

Cahiers de Linguistique Asie Orientale  
East Asian Languages and Linguistics  
February 25, 2020

# Improving data handling and analysis in the study of rhyme patterns

*Johann-Mattis List*

Max Planck Institute for the Science of Human History  
*mattis.list@shh.mpg.de*

## Abstract

By reviewing a recent quantitative study of rhyme patterns in Mandarin Chinese, this study shows how data handling and data analysis in the study of rhyme patterns can be improved. Suggestions for improvement include (a) a consistent annotation of rhyme data, which is exhaustive and facilitates data reuse, and (b) emphasizes the importance of automated approaches for exploratory data analysis, which can help to analyze rhyme data in an improved way, prior to applying statistical frameworks for hypothesis testing.

## Keywords

rhyme patterns, Mandarin Chinese, rhyme analysis, data handling

## Résumé

En examinant une récente étude quantitative des modèles de rimes en chinois mandarin, cette étude montre comment le traitement et l'analyse des données dans l'étude quantitative de rimes peuvent être améliorés. Les suggestions pour l'amélioration comprend (a) une annotation cohérente des données de rimes, qui est exhaustive et facilite la réutilisation des données, et (b) souligne l'importance des approches automatisées pour l'analyse exploratoire des données, qui peut aider à analyser les données de rimes de manière améliorée, avant d'appliquer des cadres statistiques pour les tests d'hypothèses.

## Keywords

motifs de rime – chinois mandarin – analyse de rimes – traitement des données

## 1 Introduction

Wang Zhihao's study on "A Linguistic Study on Rhyming in the Beijing Dialect" (this volume) is a considerable contribution to the growing field of quantitative studies devoted to rhyming in Chinese linguistics and beyond (Baxter 1992; List 2016; Bu 2019; List et al. 2019). The amount of data which was annotated by the author in this study is impressive and may help encourage future studies on different epochs and varieties of Chinese. There are, however, two points where the study could be further improved. First, the data and the rhyme judgments are presented in a form that makes it extremely difficult, if not impossible, to actually check the validity of the proposed rhyme judgments, not to speak of allowing scholars to build on the data in their own research. Second, while the method the author chose to analyze the data may be in concordance with analysis techniques proposed in the past by other scholars, the author missed the chance to apply techniques for exploratory data analysis which allow for an additional view on the data that may be particularly elucidating.

Given the increased interest in quantitative rhyme analysis and the potential that more, specifically young scholars, will follow up on this, I think it is useful to discuss the two problems in more detail here, as this may increase the quality of future data handling and encourage scholars to try to think out of the box when it comes to analyzing rhyme data quantitatively. In the following, I will briefly discuss each of the two points and then conclude by pointing to some future challenges in rhyme annotation.

## 2 Data handling

When sharing data and code underlying a study, we do this for multiple reasons. First, we want to make sure that the results can be *replicated*. When sharing data and code, our colleagues can apply them on their own computers and make sure they achieve the same

results. Second, we want to allow our colleagues to *learn* from our study. By investigating our data and code, our colleagues can learn how we analyzed the data, and can use our analysis to carry out similar analyses on different datasets. Third, we want to make sure our colleagues can *challenge* our research by inspecting our data and our code and searching for potential errors. As a fourth point, which is often overlooked, we may want to give our colleagues a chance to *build* on our research, for example, by using our data as a basis that they further expand.

All four aspects, *replication*, *education*, *falsification*, and *amplification*, are vital for scientific research. In fact, as scientists, we have all profited in the past from the fact that scientific research – by and large – follows these four principles of sustainability, even in those cases where studies are not based on data and code.

While scholars voluntarily subscribe to rigorous principles when it comes to the citation of previous work, which is fundamental to account for education, replication, and falsification, they are often surprisingly lax when it comes to submitting data and code. In many cases, it is hard to *find* the data underlying studies, since scholars often put the information where they can be found in a tiny footnote without further comments. Often, it is also hard to *access* the actual data, since links may be broken, or data that had been promised to be provided never made its way online. Having accessed the data, it is often difficult to actually use them, to *interoperate* them, as instructions on the basic structure are missing, or data is shared in proprietary formats that can only be used by those who bought the original software. This all makes it extremely difficult to *reproduce* a given piece of research. These four aspects of *findability*, *accessibility*, *interoperability*, and *replicability* are nowadays subsumed under the label FAIR and propagated as one key guiding principle for the compilation of research data (Wilkinson et al. 2016).

