

## **REGULAR ARTICLE**

**Open Access** 



## Segregation in religion networks

Jiantao Hu<sup>1</sup>, Qian-Ming Zhang<sup>2\*</sup> and Tao Zhou<sup>1,2\*</sup>

\*Correspondence: qmzhangpa@gmail.com; zhutou@ustc.edu ¹CompleX Lab, University of Electronic Science and Technology of China, Chengdu, China ²Big Data Research Center, University of Electronic Science and Technology of China, Chengdu, China

#### **Abstract**

Religion is considered as a notable origin of interpersonal relations, as well as an effective and efficient tool to organize a huge number of people towards some challenging targets. At the same time, a believer prefers to make friend with other people of the same faith, and thus people of different faiths tend to form relatively isolated communities. The segregation between different religions is a major factor for many social conflicts. However, quantitative understanding of religious segregation is rare. Here we analyze a directed social network extracted from weibo.com (the largest directed social network in China, similar to twitter.com), which is consisted of 6875 believers in Christianity, Buddhism, Islam and Taoism. This religion network is highly segregative. Comparative analysis shows that the extent of segregation for different religions is much higher than that for different races and slightly higher than that for different political parties. Furthermore, we study the few cross-religion links and find 46.7% of them are probably related to charitable issues. Our findings provide quantitative insights into religious segregation and valuable evidence for religious syncretism.

**Keywords:** Social networks; Religion; Mixing pattern; Segregation; Percolation

### 1 Introduction

Religious belief is a sort of mental cement that connects, blends people of diverse colors, ages and sexes. It could facilitate human cooperation [1–6], promote civic engagement [7–10], improve life satisfaction [11–15], intersect with politics [16–20], impact people's mental and/or physical health [21–24], influence social morality [25, 26], coexist with science [27–30] and even boom economic development [31, 32]. On the other side, people usually do not wish to have close relationships or frequent interactions to others of different religious beliefs. And thus people of different faiths tend to form relatively isolated communities [33]. Analogous to the separation of races [34], such religious segregation largely influences (usually negatively) culture evolution, economic development, political pattern, and so on [35, 36]. Furthermore, some different religions are even hostile to each other, which would increase prejudice between religions [37–39], and lead to regional violence, intergroup conflict and moral prejudice against atheists [40–46].

The above studies provided fruitful results about religions, but few researchers paid attention to believers in China. According to the latest data from "2012 China Family Panel Studies (CFPS)", only 10% population in China are religious. Even though it is a small proportion, it is still worthy to study since the total amount of people is over 100 million. Within the believers, 64.66% are Buddhists, 22.03% are Christians, 5.17% are Taoists, and



Hu et al. *EPJ Data Science* (2019) 8:6 Page 2 of 11

4.41% are Islamists. To date, researches on religions in China were mainly based on questionnaire surveys. Some statistical results [47] have been reported, including the number of believers with different faiths, the gender ratio of the believers, the educational level of the believers, etc. Nevertheless, the relationships among different religious groups are still unknown. The main reason is the lack of available data. During recent years, the online social platforms have recorded interactions among huge population, which makes it possible to quantify the segregation in different religions.

We focus on weibo.com, one of the largest online social platform in China. We identify 6875 believers in Christianity, Buddhism, Islam and Taoism, and construct a directed network based on the *follower*  $\rightarrow$  *followee* relationship among these believers. Through analyzing the mixing pattern of the religion network, we find most of the links are created between the individuals having same belief. This phenomenon of homophily [33] is more significant in the Moslems and Taoists. In other word, the religion network is highly segregative. There are only 1.6% of links connecting different religions. Although the few cross-religion links are apparently important to the network connectivity, it is surprised that these links are remarkably more important than links with highest betweennesses [48] or bridgenesses [49]. In particular, we also find that 46.7% of these cross-religion links are probably related to charitable issues. The contribution of this paper is two-fold. Firstly, we provide quantitative insights into religious segregation. Secondly, we claim that charitable issues might play a positive role in facilitating religious syncretism.

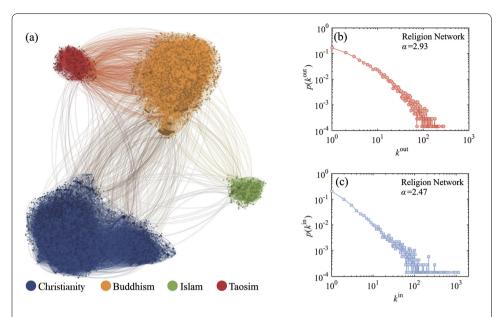
#### 2 Dataset

The dataset was collected from *Sina Weibo* (weibo.com) in April 2016. *Sina Weibo* is a Chinese microblogging website (similar to twitter.com), which has over 500 million registered users up to now. The number of monthly active users reaches about 340 million. Users can follow others, for example, if user A follows user B, B is called a followee, meanwhile A is a follower that receives B's updates. Users are allowed to describe themselves with several tags and a short description.

