Apr 24, 2020

Digital model of spatio-temporal narratives of Chinese classical narrative literature

PLOS One

Zhaoyi Ma[1], Jie He[1], Shuai Shuai Liu[1]

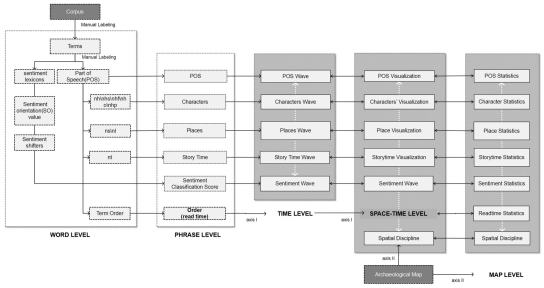
[1]School of Architecture, Tianjin University

1 Works for me dx.doi.org/10.17504/protocols.io.bdyni7ve

Zhaoyi Ma

ABSTRACT

In order to reconstruct the relationship between text and space in the tradition of Chinese classical narrative literature, a Time–Space-time–Space framework of digital models is posited, integrating spatial narrative theories and methods from literary cartography, computational narratology and geo-narrative. This Time–Space-time–Space framework enables varied structured semantics (including the spatio-temporal information) derived from the novel's text and its necessary historic context semi-manually, as the lack of mature natural language processing tools for ancient Chinese. These value-added structured semantics are then extracted and fused to map the instantaneous spatial pattern perceived by readers in the flow of reading time through data visualization, computational social science and spatial analysis. The organized spatio-temporal representations of narrative literature can help re-understand the narrative and the place. The Time–Space-time–Space framework embeds linearity, experience and significance of the narrative in an open and dialogic field.



Flows (black arrows) of variables and comparisons (white arrows) among variables in the logical loop of time–space–time–space.

EXTERNAL LINK

https://github.com/aayi/The-Tale-of-Li-Wa

THIS PROTOCOL ACCOMPANIES THE FOLLOWING PUBLICATION

Ma Z, He J, Liu S (2020) Representation of the spatio-temporal narrative of *The Tale of Li Wa*. PLoS ONE 15(4): e0231529. doi: 10.1371/journal.pone.0231529

MATERIALS TEXT

1. Electronically scanned version **S1 File.pdf** of *The Tale of Li Wa* in *Complete Library in Four Sections*
2. Raster map of Tang Chang'an with location information. **S2 File.zip**
   It can be added in Arcmap/Qgis. It is an archaeological map in -from Remote Sensing Analysis of Historical Landscape and GIS Laboratory, Northwest Institute of Historical Environment and Socio-Economic Development, Shaanxi Normal University GIS
3. . A chronicle of Bai Xingjian **31_.pdf**

BEFORE STARTING

This model only considers the narrative happening within a single city (each place can be mapped) although it may be applicable to other types of spatial narrative.

Digitalization

1 1.1 Text version from CText

1.2 Proofreading Text Edition **liwazhuang.txt** based on the version from *Complete Library in Four Sections*

Structuring

2 **Structuring of Narrative**

2.1 Text database on word level

S1 Table.xlsx (sheet1_name: *term*, sheet2_name: *POS*, sheet3_name: *so*, sheet4_name: *sentiment classification score*) we manually create this database including terms (unigram), parts of speech (POS), sentiment orientations (SO) value, and sentiment shifters by Excel.

1) *Term*

Citation: Zhaoyi Ma, Jie He, Shuai Shuai Liu (04/24/2020). Digital model of spatio-temporal narratives of Chinese classical narrative literature. https://dx.doi.org/10.17504/protocols.io.bdyni7ve

Segment of *Term* Criteria: Broadening, Dictionary, and Semantic Transparency
Reference dictionary: [illegible]
Reference Word Segmentation platform: [illegible]

2) *Part of Speech*

Tag	n	nt	nd	nl	nh	nhf	nhs	ns	nn	ni	no	nhh	v	vd	vl	vu	a	f	m	q	d	r	p	c	u	e	o
pos	Noun-general	Noun-time	Noun-direction	Noun-location	Noun-human	noun-last name	noun-first name	Noun-space	Noun-nation	Noun-institution	Noun-offical	noun-human's pronoun	Verb	Verb-direction	Verb-linking	Verb-auxiliary	adjective	difference	numeral	quantity	adverb	pronoun	preposition	conjunction	auxiliary	exclamation	or at ia

POS

3) *Sentiment oritention(so)*

The assignment of the SO value is as follows: each positive sentiment expression in the novel such as laugh ([illegible]) (v.) and magnificent ([illegible]) (a.) is given an SO value of +1 (172 in total), and each negative sentiment expression such as whimper ([illegible]) (v.) and poor ([illegible]) (a.) is assigned a SO value of −1 (177 in total).
We do two rounds of sentiment orientations (SO) value assignment(*LIU_SO value* and *MA_SO value*).
The percentage of consent of two rounds of SO value assignment is 81.5%.

