Physica A 518 (2019) 169–176

Contents lists available at ScienceDirect

Physica A

journal homepage: www.elsevier.com/locate/physa

A study of Chinese regional hierarchical structure based on surnames

Yongbin Shi [a,b], Le Li [a,b], Yougui Wang [a,b,*], Jiawei Chen [a,b,*], H. Eugene Stanley [b]

[a] School of Systems Science, Beijing Normal University, Beijing 100875, PR China
[b] Center for Polymer Studies and Physics Department, Boston University, Boston, MA 02215, USA

HIGHLIGHTS

- We construct a network of Chinese provinces whose edges are isonymic distances.
- A hierarchical tree of Chinese provinces is derived from isonymic distance matrix.
- The geographical features of these hierarchical structures are exhibited in map.
- Our empirical findings support the Tobler's First Law of Geography.

ARTICLE INFO

Article history:
Received 13 July 2018
Received in revised form 23 October 2018
Available online 4 December 2018

Keywords:
Surname
Isonymic distance
Minimum spanning tree
Single linkage cluster analysis
Hierarchical structure

ABSTRACT

We use isonymic distance to measure the dissimilarity in surname structure between populations of Chinese provinces, and we employ the minimum spanning tree (MST) and the single linkage cluster analysis (SLCA) to investigate the hierarchical structure of Chinese provinces and present its corresponding geographical features. We find diverse discrepancies in the averaged isonymic distance among provinces that are attributed to the heterogeneous surname distributions. The MST displays a core–fringe structure with Henan, Anhui, and Hubei making up the core, and several border provinces on the fringe. The degree centrality list in the MST reveals some "local centers" that act as regional economic centers. On the other hand, the geographical layout of MST reflects the historical "Rush to Northeast" mass migration, as well as the blocking effect of the Qinling–Huaihe line that separates north and south China. The clustering results derived from the SLCA show nine groups of provinces in which each group is geographically continuous.

1. Introduction

Over the last two decades, complex networks have been applied to a wide variety of research areas, many of which are interdisciplinary [1,2]. A complex network, a system of edges that connect nodes, can be used to describe interactions among individuals in social systems [3], such as airport networks [4], online social networks [5], information diffusion [6,7], and collaboration and citation networks [8].

Surnames are a data source that can be used to construct networks, and they have attracted the attention of researchers in anthropology, geography and complexity science. Two types of naming networks have been examined in recent years, (i) a bipartite network of forenames and surnames [9,10], and (ii) a network of surnames in which nodes are surnames and edges are the co-occurrence of surnames [11] or the parental relations between two families [12]. These studies indicate

* Corresponding authors at: School of Systems Science, Beijing Normal University, Beijing 100875, PR China.
E-mail addresses: ygwang@bnu.edu.cn (Y. Wang), chenjiawei@bnu.edu.cn (J. Chen).

https://doi.org/10.1016/j.physa.2018.11.059
0378-4371/

Table 1
The five highest/lowest ranking provinces, according to population, surname, and population/surname measures.

	Population		Surname		Population/Surname	
	Province	Score	Province	Score	Province	Score
The five highest ranking provinces	Henan	100 964 803	Yunnan	4361	Shandong	26 981.84
	Shandong	92 628 654	Anhui	4329	Guangdong	26 835.10
	Sichuan	86 600 969	Sichuan	4258	Henan	24 571.62
	Guangdong	79 083 035	Henan	4109	Jiangsu	23 918.89
	Jiangsu	72 187 196	Hubei	3995	Hunan	20 637.08
The five lowest ranking provinces	Tianjin	9 484 351	Ningxia	2016	Xinjiang	5 390.99
	Hainan	8 082 421	Beijing	1941	Hainan	4 561.19
	Ningxia	5 856 669	Hainan	1772	Ningxia	2 905.10
	Qinghai	5 007 531	Tianjin	1600	Qinghai	1 751.10
	Xizang	2 578 176	Shanghai	1559	Xizang	1 040.43

that a network representation of surname data has great potential for application in social, cultural, ethnic, migration and geographic studies.

Our goal here is to extend the network presentation of surname data to a spatial network and to investigate the Chinese regional hierarchical structure and geographical features behind the geographical distribution of surnames. In surname study, isonymy and its associated isonymic distances are commonly used to measure the similarity (or dissimilarity) in the surname structures of geographical locations and grouped populations and to use this as input data for constructing networks and developing clustering techniques [13]. As cheap substitutes for Y chromosome typing, surnames are a credible data source for classifying geographical population [14,15]. Because of traditional cultural constraints [16] and the long history of hereditary surname [17], Chinese surnames are an excellent data source for identifying region clusters.