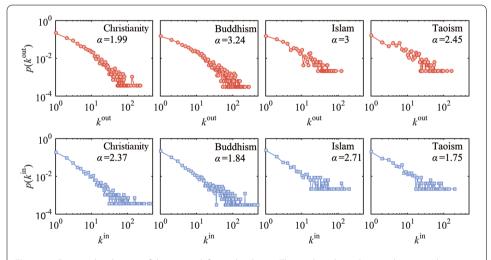
Here we focus on the users who believe in Christianity, Buddhism, Islam or Taoism. To extract the believers from weibo.com, we build up a set of religious keywords (they are Chinese keywords and readers are encouraged to see the full list in the following website <a href="http://www.dcjingsai.com/common/share/73.html">http://www.dcjingsai.com/common/share/73.html</a>), which covers the most frequently-used words related to religion. Then we searched these keywords on weibo.com and got about 170K users. Due to the ambiguity problem in word-cutting of Chinese text, we only keep the users who have at least two keywords (repeatable) in their descriptions and nicknames (similar to the name in one's profile on twitter, which can be customized by user). Then the number of remaining users is about 9K. Lastly, we check these 9K candidates one by one by hand to make sure all the 6875 users under consideration are religious users.

Focusing on these religious users, we extract a subgraph G(V,E) from weibo.com, where V and E denote the sets of nodes and links, respectively. The node set V contains 6875 believers in four major religions in China, including 3153 Christians, 2791 Buddhists, 470 Islamists and 461 Taoists. The link set contains 76,678 directed links, and the average degree is 11.15. Figure 1(a) presents a visual layout of the network, from which one can see clearly that connections inside a religion are dense while connections in-between different religions are much sparser. As shown in Fig. 1(b) and Fig. 1(c), both out-degree and indegree distributions follow power-law-like distributions, as  $p(k) \sim k^{-\alpha}$  where k denotes

Hu et al. *EPJ Data Science* (2019) 8:6 Page 3 of 11



**Figure 1** Structure of the religion network. (a) Structural layout of the network neglecting the directions of links, where blue, orange, green and red nodes denote Christians, Buddhists, Islamists and Taoists, respectively. This network is visualized by Gephi (gephi.org). (b) The out-degree distribution in a log–log plot, with an estimated power-law exponent  $\alpha \approx$  2.93. (c) The in-degree distribution in a log–log plot, with an estimated power-law exponent  $\alpha \approx$  2.47



**Figure 2** Degree distribution of the network for each religion. These plots show the out-degree and in-degree distributions in log-log plots for Christianity, Buddhism, Islam and Taosim. The power-law exponents are estimated by the maximum likelihood method [50]

the degree and  $\alpha$  is the power-law exponent. The power-law exponents, estimated by the maximum likelihood method [50], are  $\alpha \approx 2.93$  and  $\alpha \approx 2.47$  for out-degree and in-degree distributions, respectively. Further, we analyze induced subgraphs [51] of each religion, where an induced subgraph of a given religion contains users in this religion and all links connecting these users. Induced subgraph for each religion exhibits scale-free property [52] (see Fig. 2), indicating the existence of leaders (with large in-degrees) and enthusiasts (with large out-degrees).

Hu et al. *EPJ Data Science* (2019) 8:6 Page 4 of 11

#### 3 Results

## 3.1 The basic structure and mixing pattern

Neglecting the directions of links, we can obtain an undirected version G'(V, E') from G, where two nodes i and j are connected if either there is a link from i to j or there is a link from j to i. In G', there are in total 64,712 links. G' displays clustering feature as indicated by its high clustering coefficient [53] C = 0.37, and community structure with a high modularity [54] Q = 0.57 if we directly treat individuals in one religion as one community. However, neither clustering coefficient nor modularity is enough to characterize the aggregation of believers in the same religion or the segregation of believers in different religions, since the former only considers local organization and the latter is very sensitive to the community sizes [55]. Accordingly, we look into the detailed mixing pattern of the religion network. Denote  $e_{ij}$  the fraction of links from religion i to religion j (i, j = 1, 2, 3, 4),  $a_i = \sum_i e_{ii}$  the fraction of links from religion i, and  $b_i = \sum_i e_{ii}$  the fraction of links pointing to religion *j*, the corresponding mixing matrix is shown in Table 1. Obviously, the religion network is highly assortative, with most links connecting believers in the same religion. In fact, only 1.6% links are connecting believers of different faiths. We further calculate the assortativity coefficient r [56], which lies in [-1,1] with r=1 corresponding to the perfect assortative mixing (see Methods). The assortativity coefficient of the religion network is surprisingly high, as r = 0.973. In comparison, it is even higher than some wellknown social networks with remarkable segregation, such as sexual partnerships mixed by races [57] (r = 0.621) and Twitter web of politicians in democratic party and republican party [58] (r = 0.954). Besides assortativity coefficient, we also introduce some other indices which are usually used to quantify the network-level segregation. They are E-I index (S<sub>EI</sub>) [59], Gupta-Anderson-May (GAM index) [60] and Odds-ratio for within-group ties (Odds-ratio WG ties) [61]. The detailed definitions of these indices are shown in Methods. Table 2 shows the results of these indices for the religion network in comparison with

