

# Сетевой подход к анализу древнекитайских рифм

Johann-Mattis List

Department of Linguistic and Cultural Evolution  
Max Planck Institute for the Science of Human History  
Jena

2017/03/07

# Введение

## Рифмование в общем

Lose yourself in the music  
the moment you own it  
you better never let it go  
you only get one shot  
do not  
miss your chance to blow  
this opportunity comes once in a lifetime...  
(Eminem, "Lose yourself", 2002)

## Рифмование в общем

Lose yourself in the **music**  
the moment you **own it**  
you better never let it **go**  
you only get one **shot**  
do **not**  
miss your chance to **blow**  
this opportunity comes once in a lifetime...  
(Eminem, “Lose yourself”, 2002)

## Рифмование в общем

Lose yourself in the **music** [-ɪk] ?  
the moment you **own it** [-ɪt] ?  
you better never let it **go**  
you only get one **shot**  
do **not**  
miss your chance to **blow**  
this opportunity comes once in a lifetime...  
(Eminem, “Lose yourself”, 2002)

# Рифмование в общем

*music* [-ɪk]

*own it* [-ɪt]

But Germans would rhyme *employ* and *deny*!

*Mai* [-ɔi]

*neu* [-ai]

# Рифмование как доказательство

<F11> 燕	燕	于	飛，	下	上	其	音。	The swallows go flying, falling and rising are their voices;
<i>yān</i>	<i>yān</i>	<i>yú</i>	<b>fēi</b>	<i>xià</i>	<i>shàng</i>	<i>qí</i>	<b>yīn</b>	
之	子	于	歸，	遠	送	于	南。	This young lady goes to her new home, far I accompany her to the south.
<i>zhī</i>	<i>zǐ</i>	<i>yú</i>	<b>guī,</b>	<i>yuǎn</i>	<i>sòng</i>	<i>yú</i>	<b>nán</b>	
瞻	望	弗	及，	實	勞	我	心。	I gaze after her, can no longer see her, truly it grieves my heart.
<i>zhān</i>	<i>wàng</i>	<i>fú</i>	<i>jí,</i>	<i>shí</i>	<i>láo</i>	<i>wǒ</i>	<b>xīn</b>	

# Рифмование как доказательство

燕	燕	于	飛，	下	上	其	音。	The swallows go flying, falling and rising are their voices;
<i>yān</i>	<i>yān</i>	<i>yú</i>	<b>fēi</b> ,	<i>xià</i>	<i>shàng</i>	<i>qí</i>	<b>yīn</b>	
之	子	于	歸，	遠	送	于	南。	This young lady goes to her new home, far I accompany her to the south.
<i>zhī</i>	<i>zǐ</i>	<i>yú</i>	<b>guī</b> ,	<i>yuǎn</i>	<i>sòng</i>	<i>yú</i>	<b>nán</b>	
瞻	望	弗	及，	實	勞	我	心。	I gaze after her, can no longer see her, truly it grieves my heart.
<i>zhān</i>	<i>wàng</i>	<i>fú</i>	<i>jí</i> ,	<i>shí</i>	<i>láo</i>	<i>wǒ</i>	<b>xīn</b>	



# Рифмование как доказательство

燕	燕	于	飛，	下	上	其	音。
<i>yān</i>	<i>yān</i>	<i>yú</i>	<b>*pər</b>	<i>xià</i>	<i>shàng</i>	<i>qí</i>	<b>*qɾəm</b>
之	子	于	歸，	遠	送	于	南。
<i>zhī</i>	<i>zǐ</i>	<i>yú</i>	<b>*k<sup>w</sup>əj,</b>	<i>yuǎn</i>	<i>sòng</i>	<i>yú</i>	<b>*n<sup>h</sup>əm</b>
瞻	望	弗	及，	實	勞	我	心。
<i>zhān</i>	<i>wàng</i>	<i>fú</i>	<i>jí,</i>	<i>shí</i>	<i>láo</i>	<i>wǒ</i>	<b>*səm</b>

# Рифмование как доказательство

Chinese Text	RW	MCH	Gù Yánwǔ	Wáng (1980)	Baxter (1992)	OCBS-R
殷其雷	<i>léi</i> 雷	*lwoj	雷 A, 之部	雷 -	雷 -	*-uj
在南山之陽	<i>yáng</i> 陽	*yang	陽 B, 陽部	陽 A, <i>jiang</i> , 陽部	陽 A, * <i>ljang</i>	*-aŋ
何斯遠斯	<i>sī</i> 斯	*sje	斯 A, 之部	斯 -	斯 -	*-e
莫敢或遑	<i>huáng</i> 遑	*hwang	遑 B, 陽部	遑 A, <i>huang</i> , 陽部	遑 A, * <i>wang</i>	*-aŋ
振振君子	<i>zǐ</i> 子	*tsiX	子 A, 之部	子 B, <i>tziə</i> , 之部	子 -	*-əʔ
歸哉歸哉	<i>zāi</i> 哉	*tsoj	哉 A, 之部	哉 B, <i>tzə</i> , 之部	哉 -	*-ə

Comparing differences in rhyme identification for Ode 19.1 《周南·殷其雷》

# Рифмование как доказательство

Разные факторы имеют влияние на рифмование (культура, язык, познание). Поэтому это сложно использовать рифмование как доказательство, особенно для исследования тех языков, которые засвидетельствованы только в текстах.

