SCIENTIFIC REPORTS

OPEN

Stability of similarity measurements for bipartite networks

Jian-Guo Liu[1,2], Lei Hou[2,3], Xue Pan[2,3], Qiang Guo[2] & Tao Zhou[4]

Similarity is a fundamental measure in network analyses and machine learning algorithms, with wide applications ranging from personalized recommendation to socio-economic dynamics. We argue that an effective similarity measurement should guarantee the stability even under some information loss. With six bipartite networks, we investigate the stabilities of fifteen similarity measurements by comparing the similarity matrixes of two data samples which are randomly divided from original data sets. Results show that, the fifteen measurements can be well classified into three clusters according to their stabilities, and measurements in the same cluster have similar mathematical definitions. In addition, we develop a top-n-stability method for personalized recommendation, and find that the unstable similarities would recommend false information to users, and the performance of recommendation would be largely improved by using stable similarity measurements. This work provides a novel dimension to analyze and evaluate similarity measurements, which can further find applications in link prediction, personalized recommendation, clustering algorithms, community detection and so on.

Connections are everywhere and can be observed between everything in our world[1,2], such as the connections between the online users and their rated or selected objects[3,4] and the connections between neurons in the neural networks[5,6]. To characterize systems consisting of connections in which objects are represented as nodes and connections are represented as links, complex networks have been widely used in the last decade to study the relations between different objects and the structure of those kinds of systems[7–9].

However, in most real complex networks, every pair of nodes has some specific relations even there is no link between them. For example, citation networks consist of papers. If one paper cited another, there would be a link between them[10]. Besides citing relations, other potential relations may exist when the papers have same authors or cited same papers and so on[11]. Generally, *similarity* describes the connections between different objects' properties, which is the most used method to evaluate such relations. And the similarity has become an important measurement with great significance for both theoretical researches (such as biological and physical science) and practical applications (such as the e-commerce and social service). For example, in biological analysis, evaluating the similarity of genes' expression profile, one may identify similar regulations and the control processes of genes[12,13]. Co-expression networks may also be established according to the similarities between genes[14,15]. While protein-protein or metabolic interactions can only be verified by costly experiments and most of the interactions are still unknown, similarity-based link prediction method[16,17] could largely help identify the most likely pair of interacting proteins[18–20]. In addition, the similarity measurements have also found its applications in object clustering[21], community detecting[22,23], machine learning[24,25] and socio-economic dynamics[26]. As to more practical applications, in term of the similarities, recommendation systems[27,28] evaluate correlations between objects such as movies, commodities, books and so on, and accordingly make recommendations to users.

So far, dozens of similarity measurements have been developed. However, with different data, similarity measurements generally have very different performances. Even with different parts of a same data, the results may be also different. Particularly in the bipartite networks, the object similarity is determined by their natural properties, and thus, similarity should be steadfast for a definite pair of objects. On the other hand, the networks we

[1]Data Science and Cloud Service Research Centre, Shanghai University of Finance and Economics, Shanghai 200433, PR China. [2]Research Center of Complex Systems Science, University of Shanghai for Science and Technology, Shanghai 200093, PR China. [3]Informatics Research Center, Henley Business School, University of Reading, Whiteknights, RG6 6UD, United Kingdom. [4]Web Sciences Center, University of Electronic Science and Technology of China, Chengdu 610054, PR China. Correspondence and requests for materials should be addressed to J.G.L. (email: liujg004@ustc.edu.cn).

Dataset	Subject matter	NUM. of users	NUM. of objects	NUM. of links	Sparsity
MovieLens	Movie	5547	5850	698054	2.15×10^{-2}
Netflix	Movie	8608	5081	419247	9.59×10^{-3}
Amazon	Commodity	645056	99622	2036091	3.17×10^{-5}
Last.FM	Artist	1892	17632	92834	2.78×10^{-3}
Epinions	Reviews	28090	30073	422085	5.00×10^{-4}
Del.cioi.us	Bookmark	1861	1860	15328	4.43×10^{-3}

Table 1. Properties of the utilized data sets. The sparsity is the deviation between existed links and possible links, i.e. $T/(M \cdot N)$, where T, M, N are the number of links, objects and users respectively. Subject matters of those data sets are definite objects whose properties are unchangeable except Last.FM. The subjects of Last.FM are artists. However, the artists' music have definite contains and properties. Thus, the artists in Last.FM could also be regarded as objects.

investigated are mapped incompletely, which is always evolving, or contain false positives and negatives[29]. In the user-object bipartite networks, since the natural properties of the objects are unchangeable, a good similarity measurement should always return the same values for each pair of objects. To explore the stability problem of object similarity for bipartite networks, fifteen similarity measurements will be analyzed and studied in this paper. Firstly, we will report the influence of data amount on the stabilities of fifteen similarity measurements. Secondly, the comparison and classification of the fifteen similarity measurements will be analyzed. Finally, we will explore the effect of the object similarity stability on the recommendation.