The requirement that data underlying research should be FAIR is a commonplace in most branches of science now. Although linguistics is rather slow in following up, one can note a recent awareness among scholars, journal editors, and reviewers, that it is not enough to submit only the results, if a study itself was based on data and code. One can also see that in the study by Wang, since data and detailed instructions on how to arrive at the results, have been archived with Zenodo (<https://zenodo.org>), and instructions on how to carry out the calculations are also provided in the article.

Strictly speaking, however, Wang’s data is not entirely FAIR, as it is presented in a form that severely hampers interoperability, and obstructs falsification and amplification. The reasons for this problem lies in the format used for the annotation of the rhyme judgments. When dealing with annotation, there are two basic principles that can be used, namely *stand-off annotation*, where the information for a given resource is provided in another document, and *inline annotation*, where the resource is rendered completely, and information is added with the help of a markup language inside the original text.

Recommendations for rhyme annotation have recently been proposed and outlined by List (List et al. 2019). The proposed solution consists of two basic formats for rhyme annotation. The first format is an extended standoff format that renders rhyme data in tabular form along with the original rhyme texts and allows for a very detailed annotation of rhyme sequences, including phonetic alignments (List 2014) of rhyme words. The second format is based on a very straightforward inline annotation schema that was designed for the purpose of allowing scholars to annotate rhyme resources efficiently. While the inline annotation schema is less powerful than the table-based annotation, it provides enough functionality to annotate rhyme data in the form used by Wang, and it was used to digitize rhyme judgments in the work of Baxter (Baxter 1992) and Wang (Wáng 1980), as presented in List et al. (List et al. 2019).

When looking at the data prepare by Wang, we can see that neither of the two methods was used directly. Instead, rhyme sequences are represented in a rather idiosyncratic format in which rhyme sequences are rendered in a spreadsheet, with one poem (or stanza?) per line. The first three columns of the spreadsheet indicate the page number in the source, from which the rhymes were taken (Jiǎ 2009), the district from which the poem stems, and the source, from which the rhyme sequence was taken. The rhyme judgments are then indicated in the following columns of the spreadsheet in such a way that all words that rhyme with each other are placed in consecutive columns in the same row, while changes in the rhyme in a given rhyme sequence (a short poem or similar) are indicated by adding an empty cell in the row.

As an example for the data itself, Wang shows the short poem 筷子 *kuàizǐ* “Chopsticks”, on page 716 in the source, with rhyme words marked in bold font here (underlined in Wang’s text):

Chinese	Pīnyīn
朝阳区	Cháoyáng qū
哥俩一边大,	Gēliǎ yībiān dà,
见面不说话。	Jiànmiàn bù shuōhuà.
说话就顶嘴,	Shuōhuà jiù dǐngzuǐ,
顶嘴就打架。	Dǐngzuǐ jiù dǎjià.

The concrete annotation of the data, however, is rendered as follows in Wang's data:

Page Number	District	Source	Rhyme	Sequence
716	朝阳	搜集	大 d(a)4	话 h(ua)4 架 j(ia)4

Thus, instead of annotating directly which words rhyme, Wang simply lists the rhyme words in this rhyme sequence. While this may seem to be enough to *retrieve* the original rhyme judgment, provided the original resource is available, the format has several shortcomings: First, given that one rhyme sequence may have more than one rhyme, Wang has to make use of empty cells in the spreadsheet to indicate that a new rhyme sequence starts within the same poem. This also means that the format is not capable of rendering rhymes that cross, such as rhyme schemas of the form ABAB, as they can be frequently met in the literature. Second, since only rhyme words are given, but their position is not indicated, it is possible that the annotation itself is ambiguous, since words may be repeated in different parts of a poem. This essentially means that it may not be possible to identify the concrete rhyme judgments from the original resource by simply looking at the short-cut format we find in Wang's data. Third, because it is difficult to understand the concrete rhyme judgments that have been made by the author, it is also difficult to criticize and correct them efficiently. Fourth, since the format imposes huge restrictions on the kind of rhymes that can be annotated in a transparent manner, it cannot be used as a basis for an expanded, derived dataset.