# Фонология древнекитайского языка

- долгая традиция лингвистической реконструкции в Китае (началась с ученого Chén Dì 陳第, 1541 – 1606),

# Фонология древнекитайского языка

- долгая традиция лингвистической реконструкции в Китае (началась с ученого Chén Dì 陳第, 1541 – 1606),
- первый прогрессивный этап в 20ом ст. с работами Карлгрена и китайских ученых Wáng Lí 王力 (1980) и Li Fang-kuei 李方桂 (1971),

# Фонология древнекитайского языка

- долгая традиция лингвистической реконструкции в Китае (началась с ученого Chén Dì 陳第, 1541 – 1606),
- первый прогрессивный этап в 20ом ст. с работами Карлгрена и китайских ученых Wáng Lí 王力 (1980) и Li Fang-kuei 李方桂 (1971),
- второй важный этап в 80ых годах 20ого стол., когда Бакстер (1992), Старостин (1989), и Zhèngzhāng Shàngfāng (см. Zhèngzhāng 2003) предлагали реконструкции которые были очень похожи друг на друга, хотя и были независимо разработаны учеными (6 гласных, больше рифм, чем в традиционной литературе , итд.).

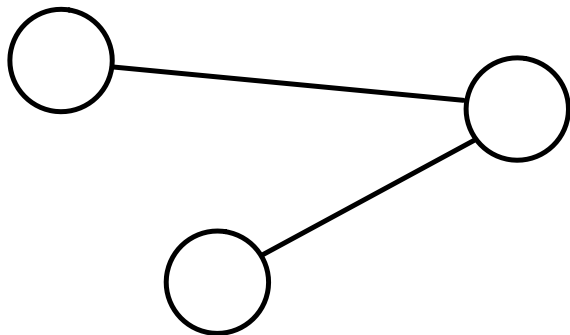
# Фонология древнекитайского языка

Общая проблема реконструкции древнекитайской фонологии — это прозрачность и соизмеримость данных. Есть много разных реконструкций древнекитайского, и много разных аннотаций рифм в «Книге песен» (Shījīng 詩經), но с тех пор никто не сравнивал эти данные конкретно.