We here use a large sampling of surname data to construct a spatial network of Chinese provinces. Network nodes are provinces, and network edges are defined by isonymic distances between two connected nodes. We use a minimum spanning tree (MST) to construct the spatial network. An MST filters information and has been used to construct networks in financial markets [18–21], the human brain [22,23], and transportation systems [24]. We also use single linkage cluster analysis (SLCA) to investigate the hierarchical structure of Chinese provinces and their geographical features. Section 2 of this paper describes the Chinese surname data sets and introduces the MST and SLCA algorithms. Section 3 presents the empirical results, and discusses their significance. Section 4 lists and discusses our conclusions.

2. Data and methodology

2.1. Data

We obtained data for this study from China's National Citizen Identity Information Center (NCIIC), which includes the surnames and administrative regions at a provincial level of all Chinese people who officially registered in the NCIIC in 2007. This includes 1.28 billion entries and encompasses 31 provincial administrative regions (22 provinces, 5 autonomous regions, and 4 municipalities), among which Hongkong, Macao, and Taiwan are excluded.

There are 7184 surnames in the sampling data. Table 1 shows the population and surname related indicators of a number of provinces. The average population of one surname in a province is determined by demography. Except for Jiangsu, the five most-populous provinces are the top five provinces with the highest population per surname. The situation in the five least-populous provinces is similar.

2.2. Isonymic distance

For province i, N_i is the population, and n_{ki} is the number of people with surname k where $\sum n_{ki} = N_i$. The share of population with surname k in province i is defined by

$$p_{ki} = \frac{n_{ki}}{N_i}. \tag{1}$$

We use the Euclidean distance [25] developed in isonymic study to quantify the dissimilarity in surname structure between two provinces, which can be mathematically expressed as

$$ED_{i,j} = \sqrt{1 - \sum_{k=1}^{S} \sqrt{p_{ki} p_{kj}}}, \tag{2}$$

where S is the number of surnames, and i and j are two provinces. The value of ED lies between 0 and 1. when $ED = 1$ there is no common surname between two provinces, indicating extreme dissimilarity. The more similar the surname structure, the smaller the value of ED. In the extreme case when two areas have exactly the same surname structure, where p_k is the

Table 2
The average *ED* of all provinces. The provinces are sorted according to the ascending order of average *ED*.

Rank	Province	Average ED	Rank	Province	Average ED
1	Beijing	0.3081	17	Chongqing	0.3733
2	Henan	0.3197	18	Yunnan	0.3750
3	Shaanxi	0.3267	19	Shanxi	0.3874
4	Tianjin	0.3291	20	Ningxia	0.3888
5	Liaoning	0.3316	21	Hunan	0.4020
6	Helongjiang	0.3323	22	Shanghai	0.4024
7	Anhui	0.3330	23	Jinagxi	0.4109
8	Hebei	0.3376	24	Zhejiang	0.4229
9	Shandong	0.3406	25	Guangdong	0.4535
10	Jilin	0.3419	26	Qinghai	0.4703
11	Gansu	0.3469	27	Hainan	0.4716
12	Hubei	0.3471	28	Fujian	0.4753
13	Sichuan	0.3576	29	Guangxi	0.5110
14	Neimenggu	0.3598	30	Xinjiang	0.5915
15	Jiangsu	0.3624	31	Xizang	0.8348
16	Guizhou	0.3680			

NOTE: The average ED is the mean of all EDs between corresponding province and other ones.

same for all individual surnames in the two provinces, $ED = 0$. In addition to the Euclidean distance, the Lasker [26] and Nei distances are [27] also popular isonymic measurements. However, Rodriguez-Larralde et al. [28] and Dipierri et al. [29] show that the Euclidean distance is better than these other two when few surnames are shared between two populations.

The spatial distribution of surnames is highly heterogeneous in China [30]. There are many localized surnames, especially in ethnic minority areas. Table 2 shows the distinct differences in the average *ED* among provinces. In general, the average *ED* of the northern provinces is smaller than that in the southern provinces, and smaller in the east than in the west. Some provinces have extremely high average *ED*, such as Guangxi, Xinjiang and Xizang, and this indicates that the surname structures in these provinces differ greatly from those in other provinces. This indicates that the Euclidean distance more accurately characterizes the dissimilarity in surname structure between these two groups.