Similarity Measurements

In many online systems, objects usually could get ratings from different users. To this kind of context, one can use the *Cosine Index* (COS) or *Pearson Coefficient* (PC) to measure the object similarity. When the ratings are unavailable, similarity can also be defined from the structure of the historical data, that is, two objects are similar if they are connected with many same users. The simplest such method is *Common Neighbour* (CN), where the similarity between two objects are directly given by the number of same neighbours who have connections with them. Considering the degree information of two objects, variations of the CN index have been proposed, including the *Salton Index* (SAL)[30], *Jaccard Index* (JAC)[31], *Sørensen Index* (SOR)[32], *Hub Promoted Index* (HPI)[33], *Hub Depressed Index* (HDI) and *Leicht-Holme-Newman Index* (LHN)[34] indices. Instead of the number of the same neighbours in the CN index, the *Adamic-Adar Index* (AA)[35] and *Resource Allocation Index* (RA)[36] indices were presented, regarding the object similarity as the summation of their common neighbours' degrees. According to the preferential attachment process[8], the *Preferential Attachment Index* (PA) was also presented. Furthermore, using the concepts from physics, the *Mass Diffusion* (MD)[37], *Heat Conduction* (HC)[38,39] and *Improved Heat Conduction* (IHC)[40] methods were also investigated. The mathematical definitions of those similarity measurements can be found in the Method Section. Generally speaking, the value of similarity is relatively high (low) if the objects are very similar (different). With these fifteen similarity measurements, we investigate the similarity stability for the user-object bipartite networks.

Data

Six different data sets are applied in this paper to study the stability of similarity measurements, differing both in the subject matter and data sparsity, as shown in Table 1. These data sets are usually modelled as the user-object bipartite networks and widely used to investigate the performance of the recommendation algorithms[41–43]. The *MovieLens* and *Netflix* data sets are movie Web Sites in which users could watch and rate movies. The *Amazon* data set is an e-commerce system in which users buy commodities. The *Last.FM* data set is a music Web site allowing users to collect different artists' music. The *Epinions* data set allows users writing reviews and on the other hand reading others' reviews. The *Del.icio.us* data set is a bookmark Web Site in which users collect and share bookmarks they interested in.

Similarity Stability

Although lots of object similarity measurements have been presented, we could not know the exact object similarity. Thus, to examine the stability of those measurements, we divide the data set into two samples to compare the similarity matrixes calculated from those two samples for each measurement. The data-dividing method can be described as follow: Every record will get a random number p from a uniform distribution ranging from 0 to 1, and this record belongs to the first sample if $p \leq \eta$ and belongs to the second sample if $p > 1 - \eta$, where η can be regarded as a data amount parameter and $0 < \eta \leq 0.5$. With this method, those two samples would have no overlaps, which means, they are totally different parts of the data set. For a specific pair of objects α and β, we use $s^1_{\alpha\beta}$ to denote their similarity in the first sample and $s^2_{\alpha\beta}$ to denote that in the second sample. Thus, if a similarity measurement can give stable evaluation of the object similarity, there would be $s^1_{\alpha\beta} = s^2_{\alpha\beta}$. Figure 1 reports the distributions of similarities of two samples for each of the fifteen similarity measurements in MovieLens data set. The dots would distribute near the diagonal if the measurement can give stable evaluation of object similarity. The PA index presents the most concentrated distribution. The reason lies in the fact that the PA index only considers the neighbour node popularity. Popular objects of a data sample are in general also popular in another data sample, and thus the object similarity is stable. Other measurements' results are not so concentrated especially for pairs of objects with low-similarity pairs of objects. Results in Fig. 1 indicate that, when the data is changed, a same pair
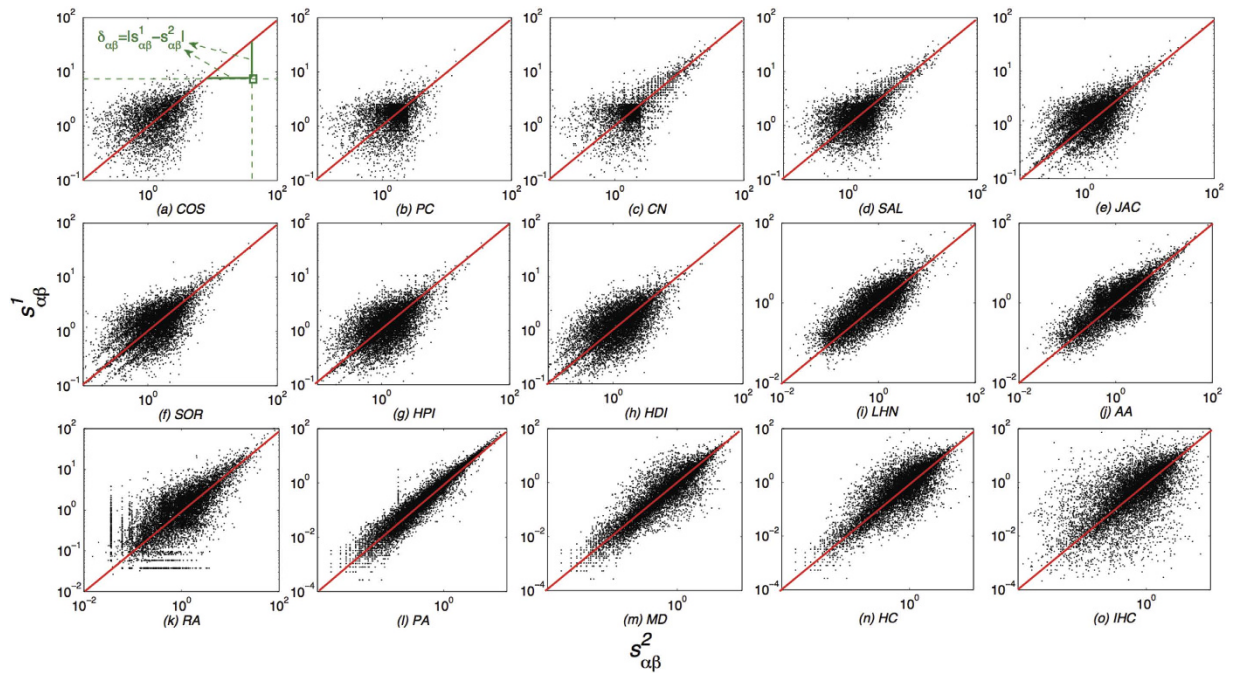
Figure 1. Typical examples of the comparison between object similarities $s^1_{\alpha\beta}$ and $s^2_{\alpha\beta}$ of the MovieLens data set for fifteen similarity measurements. When dividing the data set, we set $\eta = 0.5$, i.e. 50% ratings are divided into the first sample and the others are divided into the second one. For each calculation, we randomly select 10^4 pairs of objects' similarities of two samples to show in the figure. Thus, there are 10^4 dots in each subplot, each represents a pair of objects. The dots would locate on the diagonal if the similarities in two samples $s^1_{\alpha\beta} = s^2_{\alpha\beta}$. Consequently, the more stable the similarity measurement is, the more concentrated the dots would distribute around the diagonal.