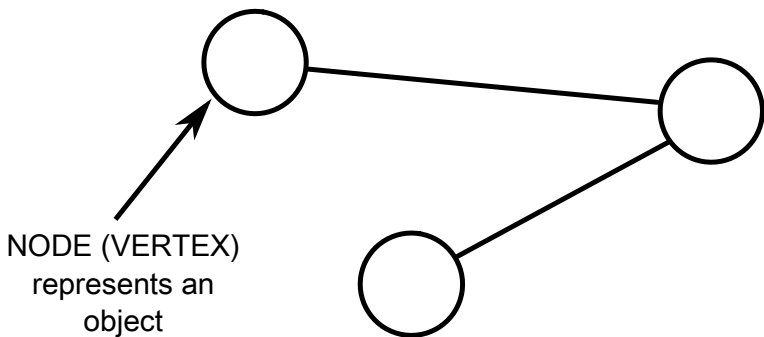
# Сети



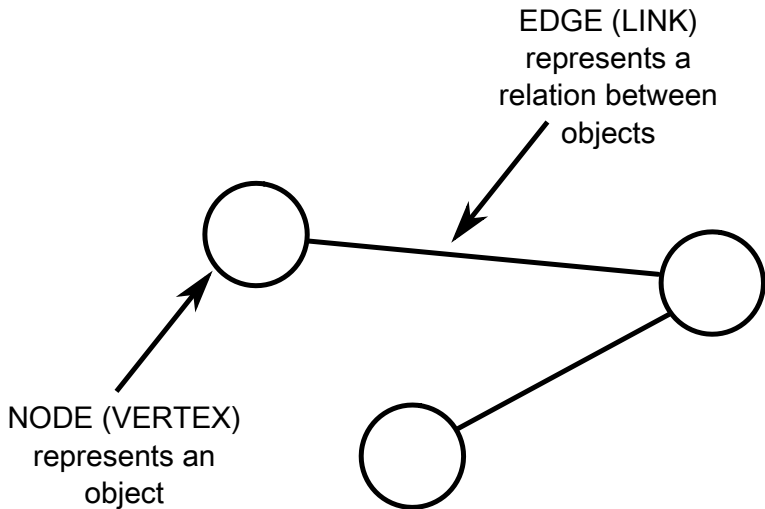
# Сети в общем



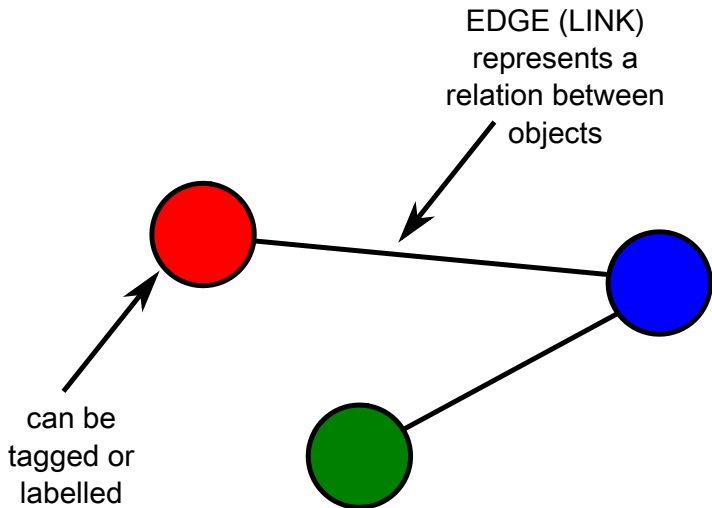
# Сети в общем



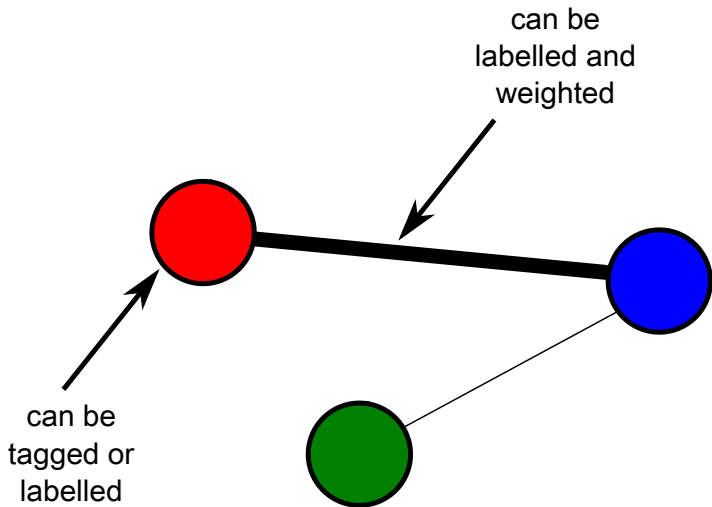
# Сети в общем



# Сети в общем



# Сети в общем



# Сети в примерах

**Большинство структур в повседневной жизни и в науке возможно моделировать как сети:**

- социальные сети: вершины (узлы) — люди, ребра — реляции (например дружба в Контакте, итд.),

# Сети в примерах

**Большинство структур в повседневной жизни и в науке возможно моделировать как сети:**

- социальные сети: вершины (узлы) — люди, ребра — реляции (например дружба в Контакте, итд.),
- филогенетические сети: вершины — языки или диалекты, ребра — генетическая близость,

# Сети в примерах

**Большинство структур в повседневной жизни и в науке возможно моделировать как сети:**

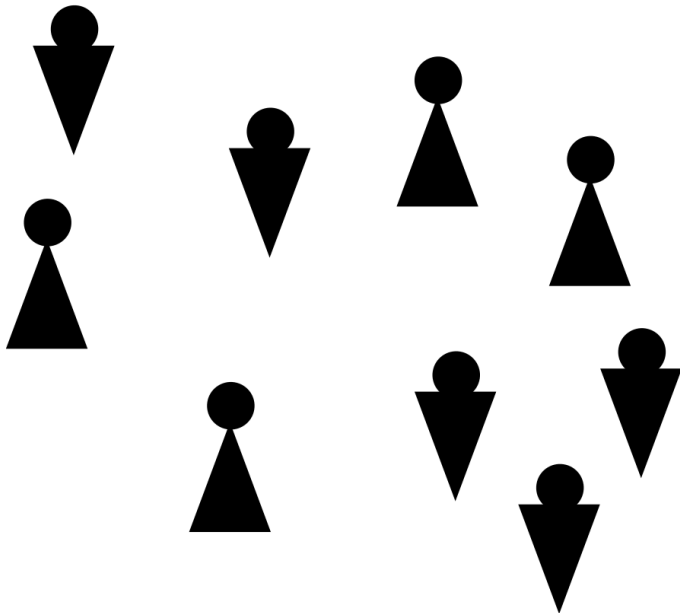
- социальные сети: вершины (узлы) — люди, ребра — реляции (например дружба в Контакте, итд.),
- филогенетические сети: вершины — языки или диалекты, ребра — генетическая близость,
- ...,