Test about the hypothesis that the sentiment words have context-dependent orientations:

word2vec_python_code&raw_data **word2vec_python_code.zip**

sentiment_network_Gephi.zip **sentiment_network.zip**

result_weighted_outdegree_distribution.xlsx **result_weighted outdegree distribution.xlsx**

We assumed that the sentiment words in ancient Chinese follow the same logic in today's sentiment analysis–– sentiment words have context-dependent orientations, i.e., the total distance among words with the same orientation sentiment expression is closer than that among different.
Based on the unigrams removing stop words, we complete the training of its word2vec model with gensim package, and get the correlation(cosine_similarity, -0.4 ~ 1) between each two words.

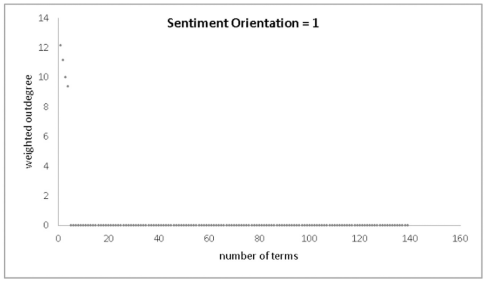
This correlation between word A (Source) and word B (Target) multiplied by the SO value of word B is taking as the weight of the edge, to set up a words' sentiment network in Gephi.

Source	Target	Weight
WordA	WordB	cosine_similarity*WordB_sentiment

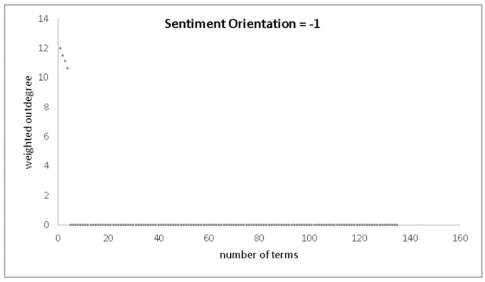
#tips for create *edge to gephi.csv*

#tips for operation step in Gephi:
File → Improt spreadsheet → *edge to gephi.csv* → charset: GB2312 → Graph Type: directed
run all analysis of "Avg. Weighted Degree" on the right manue
Data Laboratory → Nodes → Export Table#
The weighted output distribution calculated by gephi is divided into three different types of manually collected sentient words (1, - 1,0), which are made into scatter diagram by Excel.

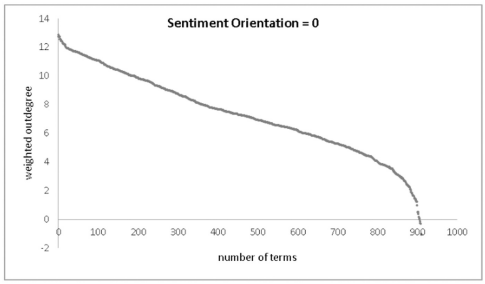
Citation: Zhaoyi Ma, Jie He, Shuai Shuai Liu (04/24/2020). Digital model of spatio-temporal narratives of Chinese classical narrative literature. https://dx.doi.org/10.17504/protocols.io.bdyni7ve



a



b



c

Weighted outdegree distribution of words' sentiment network (a. SO=1, b. SO=-1, c. SO=0)

4) *Sentiment classification score*

```
SO_value_effective = IF(sentiment_shifter_-1=-1,SO value * sentiment_shifter_-1,SO value)
for_phrase_sentiment_classification_score = IF(POS<>"w",SO_value_effective,"")
```

for_phrase_sentiment_classification_score = SUM(f*or_phrase_sentiment_classification_score*) #tips for operation step: Ctrl+G-> "Null(k)"--> "∑"(which means Automatic summation) #

```
for_phrase_sentiment_classification_score = SUM(for_phrase_sentiment_classification_score)
```