While I understand that our recommendations were published only after Wang already had collected the rhyme data, it is important to note that any project involving data should be planned in such a way that the major aspects of replication, education, falsification, and am-

plification are guaranteed. In order to guarantee them, however, any data collection needs to be preceded by a thorough planning of the data collection. Wang’s data has not been wrongly collected *per se*, but given the size and the beauty of the original collection which Wang used, it is a pity that all this work cannot be easily re-employed in future studies.

But how should one code one’s rhyme data in such a way that it *can* be reused in later studies? For all analyses, I generally recommend to start with the inline annotation schema proposed in List et al. (List et al. 2019), since this annotation can be carried out most efficiently, and is therefore ideal to get an analysis started. In this form, the poem rendered above, would look as shown in Figure 1.

```
@title: 筷子
@source: 搜集
@page: 716
```

```
朝阳区
哥俩一边 [a] 大,
见面不说 [a] 话。
说话就顶嘴,
顶嘴就打 [a] 架。
```

Figure 1: Format for inline rhyme annotation.

Once the data have been annotated in this form, one can use the PoePy Python library (List 2019) and convert them into the more flexible table format. PoePy allows for a convenient rendering of rhyme word relations that can be used in publications, as shown in Table 4, which was produced from the inline annotation (code examples are provided in the supplementary material accompanying this study).

With the expanded format, the data can be annotated in many additional ways. They can also be conveniently inspected and edited with the help of the EDICTOR tool (List 2017), which was originally designed for the handling of etymological data in historical linguistics, but is likewise apt to handle rhyme sequences, as shown in Figure 2.

ID	STANZA	LINE	R:1
1	1.1	朝阳区	
2	1.1	哥俩一边 大， 大	
3	1.1	见面不说 话。 话。	
4	1.1	说话就顶嘴，	
5	1.1	顶嘴就打 架。 架。	

Table 4: *Tabular display of rhyme data with the help of PoePy.*

ID	POEM	POEM NUMBER	LINE IN SOURCE	LINE	LINE ORDER	RHYMEIDS
1	筷子	1	朝阳 区	朝 + 阳 + 区	1	0 0 0
2	筷子	1	哥俩 一 边 [a]大，	哥 + 俩 + 一 + 边 + 大，	2	0 0 0 0 1
3	筷子	1	见 面 不 说 [a]话。	见 + 面 + 不 + 说 + 话。	3	0 0 0 0 1
4	筷子	1	说 话 就 顶 嘴，	说 + 话 + 就 + 顶 + 嘴，	4	0 0 0 0 0
5	筷子	1	顶 嘴 就 打 [a]架。	顶 + 嘴 + 就 + 打 + 架。	5	0 0 0 0 1

Figure 2: *Inspecting rhyme data in EDICTOR.*

### 3 Data analysis

The analysis reported in Wang’s study is based on the assumption that rhyming is based on strict categories. Although it is mentioned in the study itself that the traditional distinction between strict and free rhyming is unwarranted, and that an additional category should be added, the author sticks to the idea that rhyming can in some way be handled in a strict way.

For this reason, the author also refuses to apply the network-based approach for the exploratory analysis of rhyme groups in rhyme corpora, as proposed in List (2016), emphasizing that:

It is a better choice when the phonology of the object language is unknown, because any given groups of words are based on researchers’ suppositions. However, this method is unnecessary when the research object is a modern language. The nature of rhyming suggests that words with the same final and tone do rhyme with each other freely. So, finals and tones are naturally reliable

groups. The task is just to study the relation between different finals or tones. ([Wang, this volume, pp. XXX](#))