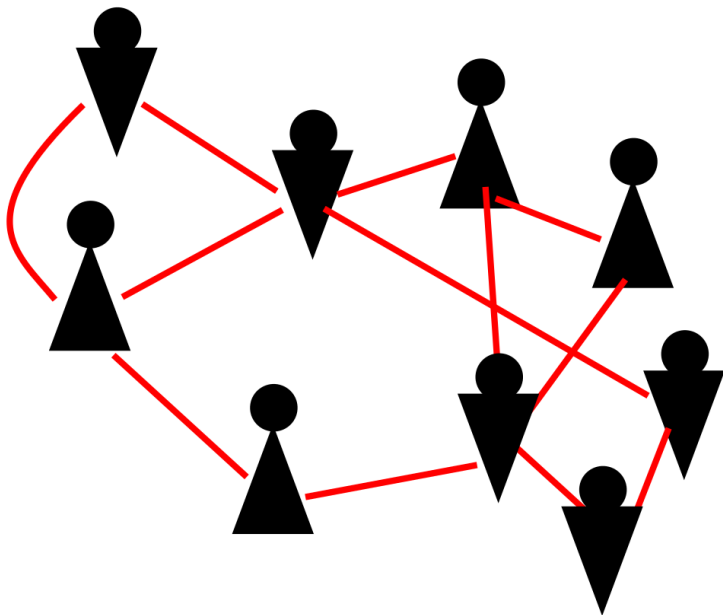
# Сетевые модели

Благодаря сетевым моделям, данные которые моделированы как графы анализируются очень быстро. Литература огромна, и часто ученым очень сложно находить лучший анализ для решения своих проблем. Поэтому иногда необходимо работать междисциплинарно, консультируя экспертов для анализа графов.