Citation: Zhaoyi Ma, Jie He, Shuai Shuai Liu (04/24/2020). Digital model of spatio-temporal narratives of Chinese classical narrative literature. https://dx.doi.org/10.17504/protocols.io.bdyni7ve

2.2 Text database on phrase level

S2 Table.xlsx (sheet1_name: *phrase*, sheet2_name: *time*, sheet3_name: *character*, sheet4_name: *character & SO*, sheet5_name: *place*, sheet6_name: *place & SO*)

The phrase-level framework assigns the recalculated value of POS and SO value to a relevant phrase by Excel. These values can be applied to the next time level because the sequence number of phrases is defined as read-time. Specific data mining approaches for the following parameters, i.e. places, story-time, and sentiment classification scores are valuable.

1) *Phrase*

sentiment_classification_score(SCS) inherits the value of *phrase_sentiment_classification_score* in S1 Table.xlsx.

2) *Time*

storytime_day

Noun-time (nt.) such as the Tianbao period (□□), 10 years later (□□), more than a month later (□□), and another day (□□), which is 2.7% of the total texts, are used to simulate the whole story-time in an interval of every single day. The entire story timeline we constructed from the texts started from when Student Zheng entered Chang'an in 747A.D. and ended around the happy ending of the novel, that is, the year Zheng is appointed to become an officer is 754 A.D., and Li Wa is conferred the title Lady Qian'guo (□□□□) in 775 A.D. The story-time is defined by the exact time record of the story that occurred during the period of 742 to 746 A.D.(□□□□), Bai Xingjian wrote the tale in August of 795 A.D. (□□□.....□□□□□□□) and the nt. phrases.

readtime_phrase

the sequence number of phrases is defined as read-time

3) *Character*

character1、*character2*、*character3*、 and *character4* contain one character in each phrase(since one phrase contains at most 4 characters).

4) *Character & SO*

SO_value_Integral_function(so_IF)= SUM(*sentiment_classification_score(SCS)*) #tips for operation step:

```
E15=SUM(D$2:D15), E18=SUM(D$2:D18), E50=SUM(D$2:D50)...
#
```

```
ZHENG_SCS = IF(character1="□□" or character2="□□" or character3="□□" or character4="□
□",sentiment_classification_score(SCS),"")
```

ZHENG_so_IF= SUM(*ZHENG_SCS*) #tips for operation step:

```
K15=SUM(J$15:J15), K18=SUM(J$15:J18), K50=SUM(J$15:J50)...
#
```

5) *Place*

A noun-space (ns.) such as Chang'an City, and specific place names inside the city such as the Buzheng Ward (□□□) and Xingyuan Garden (□□) (located in Tongshan Ward [□□□]), account for 1.1% of the total texts tagged as the level of residential wards and streets directly mentioned (e.g., Buzheng Ward) or most likely to be located (e.g., Tongshan Ward). These uniformly fine-grained places are applied to cover the corresponding story phrases of which plot takes place in these places.

6) *Place & SO*

```
Anyi_SCS = IF(ward_in_chang'an="□□□",sentiment_classification_score(SCS),"")
```

Anyi_so_IF= SUM(*Anyi_SCS*) #tips for operation step:

```
I526=SUM(H$526:H526), I535=SUM(H$526:H535), I606=SUM(H$526:H606)
#
```

2.3 Chronicle of Bai Xingjian

S3 Table.xlsx (sheet1_name: *circumstance*, sheet2_name: *poems*) is bassed on A chronicle of Bai Xingjian collated by □ □□

Detail contians Bai's specific experience every year

circumstance_orientation_value is assigned manually based on the good/bad of *Detail*. Such as "□□□□□□□□□□(Bai's grandmother died)"is assigned a value of −1, "□□□□□□□□...□□□□□...□□□□(Bai passed the Imperial Examination...Bai passed passed the Palace Examination)" is assigned a value of +2.