Com.	Memb.	Rhy.	Variants
1	341	an	en / 1, van / 30, uan / 70, ian / 112, an / 128
2	217	ang	uang / 42, iang / 67, ang / 108
3	220	ao	iao / 92, ao / 128
4	262	ing	uen / 1, uang / 1, ueng / 2, en / 2, in / 4, iong / 11, ong / 63, eng / 84, ing / 94
5	113	ou	v / 1, ao / 1, iou / 49, ou / 62
6	171	a	iao / 1, ie / 1, ua / 31, ia / 32, a / 106
7	159	i	y / 1, uei / 2, w / 12, v / 26, i / 118
8	143	uo	ua / 1, v / 1, u / 4, e / 51, uo / 86
9	97	ai	ia / 1, anr / 1, uai / 17, ai / 78
10	97	en	eng / 3, vn / 11, uen / 24, in / 27, en / 32
11	104	u	u / 104
12	95	ianr	uor / 1, ier / 1, uair / 1, uar / 2, er / 2, ar / 4, iar / 5, vanr / 6, air / 8, uanr / 13, ianr / 26, anr / 26
13	52	uei	ei / 23, uei / 29
14	50	ueir	vnr / 1, uar / 1, anr / 1, uanr / 1, er / 1, vr / 1, engr / 1, wr / 2, inr / 2, uenr / 2, r / 2, uor / 2, ver / 2, eir / 3, ur / 3, ier / 3, yr / 4, ir / 5, enr / 6, ueir / 7
15	47	enr	eir / 1, ingr / 1, yr / 2, engr / 2, r / 2, inr / 3, vnr / 3, ueir / 3, vr / 4, wr / , uenr / 5, ir / 6, en / 11
16	38	uei	v / 1, i / 1, ei / 13, uei / 23
17	41	ie	ian / 1, e / 2, ve / 6, ie / 32
18	25	ianr	ir / 1, vanr / 1, air / 1, er / 1, ier / 1, ingr / 2, uanr / 4, ar / 4, ianr / 5, anr / 5
20	21	ueir	anr / 1, vr / 1, ur / 1, enr / 2, yr / 2, uenr / 2, ir / 5, ueir / 7

Table 6: Major communities identified by the Infomap clustering.

This assumption is problematic in many regards. First, the author assumes that it is possible to know the true phonology of a given dialect variably, although the data from which the corpus is drawn itself may show multiple forms of variation, including not only *space* (diatopic variation), but also *time* (diachronic variation) and *social class* (diastratic variation), and the individual *situation*, in which speech is uttered (diaphasic variation). The only situation in which the phonology of a given language can be reliably measured is when all these dimensions of linguistic variations are reliably controlled for. Second, the author assumes that the phonology of a given language can predict the rhyming behavior, but does not elaborate on additional factors conditioning rhyming, such as, notably, *culture*, as represented by the rhyming practice which poets learn from poetry when they grow up, *cognition*, as represented by overall principles of phonetic

similarity, and *communication*, as represented by the linguistic system the poet speaks or wishes to elicit. Third, the whole *process* by which rhyme data are collected is poorly described by a process in which poets pick all meanings they wish to express and then search for rhyme words to fill the slots. This process may hold for the process of poem writing (although the whole picture is beyond doubt more complex), but the data collected in a rhyme analysis are based upon the interpretation of a linguist analyzing a given poem. What the poet itself thought – or if poets think *at all* when forming words into speech – is a question that is not accessible to us.

Given that we deal with so many uncertainties when analyzing corpus data of rhyme judgments, it seems problematic to exclusively use techniques for data analysis which assume that rhyme groups are fixed entities that can be objectively detected prior to the analysis. It is all the more surprising, since in the study by Wang itself, it is mentioned that “Semi-free rhyming and free rhyming constitute a continuum. It is difficult to decide where to draw the boundary between them.” (Wang, this volume). While there is nothing to say against the analysis presented in Wang’s study, it is a pity that the alternative approach, which allows us to inspect the data in a much less biased way, has not been pursued in this study, specifically also because it is by no means difficult to apply the approach described in List (2016).

The major idea proposed in List (2016) is to model rhyme data in form of a network in which individual rhyme words are represented as *nodes* and weighted links between the words are used to represent how many times the words rhyme in a given corpus. Once a network has been reconstructed in this way, it can be analyzed with the help of network techniques that search for *communities* in a given network, with a community being defined by a partition of a network in which the number of links inside the community is greater than that with other nodes in the network (Newman 2004). List (2016) proposed to use the Infomap algorithm by Rosvall and Bergström for this purpose (Rosvall & Bergstrom 2008). The output of this analysis is a division of the network into communities, which, in theory, correspond to distinct rhyme groups in the data. Since it is possible that not only Wang, but also other scholars find the network analysis proposed in List (2016) difficult to apply, I thought it may be useful to present the results of this analysis here.