**Table 1** Mixing matrix of the religion network. The number in ith row and jth column represents  $e_{ij}$ , the fraction of links from religion i to religion j

Religion	Religion	a <sub>i</sub>			
	Christianity	Buddhism	Islam	Taoism	
Christianity	0.5594303	0.0028952	0.000104332	0.0002739	0.5627038
Buddhism	0.0017606	0.2971778	0.000091291	0.0048254	0.3038551
Islam	0.0001956	0.0005999	0.056769869	0.0001304	0.0576958
Taoism	0.0001695	0.0046167	0.000013042	0.070946	0.0757453
bj	0.5615561	0.3052897	0.05697853	0.0761757	

**Table 2** Segregation measures applied to the religion network, the Twitter web of politicians in democratic party and republican party (political network) and the sexual partnerships network mixed by races (ethnic network). The value of Odds-ratio WG ties for the ethnic network cannot be obtained since we do not know the number of non-existent links in the ethnic network, which is necessary to apply the Odds-ratio WG ties

	Religion network	Political network	Ethnic network
-S <sub>EI</sub>	0.969	0.954	0.477
Assortativity Coefficient	0.973	0.954	0.621
GAM index	0.964	0.954	0.520
Odds-ratio WG ties	101.05	37.94	N/A

Hu et al. *EPJ Data Science* (2019) 8:6 Page 5 of 11

**Table 3** Mixing matrix of the null network. The number in ith row and jth column represents  $e_{ij}$ , the fraction of links from religion i to religion i

Religion	Religion	a <sub>i</sub>			
	Christianity	Buddhism	Islam	Taoism	
Christianity	0.3152012	0.1703618	0.032669084	0.0444717	0.5627038
Buddhism	0.1704531	0.0945904	0.016940974	0.0218707	0.3038551
Islam	0.0329169	0.0176452	0.003012598	0.0041211	0.0576958
Taoism	0.042985	0.0226923	0.004355878	0.0057122	0.0757453
$b_j$	0.5615561	0.3052897	0.05697853	0.0761757	

**Table 4** Connecting ratios of the religion network to the null network. The number in ith row and jth column represents the ratio of links from religion i to religion j in the religion network to those in the null network

Religion	Religion				
	Christianity	Buddhism	Islam	Taoism	
Christianity	1.7748355	0.0169946	0.00319361	0.0061584	
Buddhism	0.010329	3.1417345	0.00538876	0.2206321	
Islam	0.0059429	0.0339985	18.8441558	0.0316456	
Taoism	0.0039442	0.2034483	0.00299401	12.420091	

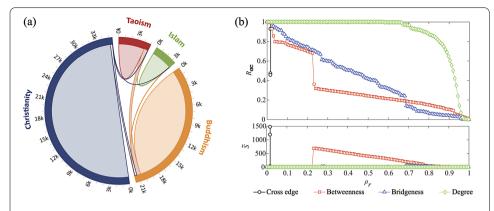
the political network and the ethnic network. In consistent to the results by assortativity coefficient, the religious network exhibits the highest degree of segregation.

We further compare the mixing matrix of the religion network *G* with its randomized counterpart G<sup>null</sup>, i.e., one null model of G, which is obtained by the degree-preserved link-rewiring process [62] (see Methods). This process can ensure each node in  $G^{\text{null}}$  has exactly the same degree as in G. Since G is heterogeneous (with power-law-like degree distribution), it is possible that such heterogeneity may lead to segregation. Since both the original network G and the null model  $G^{\text{null}}$  have the same degree heterogeneity, the direct comparison between their mixing patterns can demonstrate the role of the connecting pattern (the choices of whom to connect to) in producing the segregation structure. Table 3 shows the mixing matrix of the null network. We define the connecting ratio from religion *i* to religion *j* of *G* to  $G^{\text{null}}$  as  $\rho_{ij} = e_{ij}/e_{ii}^{\text{null}}$ , where  $e_{ii}^{\text{null}}$  is the fraction of links from religion i to religion j (i, j = 1, 2, 3, 4) in the null network. Table 4 shows such ratios, from which one can observe two remarkable phenomena: (i) Believers statistically tend to connect with others of the same faith as indicated by  $\forall i$ ,  $\rho_{ii} > 1$ , while Islam and Taoism exhibit the highest level of homophily with  $\rho_{33}$  = 18.84 and  $\rho_{44}$  = 12.42; (ii) The ratios associated with Buddhism, say  $\rho_{2\bullet}$  and  $\rho_{\bullet 2}$ , are all the largest ones in corresponding rows and columns excluded the diagonal elements, indicating that Buddhism plays the key role in cross-religion communications in China.