2.3. Minimum spanning tree (MST) and single linkage cluster analysis (SLCA)

We calculate the $ED_{i,j}$ for all pairs of provinces and get the 31×31 symmetric distance matrix (denoted by D), which we use as input data to construct the network. We have a weighted network $G(V, E, W)$, in which $V = (v_1, v_2, \ldots, v_n)$ is the set of nodes, $E = (e_1, e_2, \ldots, e_m)$ is the set of edges, and $W = (\omega_1, \omega_2, \ldots, \omega_m)$ is the weight set of edges. The edges connect nodes. A network constructed by D is fully-connected, has too many redundant connections, and the input topology is thus vague [31].

The minimum spanning tree (MST) is a sub-network in which a set of $n - 1$ edges connects all nodes without any cycles such that the sum of the edge weights is minimal [32]. To build an MST network we find the $n - 1$ most relevant connections among $n(n - 1)/2$ connections. There are two primary methods of constructing an MST: the Prim algorithm [33] and the Kruskal algorithm [34]. We here use the Prim algorithm to construct the spatial MST network of Chinese provinces. The Prim algorithm begins with an empty spanning tree and repeatedly picks the minimum weight edge from the remaining edges until all nodes are included [35]. Using MST allows us to both connect all nodes and extract key information such that all provinces are included and the most important linkages derived from the distance matrix. The MST method is also closely related to the single linkage cluster analysis (SLCA) [35].

We use SLCA to further analyze the surname-based hierarchical structure of provinces. It is an agglomerative clustering method that starts with n clusters and then sequentially combines the clusters into larger clusters until all elements are in one cluster. All the information necessary to complete SLCA is contained in the process of constructing MST [36]. Tumminello et al. applied the MST and SCLA to the financial market, and found that some connections between clusters were lost in MST presentation [37]. Thus SLCA is a more intuitive hierarchical organization showing the hierarchical tree than the MST.

3. Results and discussions

3.1. Network presentation

We use isonymic distance to construct the spatial MST of Chinese provinces. Fig. 1 shows this spatial network by overlaying it on a map of China, where the network nodes are located according to their geographical positions. The spatial MST has two backbones, the Qinghai–Gansu–Shaanxi–Henan–Beijing in the north and the Sichuan–Chongqing–Hubei–Anhui–Jiangsu in the south. The link between Henan and Anhui is their only connection. The spatial MST reveals the relevant connections among provinces, and we see that the surname structure within northern and southern provinces is more similar than that between the north and south. The only direct linkage between the north and the south is the Henan–Anhui connection, and thus Henan–Anhui is the main migratory route that connects the north and south of China. These

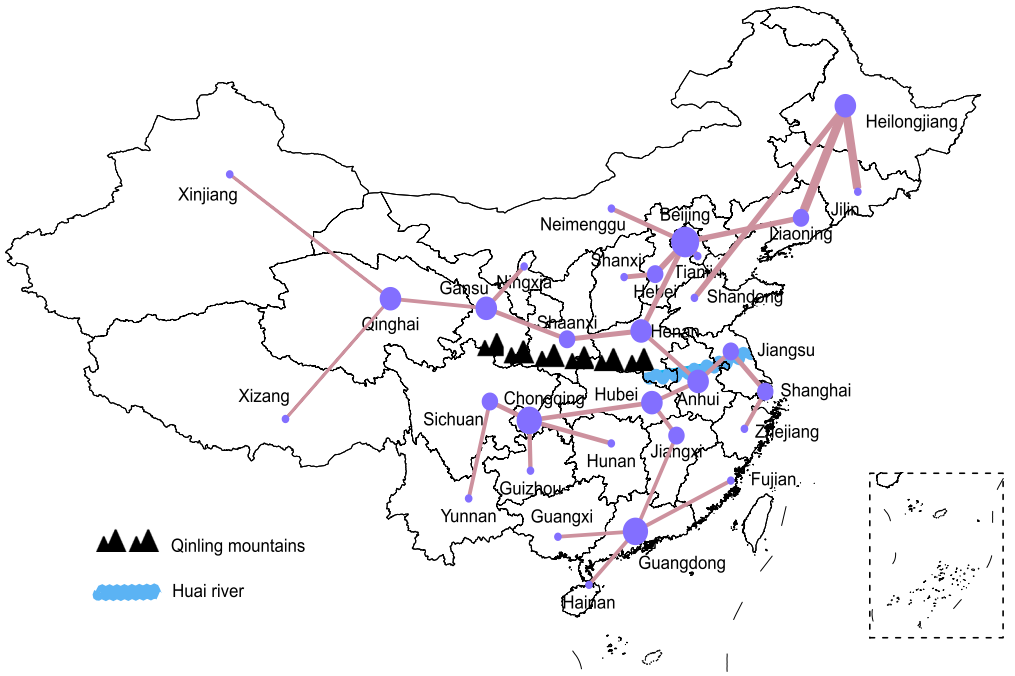

Fig. 1. A geographical presentation of the spatial MST of Chinese provinces. The nodes represent the provinces, and the size of nodes is proportional to the degree of nodes. The edges are defined by the Euclidean distance between the two connected provinces, and the thickness of an edge is proportional to its weight. The thicker the edge, the smaller the Euclidean distance.