The supplementary material accompanying this study shows ex-

plicitly how the data of Wang can be converted to a rhyme network, and how this network can be analyzed with the help of the Infomap algorithm, as proposed in List (2016). The analysis yields a network consisting of 2890 rhyme words, which consists of 102 connected components, with one very large component of 2566 rhyme words. That means that 89% of the rhyme words in the rhyme network are connected directly or indirectly. This does not only confirm the findings in List (2016), where a similarly large connected component could be identified, it also shows that rhyming is best understood as a continuum, and that completely clear rhyme groups are hard to identify.

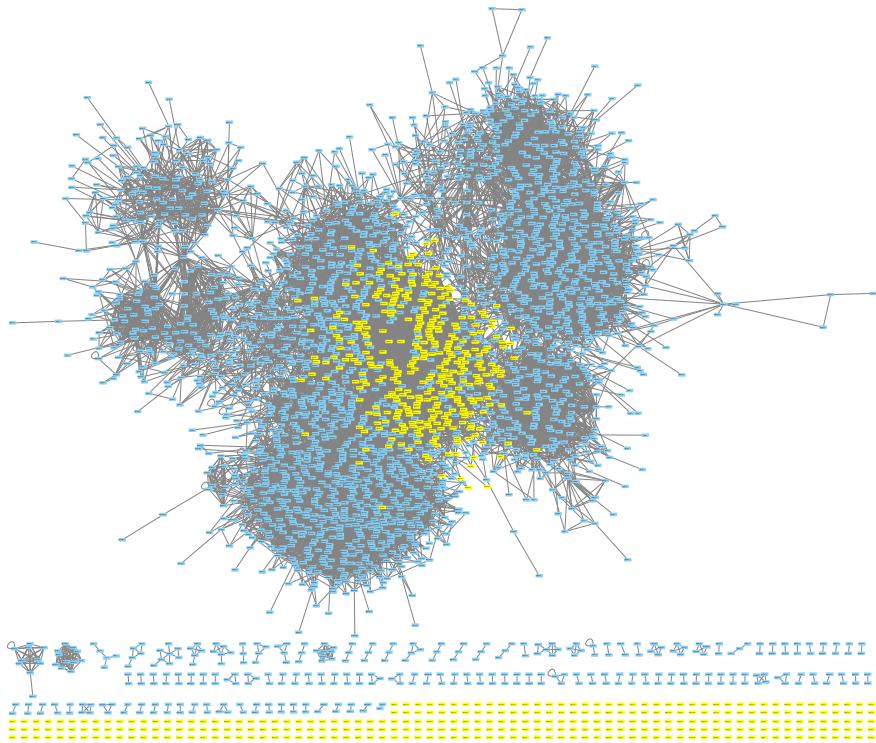


Figure 3: Full rhyme network derived from Wang's data

When applying the Infomap community detection algorithm to identify larger rhyme groups in the data, we find as many as 201 communities. In contrast to the connected components in the network, these communities are rather well-balanced. There are 19 communities with 20 or more rhyme words. Together they make up for 2293

nodes, which corresponds to 80% of the rhyme words in the data. The major components of these communities, as reflected by the Pīnyīn pronunciations of the rhyme words involved, are provided in Table 6.

What we can see from this analysis is that the network has a rather strong community structure, which can be seen as first evidence that rather well-definable rhyme groups play an important role in rhyming. However, we can also see that the Infomap community detection analysis has separated certain rhymes from each other that we would rather group together, such as communities number 13 and 16, 14 and 20, as well as 12 and 20. Apart from this, however, the communities reflect the rhyme groups as discussed by Wang rather well. That the automated community detection analysis subdivides some groups that obviously belong together results from the sparseness of the data and should not surprise us further. It seems reasonable to assume that the communities will become more pronounced with more data being considered.