of objects may be evaluated as different similarity levels and thus, the stability problem indeed exists in most similarity measurements.

The similarity values calculated by different measurements distribute in different ranges, and thus we make a simple normalization to compare the measurement stability. Suppose that $\bar{s}$ is the average value of similarities that $\bar{s} = \sum_{\alpha\beta} s_{\alpha\beta} / (M (M - 1))$ where M is the number of objects which have at least one record in the corresponding sample, the normalization is $s_{\alpha\beta} = s_{\alpha\beta} / \bar{s}$. Specifically, the similarities of the PC index distribute in the range $[-1,1]$, which may probably leads $\bar{s}$ to be 0. Hence, we make the normalization as $s_{\alpha\beta} = (s_{\alpha\beta} + 1) / (\bar{s} + 1)$ for the PC index. Henceforth, the similarities are all been normalized before used. To qualify the stability of object similarity, we define three metrics:

(1) The *average bias* μ is used to describe the average level of similarity difference between two similarity matrixes from two samples, and it reads

$$\mu = \frac{\sum_{\alpha\beta} \delta_{\alpha\beta}}{M (M - 1)}, \tag{1}$$

where $\delta_{\alpha\beta}$ is the bias of similarities between objects α and β from two samples as shown in Fig. 1(a), i.e. $\delta_{\alpha\beta} = |s^1_{\alpha\beta} - s^2_{\alpha\beta}|$. High value of average bias means, on average, the same pair of objects is evaluated as different similarities when the data is changed. Therefore, the more stable the similarity measurement is, the lower value μ would be.

(2) The *standard deviation of bias* σ reads

$$\sigma = \sqrt{\frac{\sum_{\alpha\beta} (\delta_{\alpha\beta} - \mu)^2}{M (M - 1)}}. \tag{2}$$

The deviation σ can measure the difference of susceptibility between different pairs of objects against the data change. High values of the deviation σ mean that similarities between some pairs of objects may be quite unstable. On the other hand, low values of σ indicates that, each pair of objects has similar unstable level and the bias μ may be caused by the coincident entirety changes of each pair of object similarities.

(3) The *Pearson coefficient* ρ reads
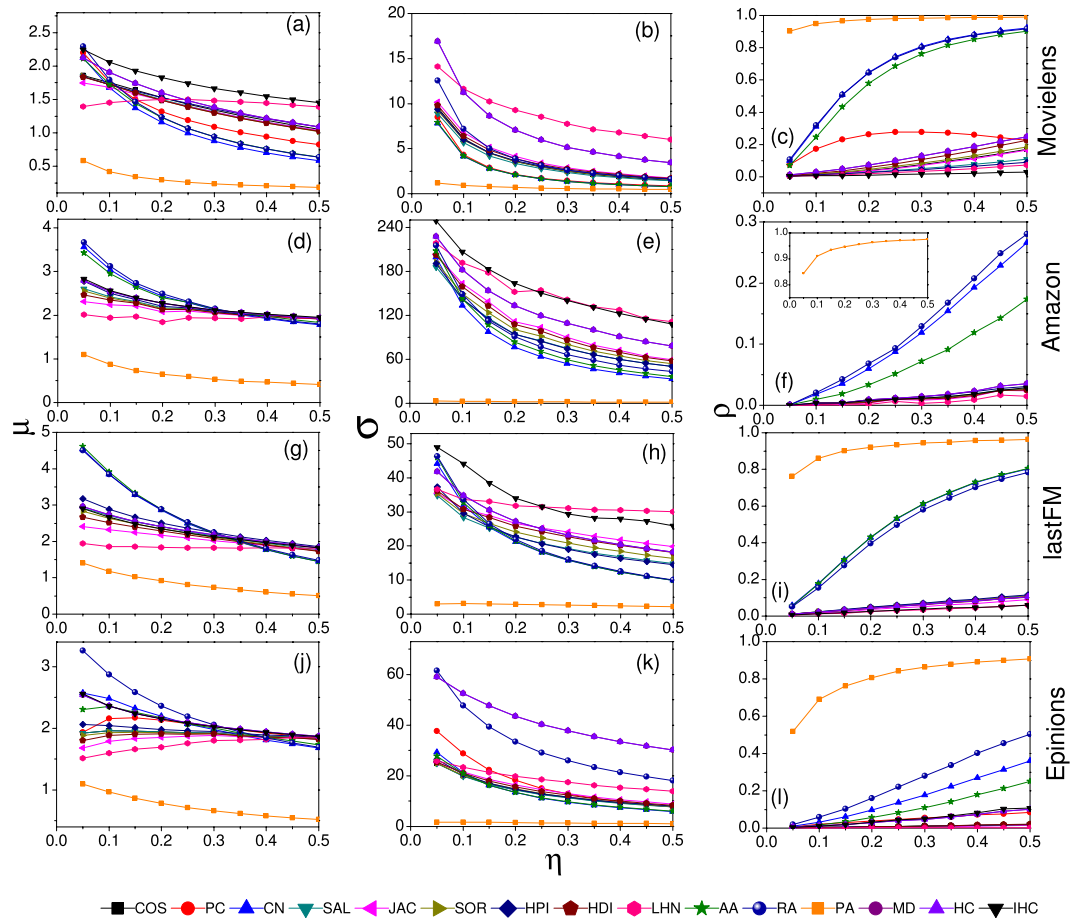
Figure 2. Average bias μ, standard deviation of bias σ and the Pearson coefficient ρ against the data amount parameter η for the MovieLens, Amazon, Last.FM and Epinions data sets respectively. Each data point is averaged over 20 independent experiments, i.e., for each level of data amount parameter η, we randomly divide the data for 20 times and calculate μ, σ and ρ of each time. Note that, there is only selecting information without ratings in the Last.FM data set. Thus, the COS and PC indices could not be performed in the Last.FM data set. As the data becomes more and more abundant, the stability of object similarity would relatively be better. However, many measurements still could not give stable evaluation of object similarity.