findings confirm the geographical definition of the Qinling–Huaihe line. The Qinling–Huaihe line runs west to east, consists of the Qinling mountains in west and the Huai river in the east, and geographically divides China into northern and southern regions [38]. Unlike the Qinling mountains, which blocks human mobility between the northern and the southern areas, the Huai river promotes it.

The Qinling mountains, a composite continental orogenic belt, have been formed by the collision between the North China Block (NCB) and the South China Block (SCB) [39–42]. Its latitude is sufficiently high to fend off the cold current from the north and the warm moist current from the south [43]. Thus the climates of the two sides differ, and this difference greatly hinders migration. As a result, the people on either side have contrasting lifestyles and differing surname structures. Fig. 1 shows this blocking and the lack of connections between the provinces on either side of the Qinling mountain range. In contrast to this blocking effect, the Huai river facilitates human mobility between the provinces on either side. The Huai river is east of the Qinling mountains, and its terrain is low-lying vast plains [44]. It flows through the south of Henan and the north of Anhui, and facilitates human mobility between these two provinces.

Note that most of the two-connected nodes in the spatial network are geographical adjacent, reflecting the Tobler's First Law of Geography that "everything is related to everything else, but nearby things are more closely related than distant things" [45]. Note that there are two links that do not connect geographically adjacent provinces, the Heilongjiang–Shandong and the Heilongjiang–Liaoning. These two outliers were due to the famous "Rush to Northeast" mass migration when tens of thousands of people in Shandong left their homeland and went to three northeastern provinces: Heilongjiang, Jilin and Liaoning. This mass migration started from the early Qing dynasty, and was intermittent until the end of the last century, lasting for about 300 years. Nowadays, many citizens of northeastern provinces can trace their ancestries to Shandong. From Fig. 1, we can see that Shandong and Heilongjiang are geographically distant from each other, but connected in the spatial network. A large influx of migrants from Shandong to the northeastern provinces made the surname structure of Shandong similar to those of northeastern provinces. However, the surname structures of these northeastern provinces are more similar to each other than to their originating province, resulting in two thickest edges of Heilongjiang–Jilin and Heilongjiang–Liaoning in the spatial network. Indeed, the three northeastern provinces are culturally unified, and people more strongly identify themselves as "Northeasterner" than citizens of an individual province. In general, the network edges are thicker in the north than in the south, which is attributed to the obvious difference in geographic and geomorphologic characteristics. The land is flat in northern China, while the terrain is mountainous in southern China.

3.2. The importance of nodes

A number of centrality measures in terms of topological characteristics are used to quantify node importance. We here examine three classical metrics: degree centrality, betweenness centrality, and closeness centrality. Each metric quantifies

Table 3
Degree centrality (*DC*), betweenness centrality (*BC*), and closeness centrality (*CC*) of 31 Chinese provinces.

Provinces	*DC*	*BC*	*CC*	Provinces	*DC*	*BC*	*CC*
Beijing	5[a]	394	0.935	Hubei	3	450[a]	0.962[a]
Tianjin	1	0	0.735	Hunan	1	0	0.649
Hebei	2	58	0.746	Guangdong	4[a]	168	0.676
Shanxi	1	0	0.613	Guangxi	1	0	0.565
Neimenggu	1	0	0.735	Hainan	1	0	0.565
Liaoning	2	162	0.769	Chongqing	4[a]	218	0.8
Jilin	1	0	0.543	Sichuan	2	58	0.658
Heilongjiang	3	114	0.645	Guizhou	1	0	0.649
Shanghai	2	58	0.68	Yunnan	1	0	0.552
Jiangsu	2	112	0.833	Xizang	1	0	0.532
Zhejiang	1	0	0.568	Shaanxi	2	250	0.885
Anhui	3	514[a]	1.053[a]	Gansu	3	214	0.746
Fujian	1	0	0.565	Qinghai	3	114	0.629
Jiangxi	2	208	0.8	Ningxia	1	0	0.613
Shandong	1	0	0.543	Xinjiang	1	0	0.532
Henan	3	558[a]	1.064[a]				