Since it would require too much time to investigate all of the communities in this sample, I will concentrate on the first and largest one identified by the Infomap algorithm. This comprises rhymes centering around rhyme words with finals pronounced as *-an*, according to the annotation provided in Wang's data, along with the finals *-ian*, *-uan*, and *-van* (phonetically *yan*). Wang's analysis seems to indicate that the whole group as such does not reflect free rhyming *per se*, but rather involves a subtype of free rhyming, which is called semi-free rhyming. Given that our community analysis included all four finals into the same group, it seems worthwhile to investigate this group further.

As a first step, consider Figure 4, which shows two views of a subgraph derived by taking all rhyme words with one of the four finals *-an*, *-ian*, *-uan*, or *-van*, following the pronunciation provided in Wang's data. The first view shows the network in a force-directed layout, while the second view distinguishes the four pronunciation groups.

What can be seen from this figure, without applying any further technique alone, is that the four subgroups seem to rhyme freely with each other. The big picture does not allow us to assume any form of strong intransitivity in rhyming, even if there may be some tendencies among the subtypes.

This can be further confirmed when applying the quantitative method proposed by List et al. (2017), where *assortativity* (Newman 2003) is

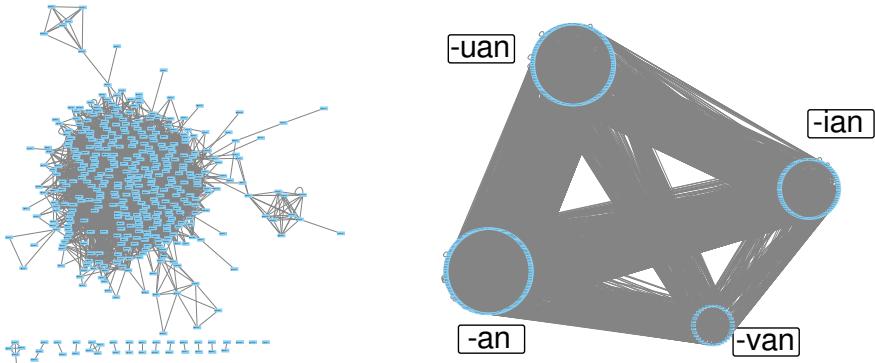


Figure 4: Two views of the subgraph of all rhyme words with finals in -(iuv)an.

used to measure to which degree rhyming is influenced by vowel purity in Old Chinese rhyme data. Assortativity tests whether a certain partition of the nodes in a network into groups is also reflected by the network structure of the network itself. While assortativity was used to test if rhyming favors identical vowels or not, we can use the same idea to test if distinctions in the medials in rhyme words with final *-an* find a reflection in the graph structure.

In order to do this, we first compute the assortativity coefficient of the subgraph shown in Figure X above, with respect to the four subgroups *-an*, *-ian*, *-uan*, and *-van*. To check if the result is significant, we then carry out 1000 random trials in which we randomly assign the nodes in the network to one of the four groups, and re-compute the assortativity coefficient for each of these trials. We then calculate the standard deviation of the 1000 trials and compute, by how many standard deviations the attested assortativity differs from the one we achieve in our random trials.

The results of this analysis (which is included as part of the supplementary material) yield an assortativity coefficient of 0.01 for the subgraph, and a mean assortativity coefficient of 0.00 for the random trials, with a sigma score of 2.21. This means on the one hand that there is a significant difference between the randomly assigned groupings and the groupings given in the data themselves. On the other hand, however, it also shows that the amount of assortative mixing in the data is extremely *low*. An assortativity coefficient of 0 indicates that there is no evidence for assortative mixing in the data, or – to apply this to our rhyme network – that there is no evidence

for different rhyme groups in a given set of nodes. A coefficient of 0.01, as we find it in this experiment, indicates a very small tendency for words from the same subgroup to rhyme with each other. But this tendency is so small that it can almost be neglected. To add some context to the obtained score of 0.01: in List et al. (2017), all reconstruction systems that were tested for vowel purity showed assortativity coefficients of 0.5 and higher.

In the light of data-driven network approaches, there is no evidence for semi-free rhyming in the group of -an finals, contrary to what was postulated by Wang. Given the continuous character of rhyming, as admitted by Wang as well, it may not even be needed to postulate such a group. Instead, it seems more useful to report results based on the data themselves, and in this context, measures like assortativity have the advantage that they are not binary, but can take rhyme groups involving multiple subgroups into account.

