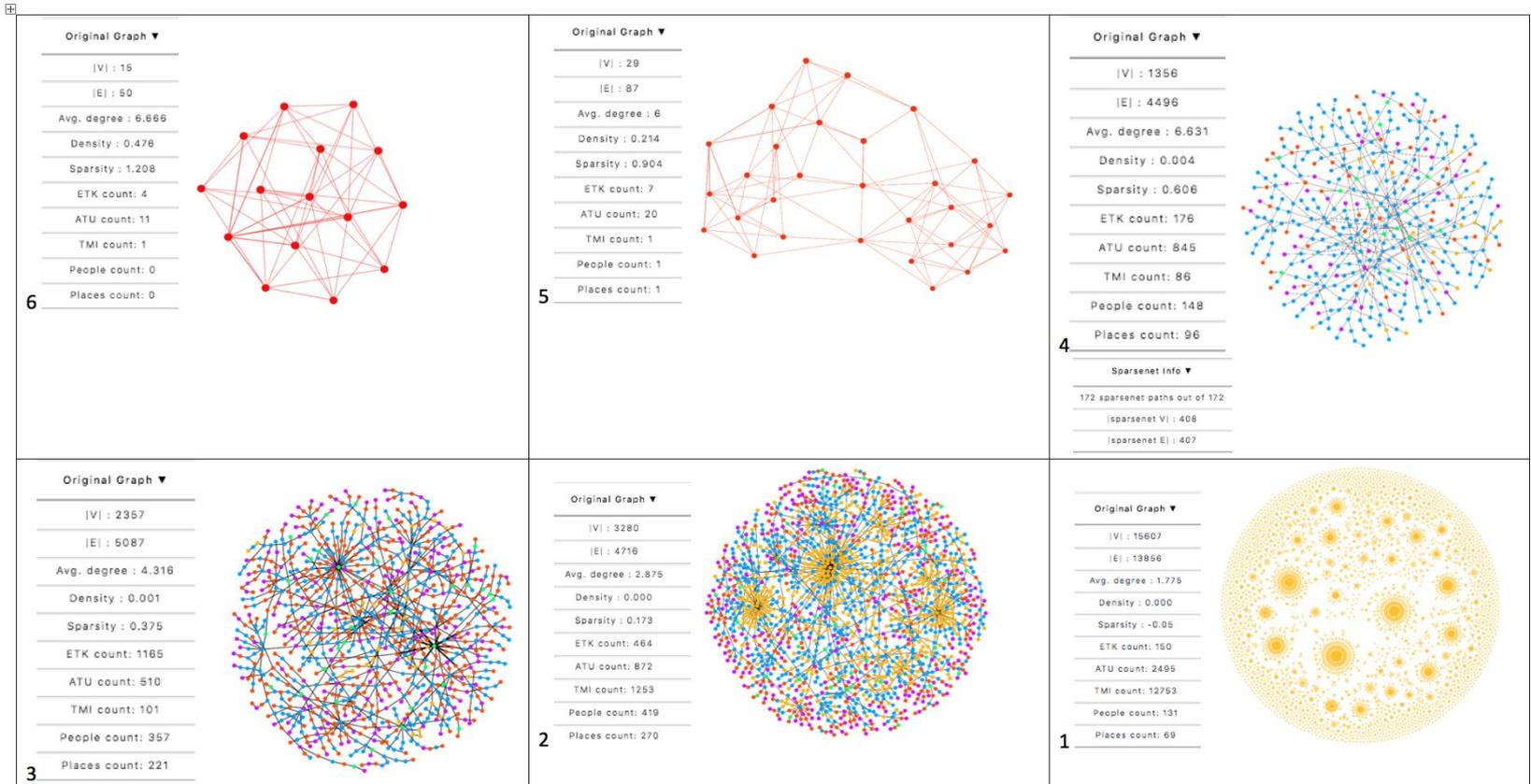
- Despite the considerable gains in understanding presented by these fairly coarse methods for reducing the hairball, the overall topology of the graph dangles just out of reach.
- While we can find unexpected pathways between stories that, in the standard index, appear largely independent of each other, and while we can discover the interdependence of stories, storytellers and geographic locations, it is still not clear what is holding the network together at its most fundamental level.
- We want to determine what the “spines” of the network are that give it its topology.