DC, *BC*, and *CC* refer to degree centrality, betweenness centrality, and closeness centrality.
[a]Represents the scores ranking at the top three.

an aspect of node importance. Degree centrality measures the number of links that a node has. Degree centrality is defined by

$$DC(e_i) = k(e_i), \tag{3}$$

where $k(e_i)$ is the degree of e_i. Betweenness centrality measures the number of shortest paths passing through a node. The betweenness centrality of e_i is

$$BC(e_i) = \sum_{j,k\neq i} \sigma(e_j, e_k|e_i), \tag{4}$$

where $\sigma(e_j, e_k|e_i)$ is the number of shortest paths between e_j and e_k that pass through e_i. Closeness centrality measures a given node's average distance from other nodes and is defined

$$CC(e_i) = \frac{1}{\sum_j d(e_i, e_j)}, \tag{5}$$

where $d(e_i, e_j)$ is the shortest path length between e_i and e_j.

Table 3 lists the results of these centrality metrics for all provinces. The provinces with the three highest degree centralities are Beijing, Guangdong and Chongqing, and their scores are 5, 4, and 4, respectively. In the spatial network, when a node has a higher degree centrality the surname structure of its province is highly similar to that of neighboring provinces. Most of edges in the spatial network connect two geographically adjacent provinces. Thus degree centrality can be used to detect "local centers". Fig. 1 shows that these three provinces are located in the north, south and southwest of China, are all regional social and economic centers, and all have frequent movements of population to and from neighboring provinces.

Betweenness centrality and closeness centrality are global centrality measurements that quantify node importance throughout the network [46]. Table 3 shows that Henan, Anhui and Hubei are the "global centers" that rank highest in betweenness and closeness centrality. Their betweenness centrality are 558, 514, and 450, respectively, and their closeness centralities 1.064, 1.053, and 0.962, respectively. They are also geographically adjacent and connected through Anhui in the spatial network. All these facts imply that their population may be the origin of modern Chinese people. The river basins in the middle-lower reaches of the Yellow and Yangze rivers are considered the cradle of Chinese civilization [47,48]. These three provinces are the global centers in the spatial network located in the middle-lower sections of these two rivers. These findings suggest clues in the search for the geographical origin of Chinese culture.

3.3. Hierarchical structure (single linkage cluster analysis)

Fig. 2 shows the hierarchical tree obtained from the distance matrix of Eq. (2) when SLCA is applied. Comparing Figs. 1 and 2, we find that all directly connected provinces in the hierarchical tree are connected in the spatial MST. This indicates a strong link between SLCA and MST algorithms. The hierarchical tree exhibits the typical hierarchical structure of Chinese provinces based on their surname structure in which most of the provinces on the bottom have similar surname structures, and several on the top have differing structures.

A cut-point is set to determine the number of clusters. If it is too large, we get few clusters, if it is too small, some similar provinces are divided into different clusters. Here we set the value of cut-point to be 0.3. Fig. 2 shows the resulting six clusters