## 3.2 Cross-religion links analysis

As indicated by the structural statistics, a tiny number of cross-religion links (i.e., links connecting individuals in different religions, see Fig. 3(a) for visualization) play a critical role in maintaining the global connectivity of the religion network. To quantify the significance of cross-religion links, we apply the link percolation dynamics [63], where links are ranked by a certain criterion and then removed one by one in order. If removing some links following a criterion (or some rules) can rapidly break the network into pieces, this criterion (or rules) is regarded as a good indicator for the importance of links.

Hu et al. *EPJ Data Science* (2019) 8:6 Page 6 of 11



**Figure 3** Quantifying the significance of cross-religion links in maintaining network connectivity. (a) Circular plot of links within and between religions, with links' directions being neglected. Inter-religion links are all colored in light colors to emphasize cross-religion links. (b) Link percolation processes by gradually removing cross-religion links in a random order (black circles), links with largest betweennesses (red squares), links with largest bridgenesses (blue triangles), and links with highest degrees (green diamonds), respectively. The upper and lower plots show the changes of giant component size and normalized susceptibility as the increase of the fraction of removed links

For convenience, we consider the undirected version G', wherein there are in total 1124 cross-religion links. The global connectivity is intuitively measured by the ratio of nodes in the giant component (i.e., the largest connected component) to the total number of nodes N, denoted by  $R_{GC}$ . Increasing  $\rho_r$ , the fraction of links being removed, the percolation dynamics may come across a phase transition where the network suddenly breaks into many small fragments at the corresponding critical point, accompanied by a sharp drop of  $R_{GC}$ . To precisely locate the critical point  $\rho_r^c$ , we adopt the normalized susceptibility  $\widetilde{S} = \sum_{s < s_{\max}} \frac{n_s s^2}{N}$  [64], where  $n_s$  denotes the number of components with size s and s runs from the smallest size to the second largest size. If there is a percolation phase transition, an obvious peak in the  $\widetilde{S}(\rho_r)$  curve can be observed that corresponds to the critical point  $\rho_r^c$ , at which the network disintegrates. A set of links whose removal leads to faster decay of  $R_{GC}$  and smaller value of  $\rho_r^c$  is considered to be more significant in maintaining the network connectivity.

We compare the following four methods in identifying significant links for connectivity maintaining: (i) Removing the 1124 cross-religion links in a random order; (ii) Removing links in a descending order of their betweennesses [48]; (iii) Removing links in a descending order of their bridgenesses [49]; (iv) Removing links in a descending order of their degrees [65]. The explicit definitions of betweenness, bridgeness and degree for an arbitrary link are presented in Methods. As shown in Fig. 3(b), as the increasing of  $\rho_r$ ,  $R_{GC}$  decreases much faster when removing cross-religion links first. Remarkable peaks are observed only for the cross-religion links and the largest-betweenness-first method, while the critical point of the former ( $\rho_r^c = 1119/64,712$ ) is one order of magnitude smaller than that of the latter ( $\rho_r^c = 14,956/64,712$ ). In comparison, the performance of the cross-religion links is far better than betweennesses, bridgenesses and degrees.

To uncover the underlying mechanism in the creation of cross-religion links, we classify all nodes into 4 types: (i) nodes not associated with any cross-religion links, (ii) nodes associated with some cross-religion out-links but none of cross-religion in-links, (iii) nodes associated with some cross-religion in-links but none of cross-religion out-links, and (iv) nodes associated with both cross-religion out-links and in-links. Table 5 shows the

Hu et al. *EPJ Data Science* (2019) 8:6 Page 7 of 11

**Table 5** Distribution and average degrees of nodes in different types. In each illustration plot, the black node is the ego under consideration and the white node(s) is (are) its neighbor(s). The first four rows show the distribution of nodes of different types in different religions. The numbers in the brackets denote the number of charitable nodes. The average degree of the religion network is 11.15