$$\rho = \frac{\sum_{\alpha\beta}(s^1_{\alpha\beta} - \bar{s}^1)(s^2_{\alpha\beta} - \bar{s}^2)}{\nu_1 \nu_2 M(M-1)}, \tag{3}$$

where $\bar{s}^1$ and $\bar{s}^2$ are the average value of similarity over every pair of objects for two samples respectively, and ν_1 and ν_2 are the standard variance of similarities for two samples respectively. In general, the value of Pearson coefficient ρ ranging from -1 to 1 measures the coherence of two similarity matrixes calculated by two samples. The upper limit of Pearson coefficient $\rho = 1$ means two similarity matrixes are totally coherent and the corresponding similarity measurement is totally stable.

For each similarity measurement, we calculate the similarity for two data samples with different data amount parameter η. The results of average bias μ, standard deviation of bias σ and the Pearson coefficient ρ of the MovieLens, Amazon, Last.FM and Epinions data sets are reported in Fig. 2 (Results of Netflix and Del.icio.us datasets can be found in the Supplementary Information). One can easily find that, the PA index is the most stable measurement regardless of the data amount η. Even with small size of data, the PA index could give stable evaluation of the object similarity. As the data amount increases, the average biases μ and the standard deviations of bias σ generally decrease. It can be observed that for both the average bias μ and deviation σ, the CN, AA and RA indices have similar decay patterns. When the data amount is small ($\eta < 0.1$) the average bias μ of the CN, AA, RA indices are almost the highest, and with the increase of η, the average bias μ rapidly decreases which means they are sensitive to the data amount. Another dynamic cluster consisting of the COS, SAL, JAC, SOR, HPI, HDI indices seem to be insensitive. Although the average bias μ and deviation σ also decrease with the increment of data, the decays are much slower than that of the former cluster (the CN, AA, RA indices). A special measurement refers to the LHN index, which has no apparent dynamic against the data amount η. Same with the results of average bias μ and deviation σ, the Pearson coefficient ρ of the PA index is the highest and larger than 0.9 even with the smallest

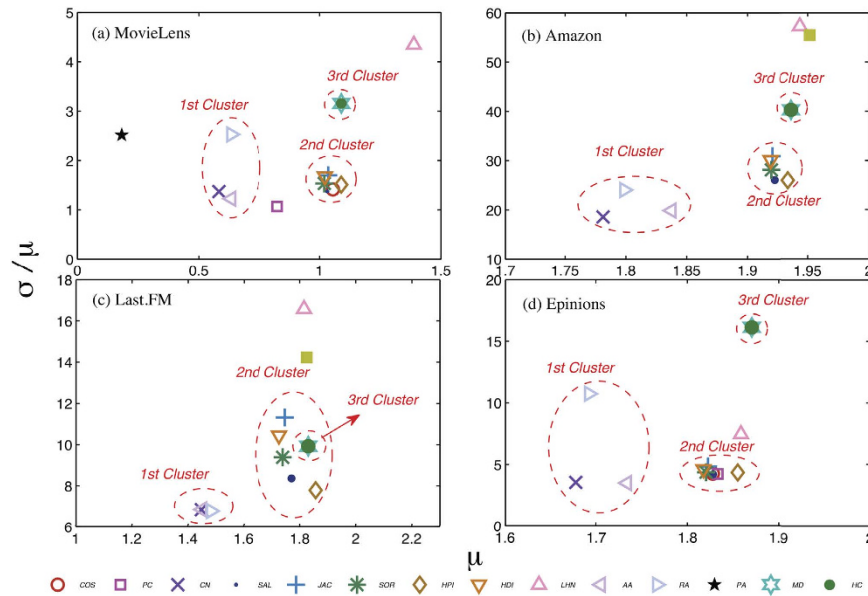**Figure 3. The μ–σ location map with data amount parameter $\eta = 0.5$.** On the location map, a measurement locating on the left side means the similarities of objects have little change at average, and the bottom means the similarities of each pair of objects have similar changes. Overall, a stable measurement generally would locate on the left bottom of the $\mu - \sigma$ location map.