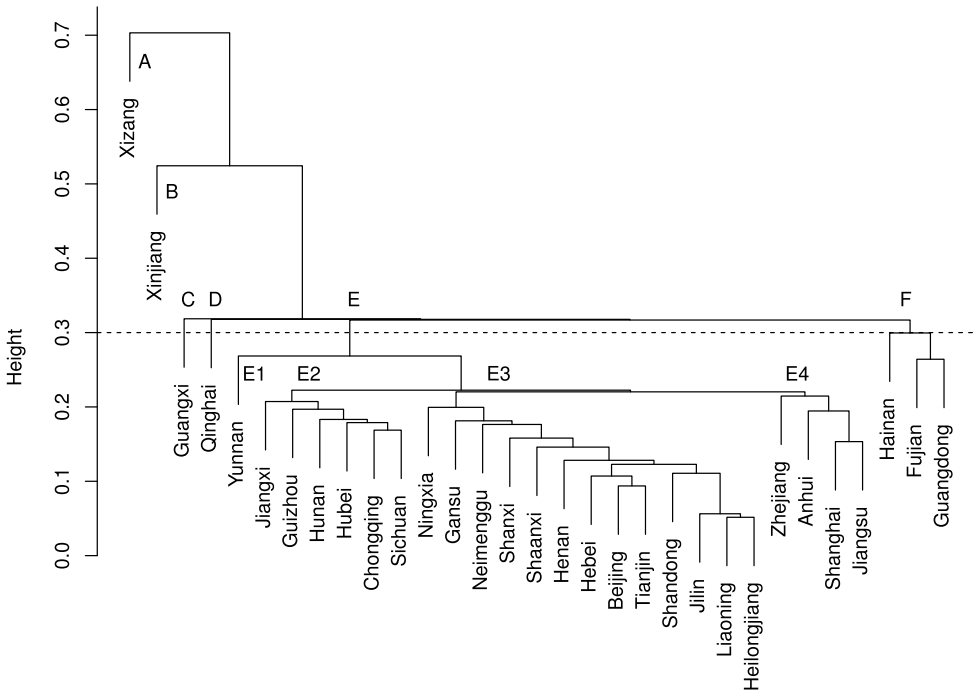

Fig. 2. Cluster dendrogram for Chinese provinces. The dendrogram is obtained based on the single linkage cluster analysis and the distance matrix computed using the Euclidean distance as a measure of dissimilarity in surname structure.

marked A through F, where a dashed line is drawn at the cut-off point of 0.3. The results show a super group (E) of provinces with similar surname structures. We further subdivide cluster E into four sub-clusters and label them E1 to E4. We find one large cluster and three large sub-clusters: F (Hainan, Fujian and Guangdong), E2 (Jiangxi, Guizhou, Hunan, Hubei, Chongqing and Sichuan), E3 (Ningxia, Gansu, Neimenggu, Shanxi, Shaanxi, Henan, Hebei, Beijing, Tianjin, Shandong, Jilin, Liaoning and Heilongjiang), and E4 (Zhejiang, Anhui, Shanghai and Jiangsu).

Fig. 3 shows a map of Chinese provinces in which colors indicate clustering results. We find that the geographical surname distribution of the Chinese population exhibits regional features. The largest group is in northern and northeastern China, and the second largest group is located in central and southwest China. These two groups are also in the north and south beyond the Qinling–Huaihe line. The E4 and the F are in eastern and southern China. The rest of the groups are minority areas and each comprises only one province. These findings indicate that there are two large areas (E2 and E3) in which the provincial populations share similar surname structures. There are also two middle-size areas (E4 and F) and five minority areas (A, B, C, D, E1) in which the surname structures differ from each other and from other provinces.

4. Conclusion

We have analyzed the hierarchical structure of Chinese administrative regions at the provincial level using a large sampling of surname data from NCIIC surveyed in 2007 as input data. We first use a minimum spanning tree algorithm to construct a spatial network of Chinese provinces in which the edges are defined by their isonymic distance. We next perform a single linkage cluster analysis of the isonymic distance matrix and obtain a hierarchical tree. Both MST and SLCA exhibit the hierarchical structure of Chinese provinces and their corresponding geographical features. The MST distinguishes "local centers" and "global centers", and the SLCA classifies the provinces into nine groups.

Our empirical results support the Tobler's First Law of Geography in that most connected provinces in the spatial MST are geographically adjacent, and all clusters derived from SLCA are geographically continuous in the map. A network presentation of surname data not only helps us identify the "local centers" and "global centers" of China, but also provides evidence of the historical "Rush to Northeast" mass migration and the blocking effect of Qinling–Huaihe line. These results indicate that the MST and SLCA methods are powerful ways of revealing the hierarchical structures of regions behind surname data.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (Grants No. 61773069 & No. 71731002) and the National Social Science Foundation of China (Grant No. 14BSH024). It was also supported by NSF, USA Grants PHY-1505000, CMMI-1125290, and CHE-1213217, and by DTRA, USA Grant HDTRA1-14-1-0017 and DOE, USA Contract DE-AC07-05Id14517.

Fig. 3. Map of China showing locations of all provinces, where the provinces in one cluster share the same color. Note the number of provinces in each group is listed in the brachet following the corresponding label.