	Type 1 ●	Type 2 $\bullet \rightarrow \circ$	Type 3	Type 4 $\circ \rightarrow \bullet \rightarrow \circ$
			$\circ \to \bullet$	
Christianity	2930	153	61 (1)	9 (1)
Buddhism	2475	170	78 (24)	68 (6)
Islam	417	41	5 (0)	7 (0)
Taoism	271	109	42 (0)	39 (1)
Average out-degree	9.1674	25.0085	22.7366	38.7236
Average in-degree	6.7857	9.3002	134.3978	48.2602

distribution of nodes of different types in different religions, as well as the average outdegree and in-degree over nodes of different types. Obviously, nodes without any crossreligion links are statistically of smaller degrees than the entire average, while nodes following or being followed by believers of other faiths are generally of higher out-degrees or in-degrees. In particular, the ones being followed by but having not followed believers of other faiths (i.e., Type 3) are usually very popular, with average in-degree more than 10 times larger than the entire average. We further look into the personal descriptions and posted microblogs of nodes of Type 3 and Type 4. There are in total 309 nodes of Type 3 or Type 4, each of which has attracted at least one believer from another religion. By content analysis, we find that there are in total 33 users whose main interest is charity. About 36.3% of these 33 users labeled themselves as philanthropists, and 45.5% of them were reported by medias that they were involved in charities. The rest of these users often promoted philanthropy through tweets. These charitable users attract 46.7% of all cross-religion links. Detailed information about these 33 users can be found in http://www.dcjingsai.com/common/share/73.html (in Chinese). To our surprise, such charitable nodes (about 10.7% of the 309 nodes) have attracted 46.7% of all cross-religion links, and most charitable nodes (30 of 33) are Buddhists.

#### 4 Discussion

In summary, though everybody has observed some evidence about religion segregation in daily life, this paper provides quantitative analysis based on an extracted religion network from weibo.com. The extent of networked segregation for different religions, measured by the assortativity coefficient, is even higher than that for different races or different political parties. In fact, to our knowledge, the present religion network exhibits the highest segregation among all previously reported social networks. Among the four religions under consideration, Buddhism plays the most significant role in promoting the cross-religion communications. We still cannot be sure this is a specific phenomenon in China as Buddhism itself is one of a few mainstays of the Chinese culture or a universal phenomenon over the world since the Buddhist doctrines are very inclusive and tolerant. A solid answer to this question asks for more data from twitter.com as well as other representative social networks at national level. We have also found that the small-scale religions in China, namely Islam and Taosim, show much higher level of cohesion (see Table 4), which probably reflects a general observation that the subculture group of smaller size usually shows a higher level of homophily and denser interactions [66].

A tiny fraction of cross-religion links maintain the global connectivity, whose removal will lead to much faster breakdown of the network in comparison with those links with

Hu et al. *EPJ Data Science* (2019) 8:6 Page 8 of 11

highest betweennesses or bridgenesses. Therefore we want to understand the underlying reasons of the generation of these cross-religion links. To our surprise, about half links point to charitable nodes. This strong evidence suggests that charity may be a common interest that can stride across the ideological barriers between religions. Accordingly, encouraging and holding charity-related activities, and at the same time inviting participants from different religions, may be an effective method to facilitate cross-religion communications.

In this paper, we demonstrate the effectiveness and validity of the data-driven paradigm in the studies of religious issues, and we believe it will turn to be the mainstream methodology in the near future [67, 68]. However, the current data is of very small size in comparison with the whole population of Chinese religious believers. Therefore the reported findings just provide a tiny and early step towards the comprehensive landscape of communicating patterns between believers of different faiths. Three open issues are left for further studies. First of all, we would like to test the universality of the present observations based on data from other countries. Secondly, we want to see the evolution of the connecting patterns of religion networks by tracing the temporal data [69]. Lastly, it would be interesting to see the role of religious believers in the whole social network, instead of the network containing only believers. For example, whether the echochamber exists between followers of different religions in which they only share similar posts and only engage in conversation with similar ones. This is of particular importance for countries like China where theists are the minority and their social inclusion needs to be promoted.

## 5 Methods

## 5.1 Segregation measures

Segregation measures are used to quantify whether and to which extent links tend to connect nodes in the same type.

- (1) Assortativity coefficient [56] is defined as  $r = \frac{\sum_i e_{ii} \sum_i a_i b_i}{1 \sum_i a_i b_i}$ , where  $e_{ii}$ ,  $a_i$  and  $b_i$  are introduced in the main text. In the case of the perfect assortative mixing, all links connecting nodes in the same type, leading to  $\sum_i e_{ii} = 1$  and r = 1.
- (2) E–I index [59] indicates differences within and between groups links, say  $S_{\text{EI}} = \frac{\sum_g \sum_{h \neq g} m_g h \sum_g m_g g}{\sum_g \sum_h m_g h}$ , where  $m_g h$  is the number of links between groups g and h, and  $m_g g$  is the number of links inside group g.
- (3) Gupta–Anderson–May [60] is defined as  $S_{\text{GAM}} = \frac{\sum_g f_{gg} 1}{K 1}$ , where  $f_{gg} = \frac{m_{gg}}{\sum_h m_{gh}}$  is the proportion of links in group g to links of group g pointing to other groups and K is the total number of groups in the network.
- (4) Odds-ratio for within-group ties (Odds-ratio WG ties) [61] is calculated through  $S_{\text{ORWG}} = \frac{\sum_g m_{gg} \sum_g \sum_{h \neq g} \overline{m}_{gh}}{\sum_g \sum_{h \neq g} m_{gh}}$ , where  $\overline{m}_{gh}$  denotes the number of disconnected links between groups g and h.