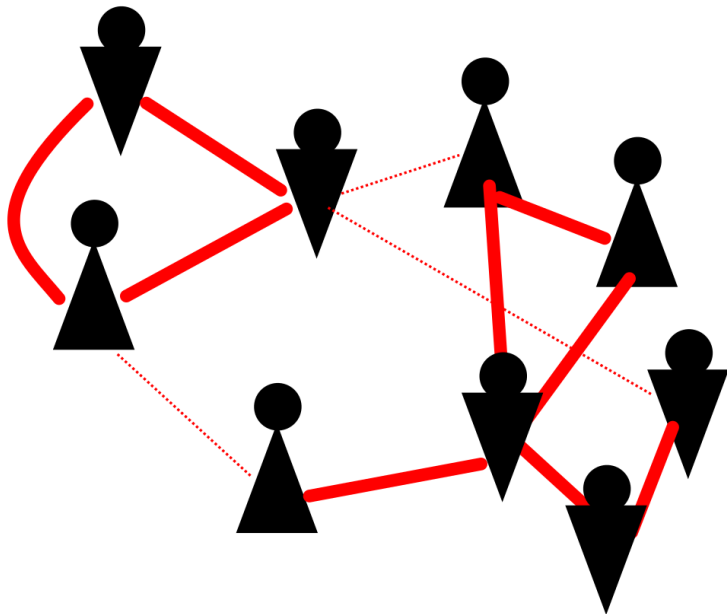
# Сетевые модели



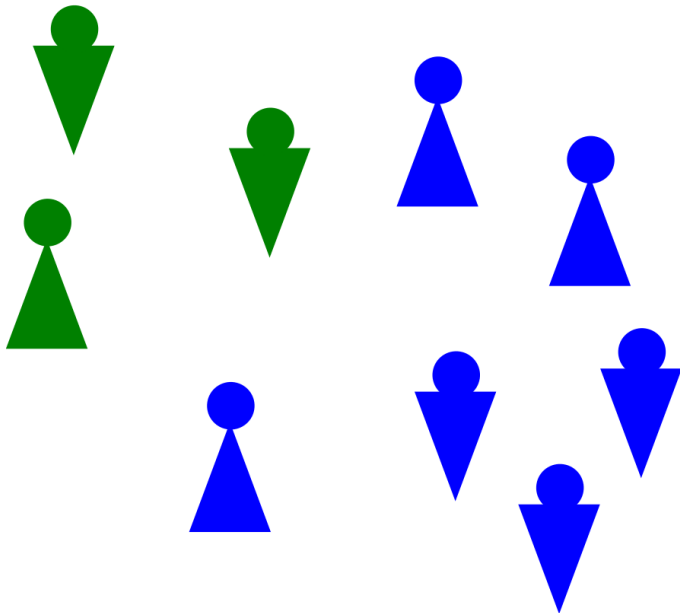
# Сетевые модели



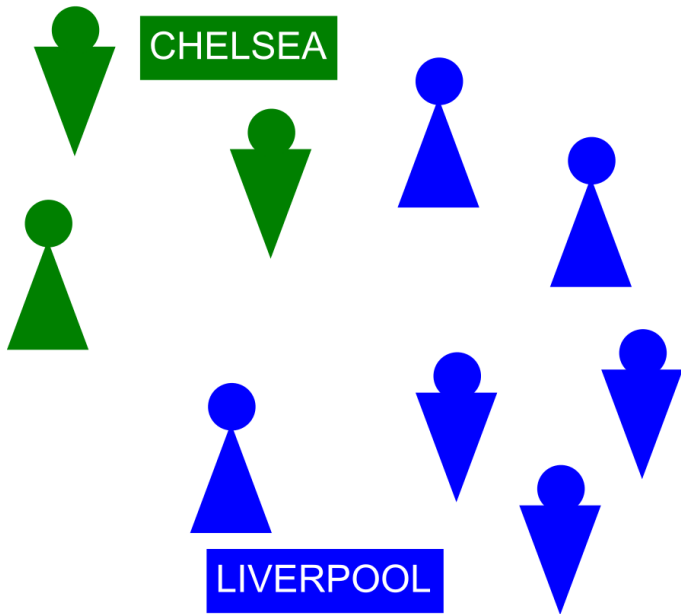
# Сетевые модели



# Сетевые модели



# Сетевые модели



# Сетевые модели

Методы для выявления сетевых сообществ (*community detection*) являются специальным типом алгоритмов для разделения графа на части (*partitioning algorithm*). Они очень полезны для поиска естественных групп в сетях.