References

[1] A.-L. Barabási, The network takeover, Nat. Phys. 8 (1) (2011) 14.
[2] M. Newman, Networks: An Introduction, Oxford university press, 2010.
[3] S.P. Borgatti, D.J. Brass, D.S. Halgin, Social network research: Confusions, criticisms, and controversies, in: Contemporary Perspectives on Organizational Social Networks, Emerald Group Publishing Limited, 2014, pp. 1–29.
[4] R. Guimera, S. Mossa, A. Turtschi, L.N. Amaral, The worldwide air transportation network: Anomalous centrality, community structure, and cities' global roles, Proc. Natl. Acad. Sci. 102 (22) (2005) 7794–7799.
[5] R.I. Dunbar, V. Arnaboldi, M. Conti, A. Passarella, The structure of online social networks mirrors those in the offline world, Soc. Networks 43 (2015) 39–47.
[6] Z.-K. Zhang, C. Liu, X.-X. Zhan, X. Lu, C.-X. Zhang, Y.-C. Zhang, Dynamics of information diffusion and its applications on complex networks, Phys. Rep. 651 (2016) 1–34.
[7] X.-X. Zhan, C. Liu, G. Zhou, Z.-K. Zhang, G.-Q. Sun, J.J. Zhu, Z. Jin, Coupling dynamics of epidemic spreading and information diffusion on complex networks, Appl. Math. Comput. 332 (2018) 437–448.
[8] A. Zeng, Z. Shen, J. Zhou, J. Wu, Y. Fan, Y. Wang, H.E. Stanley, The science of science: From the perspective of complex systems, Phys. Rep. 714–715 (2017) 1–73.
[9] P. Mateos, P.A. Longley, D. O'Sullivan, Ethnicity and population structure in personal naming networks, PLoS One 6 (9) (2011) e22943.
[10] K. Kowalska, P. Longley, M. Musolesi, Ethnic structure in global naming networks, 2015, Accessed by internet at http://www.ucl.ac.uk/~ucfamus/papers/giscience14.pdf.
[11] J. Novotnỳ, J.A. Cheshire, The surname space of the Czech Republic: Examining population structure by network analysis of spatial co-occurrence of surnames, PLoS One 7 (10) (2012) e48568.
[12] G. Ferreira, G. Viswanathan, L. da Silva, H. Herrmann, Surname complex network for Brazil and Portugal, Physica A 499 (2018) 198–207.
[13] P. Mateos, Names, Ethnicity and Populations, Springer Berlin Heidelberg, 2014.
[14] P. Darlu, G. Bloothooft, A. Boattini, L. Brouwer, M. Brouwer, G. Brunet, P. Chareille, J. Cheshire, R. Coates, K. Dräger, et al., The family name as socio-cultural feature and genetic metaphor: From concepts to methods, Hum. Biol. 84 (2) (2012) 169–214.
[15] J. Cheshire, Analysing surnames as geographic data, J. Anthropol. Sci. 92 (2014) 99–117.
[16] R. Du, Y. Yuan, J. Hwang, J. Mountain, L.L. Cavalli-Sforza, Chinese surnames and the genetic differences between north and south China, J. Chin. Linguist. (5) (1992) 1–93.
[17] P. Hanks, Dictionary of American Family Names, Oxford University Press, 2003.
[18] R.N. Mantegna, Hierarchical structure in financial markets, Eur. Phys. J. B 11 (1) (1999) 193–197.
[19] T. Mizuno, H. Takayasu, M. Takayasu, Correlation networks among currencies, Physica A 364 (2006) 336–342.
[20] J.W. Song, B. Ko, W. Chang, Analyzing systemic risk using non-linear marginal expected shortfall and its minimum spanning tree, Physica A 491 (2018) 289–304.