#### 5.2 Degree-preserved link-rewiring process

This process randomly reshuffles links while keeps the out-degree and in-degree of each node unchanged [62]. At each time step, we randomly select two links  $A \to B$  and  $C \to D$ . If the link  $A \to D$  or  $C \to B$  exists, we go back to reselect two links, otherwise these two links  $A \to B$  and  $C \to D$  are replaced by  $A \to D$  and  $C \to B$ . We repeat such operation for sufficiently long time (10<sup>6</sup> steps in this paper) to obtain the randomized counterpart (called null network) of the original network.

Hu et al. *EPJ Data Science* (2019) 8:6 Page 9 of 11

#### 5.3 Benchmark link centralities

Betweenness centrality of a link l is the fraction of shortest paths between pairs of nodes passing through l [48], say  $BC_l = \sum_{s,t \in V, s \neq t} \frac{\sigma(s,l,t)}{\sigma(s,t)}$ , where  $\sigma(s,t)$  is the number of shortest paths between nodes s and t, and  $\sigma(s,l,t)$  is the number of those paths passing through link l. The bridgeness of a link l is defined as  $B_l = \sqrt{S_x S_y}/S_l$  [49], where x and y are the two endpoints of link l.  $S_x$  and  $S_l$  are the sizes of the maximum cliques (i.e., complete subgraph) that contain node x and link l, respectively. The degree of link l is defined as  $D_l = k_x k_y$  [65], where  $k_x$  and  $k_y$  are the degrees of the two endpoints of l.

#### Acknowledgements

The authors would like to acknowledge Jun Wang for providing the raw data

#### **Funding**

This work was partially supported by the National Natural Science Foundation of China (61433014, 61673085, 61673086, 61703074) and the Science Promotion Programme of UESTC (no. Y03111023901014006).

#### Abbreviations

Not applicable.

#### Availability of data and materials

The datasets analyzed during the current study are available at http://www.dcjingsai.com/common/share/73.html.

#### **Competing interests**

The authors declare that they have no competing interests.

#### Authors' contributions

JH, QMZ and TZ designed research. JH and QMZ performed research. TZ wrote the paper. All authors read and approved the final manuscript.

#### **Endnote**

In Chinese text, we do NOT naturally have words, but characters. One, two or even more characters will form a word. A Chinese sentence looks like ABCDEFCGHHIAD, which contains a number of characters without any blanks in-between them. So in the word level, this sentence can be decomposed as "AB CD EFCG HH I AD" or "AB CDEF CGH HI AD". There are usually more than one way to decompose a sentence into words. Some decompositions will lead to meaningless sentence, but it is very possible that a sentence has multiple ways to be decomposed with reasonable meanings. Therefore, the word-cutting usually depends on semantics of the whole text and is a big challenge in natural language processing in China [70, 71].

#### **Publisher's Note**

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Received: 11 July 2018 Accepted: 27 January 2019 Published online: 01 March 2019

#### References

- 1. Sosis R, Alcorta C (2003) Signaling, solidarity, and the sacred: the evolution of religious behavior. Evol Anthropol 12:264–274
- 2. Atkinson QD, Bourrat P (2011) Beliefs about god, the afterlife and morality support the role of supernatural policing in human cooperation. Evol Hum Behav 32:41–49
- 3. Xygalatas D, Mitkidis P, Fischer R, Reddish P, Skewes J, Geertz AW, Roepstorff A, Bulbulia J (2013) Extreme rituals promote prosociality. Psychol Sci 24:1602–1605
- 4. Baumard N, Boyer P (2013) Explaining moral religions. Trends Cogn Sci 17:272–280
- Botero CA, Gardner B, Kirby KR, Bulbulia J, Gavin MC, Gray RD (2014) The ecology of religious beliefs. Proc Natl Acad Sci USA 111:16784–16789
- Purzycki BG, Apicella C, Atkinson QD, Cohen E, Mcnamara RA, Willard AK, Xygalatas D, Norenzayan A, Henrich J (2016) Moralistic gods, supernatural punishment and the expansion of human sociality. Nature 530:327–330
- 7. Wilson J, Musick M (1997) Who cares? Toward an integrated theory of volunteer work. Am Sociol Rev 62:694–713
- 8. Graham J, Haidt J (2010) Beyond beliefs: religions bind individuals into moral communities. Personal Soc Psychol Rev 14:140–150
- Lewis VA, MacGregor CA, Putnam RD (2013) Religion, networks, and neighborliness: the impact of religious social networks on civic engagement. Soc Sci Res 42:331–346
- 10. Power EA (2017) Social support networks and religiosity in rural South India. Nat Hum Behav 1:0057
- 11. Lim C, Putnam RD (2010) Religion, social networks, and life satisfaction. Am Sociol Rev 75:914–933
- 12. Okulicz-Kozaryn A (2010) Religiosity and life satisfaction across nations. Ment Health Relig Cult 13:155–169
- 13. Ritter RS, Preston JL, Hermandez I (2014) Happy tweets: christians are happier, more socially connected, and less analytical than atheists on Twitter. Soc Psychol Pers Sci 5:243–249