Circumstances_of_Bai = SUM(*circumstance_orientation_value*) #tips for operation step:

```
E2=SUM($D$2:D2), E10=SUM($D$2:D10), E52=SUM($D$2:D52)...
#
```

```
circumstance_orientation_value_chang'an = IF(Place="□□",circumstance_orientation_value,"")
```

2.4 Spatial syntax of Chang'an

S4 File.graph

Vector file of street of chang'an is created by Autocad and then imported into *Depthmap*(a technology used to analyze the spatial layouts, and human activity patterns in urban areas)

#tips for operation step of Integration analysis of the road network of Chang'an city by Depthmap:
Using Autocad to depict the main road axis map of Chang'an map(Vector file of street of chang'an) → Save as *dxf* file → Open the depthmap software and create a new workspace → Map-import-Choosing Chang'an Road Axis Chart → ap-convert drawing map → tools-axial/convex/pesh-run graph analysis-Radius/list of radii – input n,2,3,5,7-choose include choice□betweenness□/local measures/RA,RRA and total depth/weighted measures-length#

The degree of integration (a space syntax parameter) reflects the ease of access to streets, that is, it may determine which street is more likely to attract Zheng, as an explorer of Chang'an.

Citation: Zhaoyi Ma, Jie He, Shuai Shuai Liu (04/24/2020). Digital model of spatio-temporal narratives of Chinese classical narrative literature. https://dx.doi.org/10.17504/protocols.io.bdyni7ve

2.5 ## Spatially embedded semantic data

S3 File.zip (**path.xlsx**, **place.xlsx**, **shikong-vt.shp**, **link_between_places.shp**) combines **Text database** with spatial data by Arcmap.

1) path.xlsx

create **path.xlsx** semi-manually by Excel based on **Text database on phrase level**
Each time when *ward_in_chang'an* in sheet "place & SO" changes, an *ID* is added in **path.xlsx** with *Origin* to *Destination*
sentiment score between places: value difference of *sum_sentiment_effective_classification_score* between *Origin* and *Destination*
straturm classification: The classification of social *stratum* of the *character*. The classification of social stratum in the story from untouchable to nobles is as follows: (1) beggar, servant, and sex worker; (2) businessman, civilian, madam; (3) ward head (坊), candidate student, successful candidate, county judicial official (縣); and (4) Chang'an officials (京) and officials from other places.

2) place.xlsx

create **place.xlsx**(sheet1_name: *stratum_statistics*, sheet2_name: *place_statistics*) by Excel based on **Text database on phrase level**
stratum_statistics inherits the value in **path.xlsx**

place	sum_sentiment_effective_classification_score	phrase_count	averge_sentiment_classification_score	effective_sum_sentiment_classification_score	effective_phrase_count	averge_sentiment_effective_classification_score	averge_stratum	STDEV_stratum	COUNT_stratum
ward	SUM(place_SCS) in sheet" place & SO"	COUNT(place_SCS) in sheet" place & SO"	AVERAGE(place_SCS) in sheet" place & SO"	IF place_SCS<>0,SUM(place_SCS),"" in sheet" place & SO"	IF place_SCS<>0,COUNT(place_SCS),"" in sheet" place & SO"	IF place_SCS<>0,AVERAGE(place_SCS),"" in sheet" place & SO"	AVERAGE(place_stratum) in sheet" stratum_statistics"	STDEV(place_stratum) in sheet" stratum_statistics"	COUNT(place_stratum) in sheet" stratum_statistics"

tips for operation step in *place_statistics*

3) Polygon

#tips for operation step:
create shapefile of **Polygon**(Ward& Palace) based on **Raster map of Tang Chang'an with location information**#
This shapefile is not uploaded to github due to copyright issues

4) Point

shikong-vt.shp (in S3 File.zip)
#tips for operation step:
·create **Point**(centroid of **Polygon**)--> add field *ward_in_chang'an* and fill in
·**Text database on phrase level** is joined with **Point** by field *ward_in_chang'an*#

LIWA_data.shp **S6 File.zip**
#tips for operation step:
·create **Point**(centroid of **Polygon**)--> add field *ward_in_chang'an* and fill in
·**place.xlsx** is joined with **Point** by field *NAME*#

5) Polyline

temporal simulation path in space.shp **S5 File.zip**
#tips for operation step:
Arcmap--> Toolbox--> XY to Line--> import **shikong-vt.shp**#

link_between_places.shp (in S3 File.zip)
·create **polyline** The rules of the simulation are from the characteristics of the streets: first, prefer the shortest path, and second, prefer the street sections with the highest degree of spatial syntax integration.
·**path.xlsx** is joined with **polyline** by field *ID*