data amount ($\eta = 0.05$). As to the CN, AA, RA indices, the Pearson coefficients ρ are also sensitive, which is similar to the average bias μ and deviation σ. As the data amount η increases, the Pearson coefficient ρ of the CN, AA, RA indices rapidly increase to quite high levels. Other measurements' Pearson coefficients ρ, however, increase very slowly with the data amount and are in general less than 0.2 even when all the data ($\eta = 0.5$) is used. Especially in Amazon which is a very sparse (sparsity is 3.17×10^{-5}) data set, the Pearson coefficients ρ of most measurements are less than 0.03. This result indicates that, for most similarity measurements, the similarity matrixes calculated from different data samples could have no apparent coherence. Overall, more data could make it more stable for most of the measurements especially the CN, AA and RA indices.

To get deeper insight of the comparison and the classification of these similarity measurements, we analyze the results of the average bias μ and standard deviation σ when all of the data is used ($\eta = 0.5$) which is shown in Fig. 3 (Results of the Netflix and Del.icio.us data sets can be found in the Supplementary Information). Using the average bias μ and dispersion σ/μ as two dimensions, we can get the $\mu - \sigma$ location map for each data set. Surprisingly, one can find that, these fifteen similarity measurements could be well classified from the perspective of similarity stability. Except four measurements namely the PA, PC, LHN and IHC indices, the others could be classified into three clusters. Measurements in the same cluster are similar in both mathematical forms and original considerations. The first cluster consists three measurements namely the CN, AA and RA indices, all of which only take into account the information of common neighbours of two target objects. Besides the CN index which considers the number of common neighbours, the AA and RA indices calculate the total number weighted by $1/\log(k_u)$ and $1/k_u$ respectively where k_u is the degree of the common neighbour u of the two target objects. The second cluster consists of six measurements namely the COS, SAL, JAC, SOR, HPI and HDI indices. Except the COS index, the other five measurements are all variations of the CN index. However, another variation of the CN index, namely the LHN index, locates outside the second cluster. The reason may be that, when considering the degree information of two target objects, the LHN index makes the degrees of two objects multiplied, i.e. $k_\alpha k_\beta$, thus the degree information is quadratic in the LHN index. Unlike the LHN index, other variations' degree information is not quadratic, such as $\sqrt{k_\alpha k_\beta}$ of the SAL index, $k_\alpha + k_\beta$ of the SOR index, $\max(k_\alpha, k_\beta)$ of the HDI index and so on (See the Method Section for detailed mathematical definitions of these measurements). The third cluster consists the MD and HC indices which consider the degree information of both the target objects and their common neighbours. Another similar point is that, both the MD and HC indices are designed based on the spreading process on bipartite networks. Although the basic considerations are different, mathematical definitions of the MD and HC indices are very similar which leads to $s_{\alpha\beta}^{MD} = s_{\beta\alpha}^{HC}$. Overall, according to the stability of the object similarity, various similarity measurements could be well classified into three clusters. In fact, the classification can also be observed in Fig. 2, in which measurements in the same cluster always have same dynamical patterns against the data amount parameter η.

Results on the Artificial Data

To explore whether the stability pattern is due to the property of the data set or the nature of each measurement, we present two methods to test the similarity stability on the reshuffled and randomly generated data sets. 1) In the first method, we reshuffle the links between users and objects from the empirical data sets. At each step, we
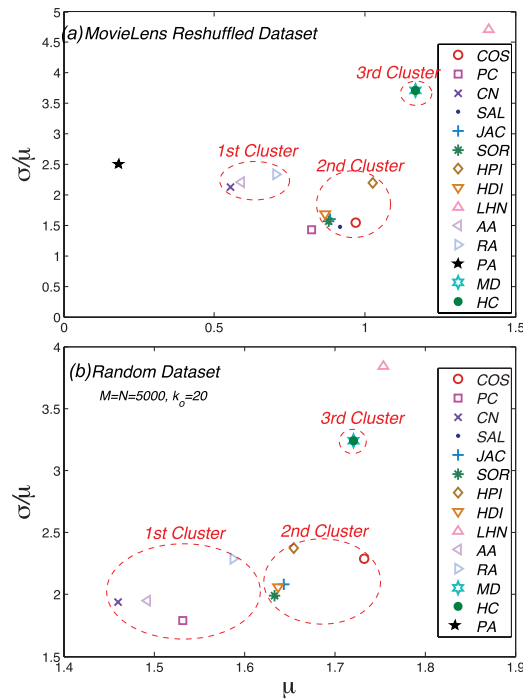
Figure 4. The μ–σ location map for (**a**) reshuffled MovieLens data set with parameter $\eta = 0.5$ and (**b**) randomly generated data set with $M = N = 5000$ and $\langle ko \rangle = 20$. On each map, we still mark the clusters we observed in the empirical results which are shown in Fig. 3. It turns out that, measurements in the same clusters still have similar similarity stabilities even when the user behavior patterns are removed. Thus, the similarity stability patterns are due to the nature of each measurement, not the properties of the data sets.