Hu et al. *EPJ Data Science* (2019) 8:6 Page 10 of 11

 Francis LJ, Ziebertz HG, Lewis CA (2003) The relationship between religion and happiness among German students. Pastor Psychol 51:273–281

- 15. Lewis CA, Cruise SM (2006) Religion and happiness: consensus, contradictions, comments and concerns. Ment Health Relig Cult 9(3):213–225
- 16. Wald KD, Calhoun-Brown A (2014) Religion and politics in the United States. Rowman & Littlefield, Totowa
- 17. Smith DE (2015) Religion and politics in Burma, vol 2338. Princeton University Press, Princeton
- 18. Mitchell C (2017) Religion, identity and politics in northern Ireland: boundaries of belonging and belief. Routledge, Abingdon
- 19. Fox J (2018) An introduction to religion and politics: theory and practice. Routledge, Abingdon
- 20. Hertzke AD, Olson LR, Den Dulk KR, Fowler RB (2018) Religion and politics in America: faith, culture, and strategic choices. Routledge, Abingdon
- 21. Koenig H, Koenig HG, King D, Carson VB (2012) Handbook of religion and health. Oxford University Press, Oxford
- 22. Levin J, Chatters LM (2008) Religion, aging, and health: historical perspectives, current trends, and future directions.

  J Relig Spirituality Aging 20(1–2):153–172
- 23. Seybold KS, Hill PC (2001) The role of religion and spirituality in mental and physical health. Curr Dir Psychol Sci 10:21–24
- 24. Shiah YJ, Chang F, Chiang SK, Lin IM, Tam WCC (2015) Religion and health: anxiety, religiosity, meaning of life and mental health. J Relig Health 54:35–45
- 25. McKay R, Whitehouse H (2015) Religion and morality. Psychol Bull 141:447–473
- 26. Mannion G (2017) Schopenhauer, religion and morality: the humble path to ethics. Routledge, Abingdon
- 27. Evans MS (2016) Seeking good debate: religion, science, and conflict in American public life. University of California Press, Berkeley
- 28. Ferngren GB (2017) Science and religion: a historical introduction. Johns Hopkins University Press, Baltimore
- 29. Drees WB (1998) Religion, science and naturalism (No. 171). Cambridge University Press, Cambridge
- 30. Plantinga A (2011) Where the conflict really lies: science, religion, and naturalism. Oxford University Press, Oxford
- 31. Campante F, Yanagizawa-Drott D (2015) Does religion affect economic growth and happiness? Evidence from Ramadan. Q J Econ 130:615–658
- 32. Norenzayan A, Shariff AF, Gervais WM, Willard AK, Mcnamara RA, Slingerland E, Henrich J (2016) The cultural evolution of prosocial religions. Behav Brain Sci 39:1
- 33. McPherson M, Smith-Lovin L, Cook JM (2001) Birds of a feather: homophily in social networks. Annu Rev Sociol 27:415–444
- 34. Lewis K (2013) The limits of racial prejudice. Proc Natl Acad Sci USA 110:18814-18819
- 35. Ruse M (2006) A natural history of religion. Nature 439:535
- 36. Bokányi E, Kondor D, Dobos L, Sebők T, Stéger J, Csabai I, Vattay G (2016) Race, religion and the city: Twitter word frequency patterns reveal dominant demographic dimensions in the United States. Palgrave Commun 2:16010
- 37. Atran S. Ginges J (2012) Religious and sacred imperatives in human conflict. Science 336:855–857
- Adida CL, Laitin DD, Valfort MA (2010) Identifying barriers to muslim integration in France. Proc Natl Acad Sci USA 107:22384–22390
- 39. Lindley J (2002) Race or religion? The impact of religion on the employment and earnings of Britain's ethnic communities. J Ethn Migr Stud 28:427–442
- 40. Appleby SR (1999) The ambivalence of the sacred: religion, violence, and reconciliation. Rowman & Littlefield, Totowa
- 41. Neuberg SL, Warner CM, Mistler SA, Berlin A, Hill ED, Johnson JD, Filip-Crawford G, Millsap RE, Thomas G, Winkelman M (2014) Religion and intergroup conflict findings from the global group relations project. Psychol Sci 25:198–206
- 42. Edgell P, Gerteis J, Hartmann D (2006) Atheists as "other": moral boundaries and cultural membership in American society. Am Sociol Rev 71:211–234
- 43. Gervais WM, Shariff AF, Norenzayan A (2011) Do you believe in atheists? Distrust is central to anti-atheist prejudice. J Pers Soc Psychol 101:1189–1206
- Gervais WM (2014) Everything is permitted? People intuitively judge immorality as representative of atheists. PLoS ONE 9:92302
- 45. Gervais WM, Xygalatas D, Mckay RT, Elk MV, Buchtel EE, Aveyard M, Schiavone SR, Dar-Nimrod I, Svedholm-Häkkinen AM, Riekki T (2017) Global evidence of extreme intuitive moral prejudice against atheists. Nat Hum Behav 1:0151
- 46. Kiernan JP (1974) Where zionists draw the line: a study of religious exclusiveness in an African township. Afr Stud
- 47. Lu Y (2014) Report on contemporary Chinese religious based on data of CFPS (2012). World Religious Cultures:11–25
- Girvan M, Newman MEJ (2002) Community structure in social and biological networks. Proc Natl Acad Sci USA 99:7821–7826
- 49. Cheng XQ, Ren FX, Shen HW, Zhang ZK, Zhou T (2010) Bridgeness: a local index on edge significance in maintaining global connectivity. J Stat Mech 2010:10011
- 50. Clauset A, Shalizi CR, Newman MEJ (2009) Power-law distributions in empirical data. SIAM Rev 51:661–703
- 51. Diestel R (2018) Graph theory. Springer, New York
- 52. Barabási AL, Albert R (1999) Emergence of scaling in random networks. Science 286:509–512
- 53. Watts DJ, Strogatz SH (1998) Collective dynamics of 'small-world' networks. Nature 393:440–442
- 54. Newman MEJ, Girvan M (2004) Finding and evaluating community structure in networks. Phys Rev E 69:026113
- 55. Fortunato S, Barthélemy M (2007) Resolution limit in community detection. Proc Natl Acad Sci USA 104:36–41
- 56. Newman MEJ (2003) Mixing patterns in networks. Phys Rev E 67:026126
- 57. Catania JA, Coates TJ, Kegeles S, Fullilove MT, Peterson J, Marin B, Siegel D, Hulley S (1992) Condom use in multi-ethnic neighborhoods of San Francisco: the population-based amen (aids in multi-ethnic neighborhoods) study. Am J Publ Health 82:284–287
- 58. Conover M, Ratkiewicz J, Francisco M, Gonçalves B, Menczer F, Flammini A (2011) Political polarization on Twitter. In: Proceedings of the fifth international AAAI conference on weblogs and social media, pp 89–96
- Krackhardt D, Stern RN (1988) Informal networks and organizational crises: an experimental simulation. Soc Psychol Q 51:123–140