Representation

3 **Spatio–temporal Representations of Narrative**

Citation: Zhaoyi Ma, Jie He, Shuai Shuai Liu (04/24/2020). Digital model of spatio-temporal narratives of Chinese classical narrative literature. https://dx.doi.org/10.17504/protocols.io.bdyni7ve

3.1 Time

#tips for operation step of **Trajectory of the integral function of SO value by sigmaplot**:
create graph--> simple straight line--> data format--> XY Pair--> select data
data for X: *storytime_day* data for Y: *SO_value_Integral_function(so_IF)*#

#tips for operation step of **Trajectory of the integral function of SO value and characters' appearance by sigmaplot**:
create graph--> multiple straight line--> data format--> XY Pair--> select data
data for X: *storytime_day* data for Y: *SO_value_Integral_function(so_IF)*
data for X: *storytime_day* data for Y: *ZHENG_so_IF*
data for X: *storytime_day* data for Y: *LI_Wa_so_IF*
data for X: storytime_day data for Y: *LI_Wa's_mother_so_IF*
data for X: *storytime_day* data for Y: *ZHENG's_father_so_IF*#

#tips for operation step of **Trajectory of the integral function of SO value and places' appearance by sigmaplot**:
create graph--> simple straight line--> data format--> XY Pair--> select data
data for X: *storytime_day* data for Y: *SO_value_Integral_function(so_IF)*

add new plot--> graph types--> vertical bar chart--> graph styles--> grouped bars--> data formats--> many Y
data for Y: *Anyi_so_IF*
data for Y: *Buzheng_so_IF*
data for Y: *Chongren_so_IF*
data for Y: *EastMarket_IF*
data for Y: *Pingkang_IF*
data for Y: *DepartmentOfStateAffairs_IF*
data for Y: *TianmenStreet_so_IF*
data for Y: *Tongshan_so_IF*
data for Y: *Tongyi_so_IF*
data for Y: *WestMarket_so_IF*
data for Y: *XingqingPalace_IF*
data for Y: *Xuanyang_so_IF*

graph page--> add axist--> Y#

#tips for operation step of **Trajectory of the integral function of SO value versus the story-time's appearance by sigmaplot**:
create graph--> multiple straight line--> data format--> XY Pair--> select data
data for X: *storytime_day* data for Y: *SO_value_Integral_function(so_IF)*
data for X: *storytime_day* data for Y: *readtime_phrase*
graph page--> add axist--> Y#

#tips for operation step of **Bai Xingjian's up and down by sigmaplot**:
create graph--> vertical bar chart--> graph styles--> grouped bar--> data format--> many Y
data for X: *Age* data for Y: *Circumstances_of_Bai* data for X: *Age* data for Y: *Circumstances_of_Bai_Chang'an*#

3.2 Space-time

import **shikong-vt.shp**, **temporal simulation path in sapce.shp** and **link_between_places.shp** into Arcscene

#tips for operation step of **Visualization of the integral function of SO value versus places' appearance by Arcscene**:
"shikong-vt.shp"--> properties--> Element--> Single symbol "temporal simulation path in sapce.shp"--> properties--> Symbolic System--> Graded colour--> value--> *sheet1_em(SO_value_Integral_function(so_IF))*#

#tips for operation step of **Visualization of path trajectory based on spatial discipline: characters' appearance versus places' appearance by Arcscene**:
"shikong-vt.shp"--> properties--> Element--> Single symbol
"link_between_places.shp"--> properties--> Symbolic Systems--> category--> Unique value--> value--> *(character)*#

#tips for operation step of **Visualization of path trajectory based on spatial discipline: characters' appearance and the integral function of SO value versus places' appearance by Arcscene**:
"shikong-vt.shp"--> properties--> Element--> Single symbol "link_between_places.shp"--> properties--> Symbolic Systems--> Graded colour--> value--> *(sentiment score between places)*#

#tips for operation step of **Visualization of path trajectory based on spatial discipline: characters' appearance of different stratum versus places' appearance and the integral function of SO value by Arcscene**:
"shikong-vt.shp"--> properties--> Symbolic System--> Graded colour--> value--> *sheet1_em(SO_value_Integral_function(so_IF))*
"link_between_places.shp"--> properties--> Symbolic Systems--> category--> Unique value--> value--> *(stratum)*#

3.3 Space

POS map

#tips for operation step of **Statistics of different POS and places by Photoshop**:
create the wordcloud image of POS of each place by wordart --> imported in Photoshop --> move to relevant place --> export into png#