randomly select two records, say $u_1 - o_1 - r_1$ and $u_2 - o_2 - r_2$, and exchange two objects o_1 and o_2 or two ratings r_1 and r_2 with equal probabilities. After enough steps, the data set would be reshuffled greatly. In our case, we perform $10T$ steps for reshuffles, where T is the number of records in the corresponding empirical data set. 2) The second method is randomly generated data sets. Initially, we suppose there is an empty bipartite network with M objects and N users. Then we randomly generate T links between users and objects. At each step, we randomly select an object and a user, where the selection probabilities of each object and user are $p_{(o)} = (k_o^\gamma + 1)/\sum_{i \in \Gamma_o}(k_i^\gamma + 1)$ and $p_{(u)} = (k_u^\gamma + 1)/\sum_{i \in \Gamma_u}(k_i^\gamma + 1)$ respectively, where Γ_o and Γ_u are the sets of objects and of users. Furthermore, considering the COS and PC measurements, we randomly generate an integer rating ranging from 1 to 5 for each record. In our study, we have $\gamma = 1.4$. Under this scheme, after enough steps, the bipartite network would emerge power-law degree distributions for both users and objects. Note that, with fixed M and N, the number of links T could control the average degree of the bipartite network, say, $\langle k_o \rangle = T/M$ and $\langle k_u \rangle = T/N$.

For the reshuffled and randomly generated data sets, we perform the same calculations used for the empirical data sets. Figure 4 reports the $\mu - \sigma$ location maps of the data sets. The subplot (a) shows the location map of reshuffled data set. Taking the MovieLens data set as an example, the measurements classified into the same cluster are still in the same area on the $\mu - \sigma$ location map. The reshuffle process does not change the stability pattern of each measurement. Actually, the measurements in the same cluster have similar stability trends against the data amount parameter η which is shown in Fig. S7 of the Supplementary Information. For the randomly generated data set, the subplot (b) shows the location map with parameter $M = N = 5000$ and $\langle k_o \rangle = 20$. Although the differences may be larger than the empirical results, the locations of measurements in the same cluster are still close to each other. Since the structure of the randomly generated data set could be controlled, we explore the effect of the structure influence on the similarity stability, and the results are shown in the Supplementary Information. In summary, the results of the reshuffled and randomly generated data sets suggest that the stability patterns come from the nature of each measurement.

Effect on the Recommendation

Object similarities of the user-object bipartite networks are generally used for recommendations[3]. Although the fifteen similarity measurements are widely used in the recommendation systems, the stability of the recommendation regarding to the similarity measurements is still unknown. In this section, we analyze the effect of object similarity stability on the recommendation results. Generally speaking, the goal of a recommendation system is to generate a recommendation list of L objects and voluntarily to display on each user's interface based on the target user's historical selections. The system predict the scores for every unselected objects to a target user u, and rank the objects from high scores to low ones. The score of an object β for the target user u, $w_{u\beta}$, is given by

$$w_{u\beta} = \sum_{\alpha \in \Gamma_u} s_{\alpha\beta}, \tag{4}$$

where Γ_u is the object set which consists with the historical selections of the user u. A high score means that, the system evaluates it as what the target user interests in. To quantify the stability of recommendation results, we divide the data sets as two samples according to the former method with $\eta = 0.5$. For a target user u, there would be two ranking lists of objects. If an object α is ranked at the ith position of the ranking list, we define $R_{u\alpha} = i/M$ as the ranking score where i is object α's position in another ranking list and M is the number of objects. Hence, we can use the average ranking position $\langle R \rangle$ to describe the stability of the recommendation results and $\langle R \rangle$ reads

$$\langle R \rangle = \sum_u \sum_{\alpha \in O_u} \frac{R_{u\alpha}}{|O_u|}, \tag{5}$$

where O_u is the object set ranking at the top L positions of the ranking list and at the same time have not been selected by the target user u in both of the samples, and $|O_u|$ is the number of objects in the set O_u. According to this definition, stable measurements would have small average ranking position $\langle R \rangle$.

According to Eq. (5), we calculate the recommendation stability and find that, many of the recommendations are quite unstable (Table S1). The average ranking position $\langle R \rangle$ of SAL, HPI, LHC, HC and IHC indices are even larger than 0.1 in each data set. Taking the MovieLens data set as an example, $\langle R \rangle = 0.1$ means that, when using the similarities of another sample data, the objects recommended at the top L positions using a data sample are ranked at 585th position (there are 5850 objects in the MovieLens data set). Theoretically, the average ranking position $\langle R \rangle$ of the totally random case is 0.5 and the most stable results is $\langle R \rangle = L/2M$ where M is the number of the objects. Thus, the theoretical best stability is 4.3×10^{-3}, 4.9×10^{-3}, 8.3×10^{-4} and 1.4×10^{-3} for the MovieLens, Netflix, Epinions and Last.FM data sets respectively, which means the recommendation lists of the two data samples are very close. Furthermore, if compare the similarity stability ρ with the recommendation stability $\langle R \rangle$, one may find that, the more unstable the similarity quantification is, the more unstable the recommendation generally would be (Fig. S3).

To improve the stability of the recommendation and explore the effect of the similarity stability, here we present a top-n-stability method. For an object α, the similarity bias $\delta_{\beta\alpha}$ between object α and β, is calculated and ranked from the lowest value (stable) to the highest one (unstable). According to Eq. (4), when adding the score of object α, we only take n objects which have the most stable similarities i.e., which ranks at the top n positions to object α. This could be explained as

$$s_{\beta\alpha} = \begin{cases} s_{\beta\alpha} & P_{\beta\alpha} \leq n \\ 0 & P_{\beta\alpha} > n \end{cases}, \tag{6}$$

where $P_{\beta\alpha}$ is the position of object β in object α's stability list. Note that, unlike the classical top-n-similarity recommendation algorithm in which n objects with the highest similarities to object α would be counted[44,45], here we ignore the exact value of similarity, just consider the stability. The basic consideration is that, if one pair of objects' similarity has poor stability, the similarity would be meaningless regardless of the value of similarity. Through the experiments, the classical top-n-similarity method can also improve the recommendation's stability for a little bit, but the improvement of our top-n-stability method is much bigger (See the Supplementary Information).