[21] P. Coletti, Comparing minimum spanning trees of the Italian stock market using returns and volumes, Physica A 463 (2016) 246–261.
[22] P. Tewarie, E. van Dellen, A. Hillebrand, C.J. Stam, The minimum spanning tree: An unbiased method for brain network analysis, Neuroimage 104 (2015) 177–188.
[23] E. van Dellen, I.E. Sommer, M.M. Bohlken, P. Tewarie, L. Draaisma, A. Zalesky, M. Di Biase, J.A. Brown, L. Douw, W.M. Otte, et al., Minimum spanning tree analysis of the human connectome, Hum. Brain Mapp. 39 (6) (2018) 2455–2471.
[24] N. Akpan, I. Iwok, A minimum spanning tree approach of solving a transportation problem, Int. J. Math. Stat. Invent. 5 (3) (2017) 09–18.
[25] L.L. Cavalli-Sforza, A.W. Edwards, Phylogenetic analysis: Models and estimation procedures, Evolution 21 (3) (1967) 550–570.
[26] A. Rodriguez-Larralde, C. Scapoli, M. Beretta, C. Nesti, E. Mamolini, I. Barrai, Isonymy and the genetic structure of Switzerland. ii. isolation by distance, Ann. Hum. Biol. 25 (6) (1998) 533–540.
[27] M. Nei, Genetic distance between populations, Am. Nat. 106 (949) (1972) 283–292.
[28] A. Rodriguez-Larralde, A. Gonzales-Martin, C. Scapoli, I. Barrai, The names of Spain: A study of the isonymy structure of Spain, Am. J. Phys. Anthropol. 121 (3) (2003) 280–292.
[29] J. Dipierri, E. Alfaro, C. Scapoli, E. Mamolini, A. Rodriguez-Larralde, I. Barrai, Surnames in Argentina: A population study through isonymy, Am. J. Phys. Anthropol. 128 (1) (2005) 199–209.
[30] Y. Liu, L. Chen, Y. Yuan, J. Chen, A study of surnames in China through isonymy, Am. J. Phys. Anthropol. 148 (3) (2012) 341–350.
[31] S. Xingjian, Z. Chen, H. Wang, D.-Y. Yeung, W.-K. Wong, W.-C. Woo, Convolutional LSTM network: A machine learning approach for precipitation nowcasting, in: Adv. Neural Inf. Process. Syst., 2015, pp. 802–810.
[32] B. Saoud, A. Moussaoui, Community detection in networks based on minimum spanning tree and modularity, Physica A 460 (2016) 230–234.
[33] R.C. Prim, Shortest connection networks and some generalizations, AT&T Tech. J. 36 (6) (1957) 1389–1401.
[34] J.B. Kruskal, On the shortest spanning subtree of a graph and the traveling salesman problem, Proc. Amer. Math. Soc. 7 (1) (1956) 48–50.
[35] S.K. Jo, M.J. Kim, K. Lim, S.Y. Kim, Correlation analysis of the Korean stock market: Revisited to consider the influence of foreign exchange rate, Physica A 491 (2018) 852–868.
[36] J.C. Gower, G.J. Ross, Minimum spanning trees and single linkage cluster analysis, Appl. Stat. (1969) 54–64.
[37] M. Tumminello, F. Lillo, R.N. Mantegna, Correlation, hierarchies, and networks in financial markets, J. Econ. Behav. Organ. 75 (1) (2010) 40–58.
[38] W. Wang, H. Liu, X. Yue, H. Li, J. Chen, L. Ren, D. Tang, S. Hatakeyama, A. Takami, Study on acidity and acidic buffering capacity of particulate matter over Chinese eastern coastal areas in spring, J. Geophys. Res. Atmos. 111 (D18207) (2006) 1–11.
[39] M. Faure, W. Lin, N. Le Breton, Where is the north China–south China block boundary in eastern China?, Geology 29 (2) (2001) 119–122.
[40] Z. Guowei, Y. Zaiping, S. Yong, C. Shunyou, L. Taohong, X. Feng, Z. Chengli, The major suture zone of the Qinling orogenic belt, J. Southeast Asian Earth Sci. 3 (1–4) (1989) 63–76.
[41] Y. Dong, M. Santosh, Tectonic architecture and multiple orogeny of the Qinling orogenic belt, central China, Gondwana Res. 29 (1) (2016) 1–40.
[42] Y. Dong, X. Liu, F. Neubauer, G. Zhang, N. Tao, Y. Zhang, X. Zhang, W. Li, Timing of paleozoic amalgamation between the north China and south China blocks: Evidence from detrital zircon u–pb ages, Tectonophysics 586 (2013) 173–191.
[43] J. Chen, C. Li, G. He, A diagnostic analysis of the impact of complex terrain in the eastern Tibetan Plateau, China, on a severe storm, Arct. Antarct. Alp. Res. 39 (4) (2007) 699–707.
[44] M. Zhang, J. Xia, New challenges and opportunities for flood control in the Huai river: Addressing a changing river-lake relationship, Bull. Acad. Mil. Med. Sci. (1) (2012) 42–49.
[45] W.R. Tobler, A computer movie simulating urban growth in the detroit region, Econ. Geogr. 46 (sup1) (1970) 234–240.
[46] A. Abbasi, L. Hossain, L. Leydesdorff, Betweenness centrality as a driver of preferential attachment in the evolution of research collaboration networks, J. Inform. 6 (3) (2012) 403–412.
[47] G. Zhao, X. Mu, A. Strehmel, P. Tian, Temporal variation of streamflow, sediment load and their relationship in the Yellow River basin, China, PLoS One 9 (3) (2014) e91048.
[48] J. Guo, S. Guo, Y. Li, H. Chen, T. Li, Spatial and temporal variation of extreme precipitation indices in the Yangtze River basin, China, Stoch. Environ. Res. Risk Assess. 27 (2) (2013) 459–475.