Hu et al. *EPJ Data Science* (2019) 8:6 Page 11 of 11

- Gupta S, Anderson RM, May RM (1989) Networks of sexual contacts: implications for the pattern of spread of HIV. AIDS 3:807–817
- 61. Moody J (2001) Race, school integration, and friendship segregation in America. Am J Sociol 107:679–716
- 62. Maslov S, Sneppen K (2002) Specificity and stability in topology of protein networks. Science 296:910–913
- 63. Onnela J-P, Saramäki J, Hyvönen J, Szabó G, Lazer D, Kaski K, Kertész J, Barabási AL (2007) Structure and tie strengths in mobile communication networks. Proc Natl Acad Sci USA 104:7332–7336
- 64. Aharony A, Stauffer D (2003) Introduction to percolation theory. Taylor & Francis, London
- 65. Holme P, Kim BJ, Yoon CN, Han SK (2002) Attack vulnerability of complex networks. Phys Rev E 65:056109
- 66. Gelder K (2007) Subcultures: cultural histories and social practice. Routledge, London
- 67. Campbell HA (2013) Religion and the Internet: a microcosm for studying Internet trends and implications. New Media Soc 15:680–694
- 68. Chen L, Weber I, Okulicz-Kozaryn A (2014) U.S. religious landscape on Twitter. In: Proceedings of the international conference on social informatics. pp 544–560
- 69. Holme P, Saramäki J (2012) Temporal networks. Phys Rep 519:97–125
- 70. Gao J, Li M, Wu A, Huang C-N (2005) Chinese word segmentation: a pragmatic approach. Comput Linguist 31:531–574
- 71. Xu J, Zens R, Ney H (2004) Do we need Chinese word segmentation for statistical machine translation. In: Proceedings of the third SIGHAN workshop on Chinese language processing, pp 122–128

# Submit your manuscript to a SpringerOpen journal and benefit from:

- ► Convenient online submission
- ► Rigorous peer review
- ▶ Open access: articles freely available online
- ► High visibility within the field
- ► Retaining the copyright to your article

Submit your next manuscript at ▶ springeropen.com