With different number of stable objects n, Fig. 5 shows the average ranking position of the recommended objects $\langle R \rangle$, which is summarised according to the similarity measurement clusters. The results of similarity measurements the PC, LHN, PA and IHC indices could be found in the Supplementary Information. One can find that, there is no apparent recommendation stability improvement for the first cluster (the CN, AA, RA indexes) except in the Epinions data set in which the recommendation stability is poor for every similarity measurement. On the other hand, recommendation stability of measurements of the second cluster could be well improved by the top-n-stability method especially for the SAL and HPI indices whose average ranking position $\langle R \rangle$ are over 0.1. However, measurements in the third cluster, i.e. the MD and HC indices, have different patterns against the top-n-stability method. Although the HC index's recommendation stability could be largely improved, the MD index has no apparent improvement. We can conclude that, when the recommendation is unstable, our top-n-stability method could largely improve (See Table S2 for detailed improvement ratio) the stability by taking only the stable similarities into account. For most similarity measurements, when considering around 10% of the similarities, the optimized stability could be reached. And for the poor-stability measurements, the counted ratio may even be about 5%. The improvement indicates that, those unstable similarities are more like false information which would lead to the deflected evaluation of users' true interests.

Conclusion and Discussions

The similarity measurements can evaluate the potential relations between objects in the biological, social, commerce systems, they are meaningful only if the evaluated similarities are stable when the nature of the objects are unchangeable. Unstable similarities are generally false information which would lead to the misunderstanding of the relations between objects. We investigated the stabilities of fifteen similarity measurements for user-object bipartite networks, and found that when measuring the object similarity, most similarity measurements except the PA, CN, AA, RA indices, are quite unstable. The Pearson coefficient ρ of two similarity matrixes calculated from two data samples may be even smaller than 0.2, which means the two matrixes have little correlation. Generally
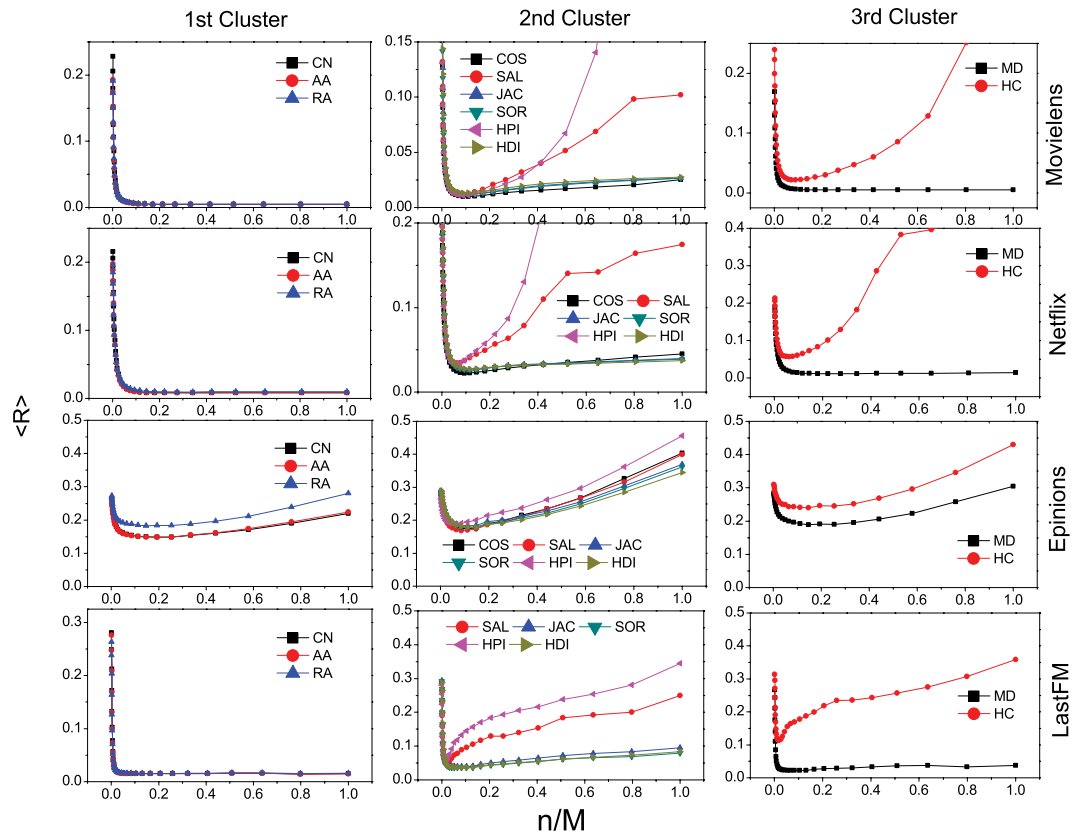
Figure 5. The average ranking position $\langle R \rangle$ of the recommended objects, against number of objects that counted in the top-*n*-stability method. The length of the recommendation list in the simulation is $L = 50$, and the results are averaged over 10 independent simulations. In general, the recommendation stability could be improved by considering only the stable similarities.

speaking, measurements with simple considerations can describe the natural properties of objects and are stable. The CN, AA, RA indices considering only the information of two objects' common neighbours are stable and can be regarded as one cluster. On the other hand, variations of the CN index, namely the SAL, JAC, SOR, HPI, HDI indices, considering further the degree information of two objects, are less stable than the CN index and can be regarded as another cluster. Measurements in the same clusters have in general similar considerations and mathematical definitions, and thus have similar stabilities and even the dynamic against the data amount. In other words, while dozens of measurements have been developed, those similarity measurements can be well classified according to their object similarity stability. When a new measurement is proposed, one just need to analyze its stability to identify which cluster it belongs to, and then could get deeper insight to this measurement by comparing with other measurements within the same cluster. In addition, we presented a top-*n*-stability method to investigate the effect of object similarity on the recommendations. By considering only the stable similarities i.e. deleting the unstable, false information, the stability of the recommendation could be improved.