# Рифмы и сети



# Моделирование

## Shijing 39 (泉水)

愔彼泉水

亦流于淇

有懷于衛

靡日不思

變彼諸姬

聊與之謀

# Моделирование

## Shijing 39 (泉水)

愔彼泉水

亦流于淇

有懷于衛

靡日不思

變彼諸姬

聊與之謀

# Моделирование

## Shijing 39 (泉水)

愬彼泉水

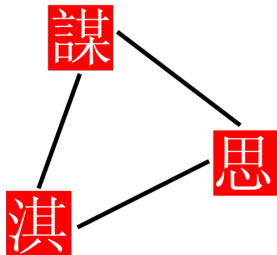
亦流于淇

有懷于衛

靡日不思

變彼諸姬

聊與之謀



# Моделирование

## Shijing 39 (泉水)

毖彼泉水

亦流于淇

有懷于衛

靡日不思

變彼諸姬

聊與之謀

## Shijing 54 (載馳)

我行其野

芄芄其麥

控于大邦

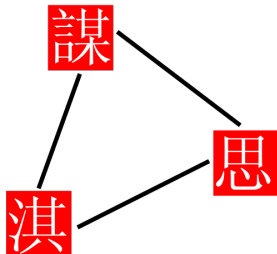
誰因誰極

大夫君子

無我有尤

百爾所思

不如我所之



# Моделирование

## Shijing 39 (泉水)

愬彼泉水

亦流于淇

有懷于衛

靡日不思

變彼諸姬

聊與之謀

## Shijing 54 (載馳)

我行其野

芄芄其麥

控于大邦

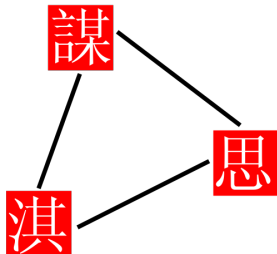
誰因誰極

大夫君子

無我有尤

百爾所思

不如我所之



# Моделирование

## Shijing 39 (泉水)

愆彼泉水

亦流于淇

有懷于衛

靡日不思

變彼諸姬

聊與之謀

## Shijing 54 (載馳)

我行其野

芄芄其麥

控于大邦

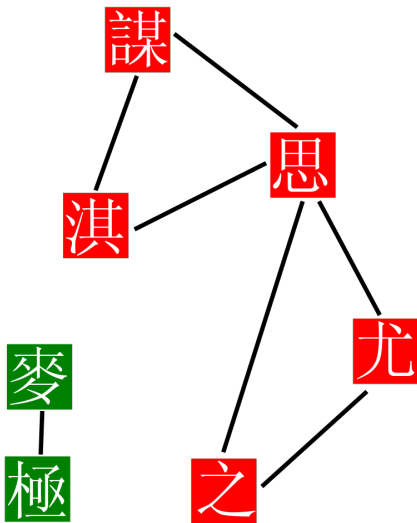
誰因誰極

大夫君子

無我有尤

百爾所思

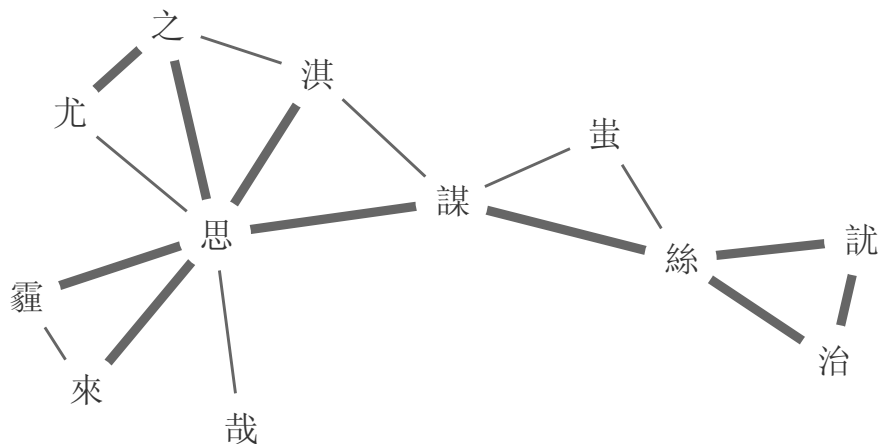
不如我所之



# Моделирование

27.3.A		sī 丝						
30.2.A	lái 来	sī 思						
33.3.A	lái 来	sī 思						
39.1.A		sī 思						
54.4.B		sī 思			zhī 之			yóu 尤
58.1.A		sī 丝	qī 淇	móu 谋				
58.6.B		sī 思			zāi 哉		qī 期	
59.1.A		sī 思	qī 淇	móu 谋				
66.1.A	lái 来	sī 思			zāi 哉		qī 期	
130.1.A					zāi 哉		méi 梅	
204.4.A			qī 淇			zhī 之	méi 梅	yóu 尤
227.2.A					zāi 哉			

# Моделирование





# Моделирование

## 4. 樛木

南有樛木、葛藟纍之。  
樂只君子、福履綏之。

南有樛木、葛藟荒之。  
樂只君子、福履將之。

南有樛木、葛藟縈之。  
樂只君子、福履成之。

Poem	Stanza	Verse	Sect.	Text	Rhyme	Pattern	MCH	OCBS
4	1	1	1	南有樛木、	木	-	muwk	C.m <sup>5</sup> ok
4	1	1	2	葛藟纍之。	纍	A	lwij	[r]uj
4	1	2	1	樂只君子、	子	-	tsiX	tsəʔ
4	1	2	2	福履綏之。	綏	A	swij	s.nuj

# Моделирование

Реконструкция рифмовой сети от аннотированных поэм проста. Проблемы возникают с взвешиванием если иероглифы рифмуются часто или редко друг с другом, с нормализацией длинных поэм, и с аннотацией специальных рифм (рифмы на двух слогах, рифмы в самой строке, итд.).

# Моделирование

- если два слова рифмуются друг с другом много раз в разных поэмах, это более важное доказательство чем спорадическое рифмование (→ use weighted networks to represent frequency)
- если куча слов рифмуется в длинных строфах, у нас больше связей чем в коротких строфах, и мы должны нормализовать это (→ normalize rhyme connections in each stanza)
- рифмы в самой строке и на двух слогах часто следуют специальным законам рифмования (question not yet solved)

# Shījīng Rhyme Browser

- интерактивное веб-приложение
- показывает рифмы «Книги песен» согласно Бакстеру (1992) вместе с реконструкциями Бакстера и Сагара (2014) и Pān (2000, из *Thesaurus Linguae Sericae*).
- аппликация позволяет быструю инспекцию данных древнекитайского
- URL: <http://digling.org/shijing>

## Гипотеза коды -r (List forthcoming)

- теория была впервые предложена С. А. Старостином (1989: 338-340), который предположил, что связи между рифмами с кодами на \*-n и \*-j объясняются добавочной кодой \*-r с регулярным рефлексом -n в среднекитайском,

## Гипотеза коды -r (List forthcoming)

- теория была впервые предложена С. А. Старостином (1989: 338-340), который предположил, что связи между рифмами с кодами на \*-n и \*-j объясняются добавочной кодой \*-r с регулярным рефлексом -n в среднекитайском,
- дополнительные доказательства можно найти в родственных с китайском словах Тибетского (Hill 2014),

## Гипотеза коды -r (List forthcoming)

- теория была впервые предложена С. А. Старостином (1989: 338-340), который предположил, что связи между рифмами с кодами на \*-n и \*-j объясняются добавочной кодой \*-r с регулярным рефлексом -n в среднекитайском,
- дополнительные доказательства можно найти в родственных с китайском словах Тибетского (Hill 2014),
- Бакстер и Сагарт (2014) адаптировали эту идею в своей реконструкции древнекитайского, но количественная проверка до сих пор отсутствует.