Sentiment map

#tips for operation step of **Inverse distance weighted (IDW) interpolation by ArcGIS of the sentiment classification score and places by Arcmap** :
Toolbox → geostatistical analyst- interpolation analysis – IDW→ input layer"(**LIWA_data.shp**)" → Environment-range-"

"#

Social network

Citation: Zhaoyi Ma, Jie He, Shuai Shuai Liu (04/24/2020). Digital model of spatio-temporal narratives of Chinese classical narrative literature. https://dx.doi.org/10.17504/protocols.io.bdyni7ve

S7 File.zip (edge.csv, node.csv)
if two *character* co-occurr within two adjacent phrases, one edge will be added between them.

#tips for operation step of **Statistics of characters in co-occurrence network, modularity class, betweenness centrality, and Girvan-Newman clustering** in Gephi:
File → Improt spreadsheet → *edge.csv* → charset: GB2312 → Graph Type: indirected
Tools → plugins → Availible Plugins → Install
run all analysis of "Statistics" on the right manue
appearence → nodes → color → Partition → Modularity class → run
appearence → nodes → color → Partition → Cluster-ID → run
appearence → nodes → size → Ranking → Betweenness centrality → run#
Network parameters:
node：26
edge：38
Average Degree: 1.462
Average Weighted Degree: 4.731
Diameter: 5
Radius: 1
Average Path length: 2.265625
Density: 0.095
Randomize: On (Modularity)
Use edge weights: Off (Modularity)
Resolution: 0.8 (Modularity)
Modularity: 0.351 (Modularity)
Modularity with resolution: 0.218 (Modularity)
Number of Communities: 6 (Modularity)
Number of communities:4 (*Girvan-Newman* Clustering)
Maximum found modularity:0.39129266 (*Girvan-Newman* Clustering)
Network Interpretation: undirectedAverage
Clustering Coefficient: 0.506
Total triangles: 6

Spatially embedded network

S8 File.zip (direct-node.csv, edge.csv, indirect-node.csv) is based on **path.xlsx**
#tips for operation step of **Network analysis of characters and places in Girvan-Newman clustering analysis, modularity class analysis (use edge weights: On) and weighted degree centrality based on a full-text, spatially embedded, undirected network of characters** in Gephi:
File → Improt spreadsheet → *edge.csv* → charset: GB2312 → Graph Type: indirected
Tools → pluginss → Availible Plugins → Install
run all analysis of "Statistics" on the right manue
appearence → nodes → color → Partition → Cluster-ID → run
appearence → nodes → color → Partition → Modularity class → run
appearence → nodes → size → Ranking → Weighed Degree → run#

#tips for operation step of **Network analysis of characters and places in closeness centrality and betweenness centrality based on a full-text spatially embedded undirected network of characters** in Gephi:
File → Improt spreadsheet → *edge.csv* → charset: GB2312 → Graph Type: indirected
run all analysis of "Statistics" on the right manue
appearence → nodes → color → Ranking → Closeness centrality → run
appearence → nodes → size → Ranking → Betweenness centrality → run#

#tips for operation step of **Network analysis of characters and places in authority and hub analysis based on a full-text spatially embedded directed network of characters** in Gephi:
File → Improt spreadsheet → *edge.csv* → charset: GB2312 → Graph Type: directed
run all analysis of "Statistics" on the right manue
appearence → nodes → color → Ranking → Authority → run
appearence → nodes → size → Ranking → Hub → run#
Network parameters:
node：16
edge：24
Average Degree: 3.000
Average Weighted Degree: 7.750
Diameter: 6
Radius: 3
Average Path length: 2.265625
Density: 0.200
Randomize: On (Modularity)
Use edge weights: On (Modularity)
Resolution: 0.8 (Modularity)
Modularity: 0.326 (Modularity)
Modularity with resolution: 0.194 (Modularity)
Number of Communities: 4 (Modularity) Number of communities:4 (*Girvan-Newman* Clustering)
Maximum found modularity:0.3949653 (*Girvan-Newman* Clustering)
Network Interpretation: undirectedAverage
Clustering Coefficient: 0.474
Total triangles: 9

Citation: Zhaoyi Ma, Jie He, Shuai Shuai Liu (04/24/2020). Digital model of spatio-temporal narratives of Chinese classical narrative literature. https://dx.doi.org/10.17504/protocols.io.bdyni7ve