# A novel decomposition

- Because the edges as well as the nodes contain information about the network, we want to decompose the network based on these relationships
- We develop a method of edge-peel decomposition, that allows us to discover the underlying topologies at different levels of edge degree
- Each “layer” of this peeling algorithm tells us something different about the underlying structures of the network
- By examining the different layers—and the different “waves”—of the peel, we can discover different types of structures that hold the otherwise intractable network together

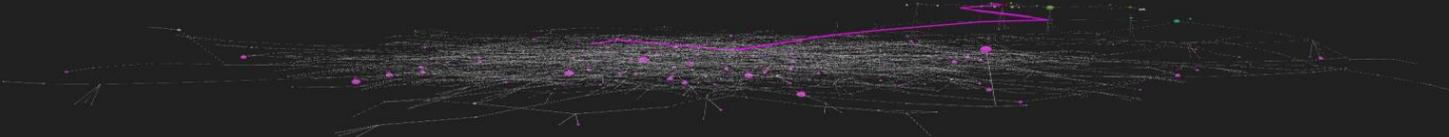


# Six distinct levels



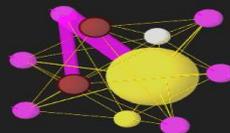
# What we find (confirmatory)

- When we do this, we discover
  - Confirmation of existing theoretical discussions about folk tale tellers the motivation for tale telling in the late 19<sup>th</sup> century (Holbek 1987)
    - Among the ATU 300-749 tales, there is a preponderance of male active tales
  - Certain aspects of the contours of the tradition that have been commented in other works
    - Holbek's main 128 story tellers all appear in layers 5 through 2 in the decomposed graph
    - Women tellers are equally well represented in layer 4 and layer 5, suggesting that female tellers played a significant, yet substantively different role in the tradition



# More importantly, we find (novel)

- We discover certain features of the Danish folktale tradition that have not been commented on before
- At the two very highest level of the decomposition, we find several groups of ATU tale types and a small number of folk tales that:
  - Emphasize labor contracts
  - Emphasize deceit and trickery
  - Consider challenges of farm work, including plowing, baking, animal care
  - Put considerable emphasis on the existence & machinations of hidden folk

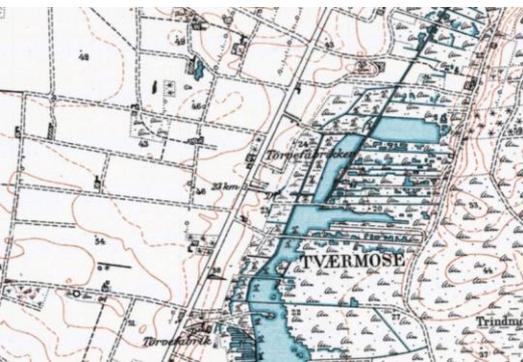


# Some tantalizing clues

- While Holbek makes a convincing case for the idea that many of these tales explore the potential for incremental economic advancement through strategic marriage arrangements
- Many of the other tales clearly explore the exigencies of the shifting agricultural organization of late 19<sup>th</sup> century Denmark including:
  - Labor arrangements
  - Aftægt (elderly) arrangements and dealing with poor/infirm
  - Land use and the shift toward dairy/swine production
- These stories may not be as much about sex and marriage as hypothesized, and may instead be a lot more about economic change and the challenges faced by the (very) small holders as a result of the reorganization of the economy

# Additionally, we find

- Certain places, such as Brandstrup and Tværmose, both of which are in North Central Jutland, play influential roles on the highest levels of the decompositions
  - Brandstrup reveals the degree to which ETK relied on local informants for his collections, allowing him to visit repeatedly certain narrators
  - Tværmose, in contrast, was a fraught location where the draining of the fields was both a potential economic boon but also a potential calamity



# Intriguing story tellers

- The most productive storytellers, as measured by the number of stories they told, may not be the only ones to focus on
- On level 5 of the decomposition, we discover Ole Melchiorson:
- Ole Melchiorson Uglkjær was born December 19, 1822 in Vester-Skovlund, Sale parish. He left home when he was thirteen and became a tailor's apprentice. He worked as a tailor until he was 24, got married and moved to Tyrre heath houses in Borbjerg. That's where I met him and he told me his stories... In 1895, he moved to his daughter's house in Kyllinglund where he lives now. In the heath houses, the family suffered from a lot of illness, as his wife and son were in bed most of the time I was there, only he was healthy. **He learned his stories from many different places, wherever his work took him. He had, as he said it himself, a good aptitude for these things and could learn all sorts of bs.** These days he makes reins and weaves baskets, and then works at various odd jobs. I've noticed he always has his pipe, he smokes early and late



# Preliminary conclusions

(What can I do now that I couldn't do before?)

- Even with limited information, a network model, properly interrogated yields non-trivial insight into a tradition *as a whole*
- The network decomposition holds the promise of making large collections or substantial representations of tradition groups *comparable*
- The ability to identify *spines in the tradition* at varying levels provides insight into how a tradition holds together *semantically*
- The identification of important themes, even if based on sparse descriptions such as ATU, allow one to explore *why* people told the stories they did *when* they did