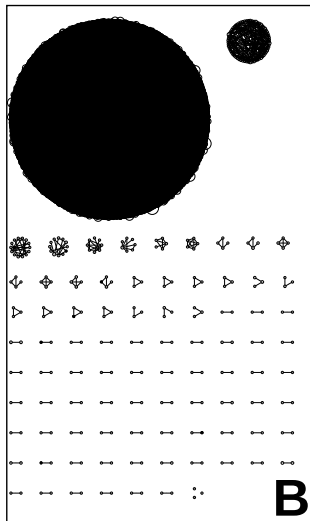
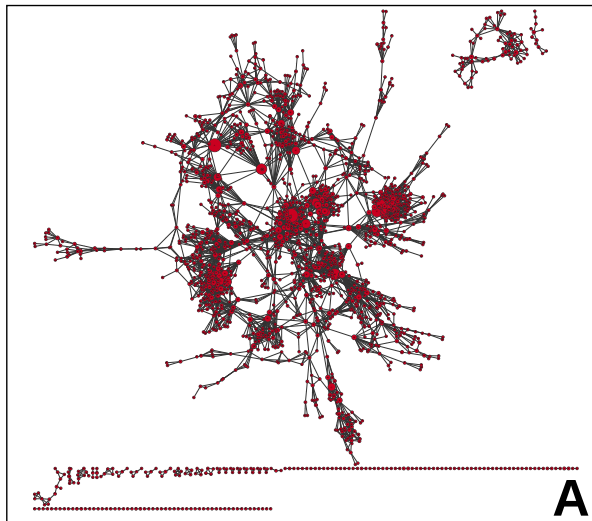
## Гипотеза коды -r (List forthcoming)

- теория была впервые предложена С. А. Старостином (1989: 338-340), который предположил, что связи между рифмами с кодами на \*-n и \*-j объясняются добавочной кодой \*-r с регулярным рефлексом -n в среднекитайском,
- дополнительные доказательства можно найти в родственных с китайском словах Тибетского (Hill 2014),
- Бакстер и Сагарт (2014) адаптировали эту идею в своей реконструкции древнекитайского, но количественная проверка до сих пор отсутствует.

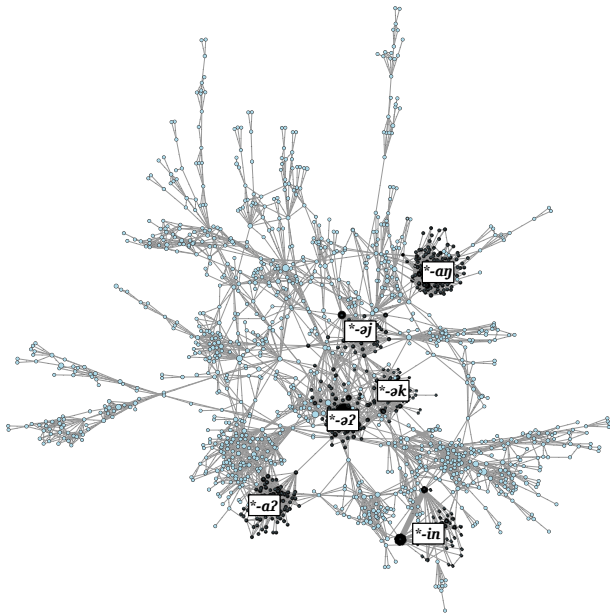
**Можно ли использовать граф «Ши цзин» для количественной проверки гипотезы Старостина?**



# Гипотеза коды -r: Граф «Ши цзин» в целом



# Гипотеза коды -r: Граф «Ши цзин» в целом



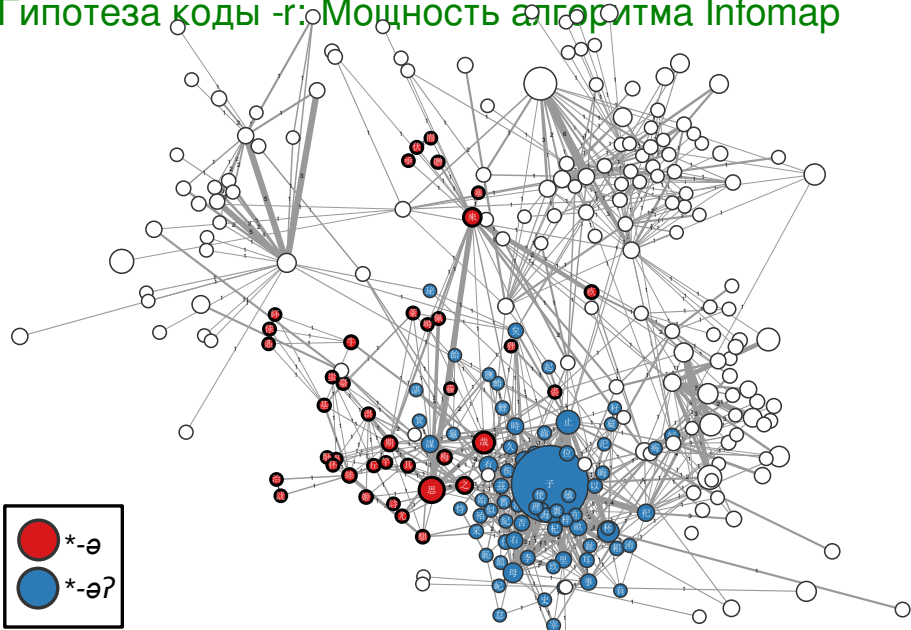
## Гипотеза коды -r: Сетевые сообщества

Рассмотрение графа в целом не достаточно для поиска древнекитайских код. Но разделение графа на части с помощью классических методов для выявления сетевых сообществ может дать нам новую перспективу. Если сообщества которые были идентифицированы автоматически отражают разбиение слов с кодами \*-j или \*-n на три класса, это может служить дополнительным доказательством для коды \*-r.