#### 4 Conclusion

In this small review of Wang's quantitative study on rhyming in Beijing Chinese, I have tried to show how data handling and analysis in the study of rhyme patterns can be improved. With respect to data handling, I have concentrated on recently proposed annotation frameworks which have the advantage of reflecting rhyme analyses in a replicable way that can be easily expanded, criticized, and modified. With respect to the analyses, I have tried to illustrate the importance of network approaches when trying to identify consistent rhyme groups in a given dataset. By showing how tests for assortative mixing can be used to assess whether a rhyme group should be further subdivided into smaller groups, I have tried to propose an alternative for the binary test procedure used in Wang's study. While this is not intended to reject Wang's analysis, which is valuable in its own right, I think it is desirable for future analyses to include network techniques as well, since these techniques offer an important additional perspective on rhyme data.

## Supplementary Material

The supplementary material accompanying this study contains the source code and the data required to replicate the studies reported here, along with a description explaining how to apply the code. It has been uploaded to the Open Science Framework, where it can be downloaded from <https://doi.org/10.17605/OSF.IO/C9VDP>.

## Acknowledgments

This research was funded by the ERC Starting Grant 715618 “Computer-Assisted Language Comparison” (<http://calc.digling.org>). The author thanks one anonymous reviewer for help with the manuscript, and expresses his gratitude to Wang Zhihao for sharing the data underlying the study transparently.

## References

- Baxter, William H. 1992. *A handbook of Old Chinese phonology*. Berlin: de Gruyter.
- Bu, Aisong. 2019. *An optimality-theoretical analysis of Chinese Great Vowel Shift*. Beihang University. (Master thesis).
- Jiǎ, Zhī 贾芝 (ed.). 2009. *Zhōngguó gēyáo jíchéng, Běijīng juàn* 中国歌谣集成 · 北京卷 [*China Folk Rhymes Collection, Vol. Beijing*]. Beijing: China ISBN Center.
- List, Johann-Mattis. 2014. *Sequence comparison in historical linguistics*. Düsseldorf: Düsseldorf University Press.
- List, Johann-Mattis. 2016. Using network models to analyze Old Chinese rhyme data. *Bulletin of Chinese Linguistics* 9(2). 218–241.
- List, Johann-Mattis. 2017. A web-based interactive tool for creating, inspecting, editing, and publishing etymological datasets. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics. System Demonstrations*, 9–12. Valencia: Association for Computational Linguistics.
- List, Johann-Mattis. 2019. *PoePy. A Python library for handling annotated rhymes*. Jena: Max Planck Institute for the Science of Human History. <https://github.com/lingpy/poepy/releases>.
- List, Johann-Mattis & Hill, Nathan W. & Foster, Christopher J. 2019. Towards a standardized annotation of rhyme judgments in Chinese historical phonology (and beyond). *Journal of Language Relationship* 17(1). 26–43.

- List, Johann-Mattis & Pathmanathan, Jananan Sylvestre & Hill, Nathan W. & Bapteste, Eric & Lopez, Philippe. 2017. Vowel purity and rhyme evidence in Old Chinese reconstruction. *Lingua Sinica* 3(1). 1–17.
- Newman, M. E. J. 2003. Mixing patterns in networks. *Physical Review E* 67. 026126.
- Newman, M. E. J. 2004. Analysis of weighted networks. *Physical Review E* 70(5). 056131.
- Rosvall, M. & Bergstrom, C. T. 2008. Maps of random walks on complex networks reveal community structure. *Proc. Natl. Acad. Sci. U.S.A.* 105(4). 1118–1123.
- Wáng, Lì 王力. 1980. *Shījīng Yùndú* 詩經讀 [Rhyme readings in the Book of Odes]. Shànghǎi 上海: Shànghǎi Gǔjí 上海古籍.
- Wilkinson, Mark D. & Dumontier, Michel & Aalbersberg, IJsbrand J. & Appleton, Gabrielle & Axton, Myles & Baak, Arie & Blomberg, Niklas & Boiten, Jan-Willem & da Silva Santos, Luiz B. & Bourne, Philip E., et al. 2016. The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data* 3.