## Гипотеза коды -r: Сетевые сообщества

- выявление сетевых сообществ с помощью Infomap (Rosvall and Bergstrom 2007)
- алгоритм разделяет граф «Ши цзин» в 345 сообществ
- все данные можно осматривать в интерактивной манере:  
<http://digling.org/shijing/communities.html>

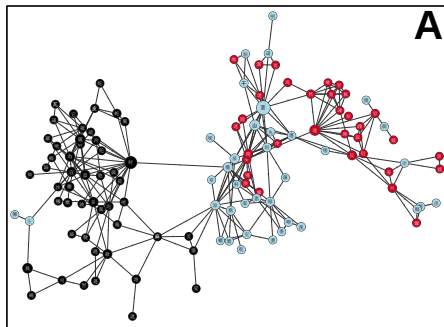
## Гипотеза коды -r: Мощность алгоритма Infomap



## Гипотеза коды -r: Мощность алгоритма Infomar

Если сравнить красную группу с синей группой, которые были идентифицированы алгоритмом Infomar, предложение очень близко к реконструкции Бакстера и Сагара (2014): из 74 слов которые были автоматически идентифицированы «красными», 59 реконструированы как -э? в системе Бакстера и Сагара. Из 39 слов, автоматически идентифицированных как «синие», 30 реконструированы как -э.

# Гипотеза коды -r: Коды на \*-a[nrj]



\*-an



\*-aj

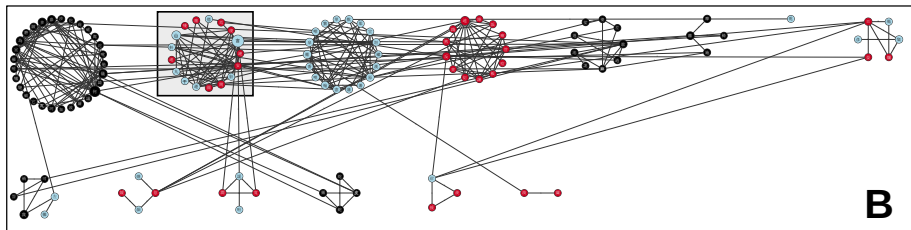


\*-ar



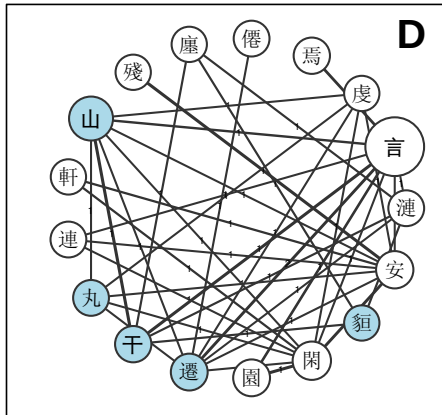
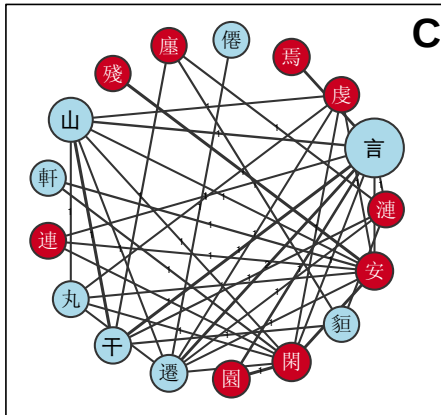
\*-an / \*-ar

# Гипотеза коды -r: Коды на \*-a[nrj]





## Гипотеза коды -r: Коды на \*-a[nrj]



## Гипотеза коды -r: Коды на \*-a[nrj]

Благодаря анализу сообществ в графе «Ши цзин», мы можем исправить реконструкцию Бакстера и Сагара и предложить коррекцию сомнительных реконструкций данной системы. Структура сообществ которые мы нашли с помощью Infomap указывает что кода -r С. А. Старостина оставила свои следы в рифмах «Книги песен». Дополнительные исследования нужны для включения автоматического анализа в реконструкцию древнекитайского языка.

# Перспектива



С. А. Старостин был пионером не только в классической, но и в автоматизированной лингвистической компаративистике. В свое время использовать компьютер было намного сложнее чем сегодня, и препятствий было много.

Сегодня, это все просто, но большинство лингвистов все еще избегает автоматизацию, в то время как неспециалисты анализируют лингвистические данные автоматически. Пора это изменить.

Сетевые методы — хороший пример для мощности количественных методов в исторической лингвистике. Но нельзя забывать что это только первый шаг: без лингвистического анализа, автоматические методы нас ничему не выучат.

**Спасибо за Ваше внимание!**