The investigations and considerations in this paper only focused on the objects. Actually, similarity is also an important method measuring the potential relations of human beings in the social systems and users in the online systems[2,46]. However, different with objects whose natural properties are definitely unchangeable, evidences have been found to prove that, the behaviors and interests of human behavior are temporal[4,47]. Thus, the stabilities of object similarity and human-to-human similarity may have totally different meanings. Additionally, the stability of those similarity measurements should be also studied in one-mode systems, which contain only one kind of nodes. Especially for the objects like genes, proteins etc., the investigations of similarity stability are still urgently needed because those objects may have different properties.

Methods

The data sets used in the this paper are usually modelled as user-object bipartite networks in which nodes can be divided into two groups, representing users and objects respectively. In such kind of system, links only exist between different kinds of nods, i.e. between users and objects. We use α and β to denote the target pair of objects and $U_{\alpha\beta}$ is the set of users who select both objects α and β. The popularity k_α and k_β represent the selection times of object α by users respectively, and the activity k_u is the number of objects user u have selected. We suppose that, the function $\min(x, y)$ equals to the minimum value between x and y and $\max(x, y)$ equals to the maximum value between x and y. In addition, r_α and r_β are rating vectors in the N-dimensional user space and $r_{u\alpha}$ and $r_{u\beta}$ are the ratings user u given to the object α and β respectively. With these defined parameters, the fifteen similarity measurements referred in this paper read:

$$COS: s_{\alpha\beta} = \frac{1}{|r_\alpha||r_\beta|} \sum_{u \in U_{\alpha\beta}} r_{u\alpha} r_{u\beta},$$

$$PC: s_{\alpha\beta} = \frac{\sum_{u \in U_{\alpha\beta}} (r_{u\alpha} - \bar{r}_\alpha)(r_{u\beta} - \bar{r}_\beta)}{\sqrt{\sum_{u \in U_{\alpha\beta}} (r_{u\alpha} - \bar{r}_\alpha)^2} \sqrt{\sum_{u \in U_{\alpha\beta}} (r_{u\beta} - \bar{r}_\beta)^2}},$$

$$CN: s_{\alpha\beta} = \sum_{u \in U_{\alpha\beta}} 1,$$

$$SAL: s_{\alpha\beta} = \frac{1}{\sqrt{k_\alpha k_\beta}} \sum_{u \in U_{\alpha\beta}} 1,$$

$$JAC: s_{\alpha\beta} = \frac{1}{k_\alpha + k_\beta - \sum_{u \in U_{\alpha\beta}} 1} \sum_{u \in U_{\alpha\beta}} 1,$$

$$SOR: s_{\alpha\beta} = \frac{2}{k_\alpha + k_\beta} \sum_{u \in U_{\alpha\beta}} 1,$$

$$HPI: s_{\alpha\beta} = \frac{1}{\min(k_\alpha, k_\beta)} \sum_{u \in U_{\alpha\beta}} 1,$$

$$HDI: s_{\alpha\beta} = \frac{1}{\max(k_\alpha, k_\beta)} \sum_{u \in U_{\alpha\beta}} 1,$$

$$LHN: s_{\alpha\beta} = \frac{1}{k_\alpha k_\beta} \sum_{u \in U_{\alpha\beta}} 1,$$

$$AA: s_{\alpha\beta} = \sum_{u \in U_{\alpha\beta}} \frac{1}{\log k_u},$$

$$RA: s_{\alpha\beta} = \sum_{u \in U_{\alpha\beta}} \frac{1}{k_u},$$

$$PA: s_{\alpha\beta} = k_\alpha k_\beta,$$

$$MD: s_{\alpha\beta} = \frac{1}{k_\alpha} \sum_{u \in U_{\alpha\beta}} \frac{1}{k_u},$$

$$HC: s_{\alpha\beta} = \frac{1}{k_\beta} \sum_{u \in U_{\alpha\beta}} \frac{1}{k_u},$$

$$IHC: s_{\alpha\beta} = \frac{1}{k_\beta^2} \sum_{u \in U_{\alpha\beta}} \frac{1}{k_u}.$$

References

1. Barabási, A. L. Network science: Luck or reason. *Nature* **489,** 1–2 (2012).
2. Papadopoulos, F., Kitsak, M., Serrano, M. Á, Boguná, M. & Krioukov, D. Popularity versus similarity in growing networks. *Nature* **489(7417),** 537–540 (2012).
3. Zhou, T. *et al.* Solving the apparent diversity-accuracy dilemma of recommender systems. *Proc. Natl. Acad. Sci. USA* **107(10),** 4511–4515 (2010).
4. Hou, L., Pan, X., Guo, Q. & Liu, J.-G. Memory effect of online user preference. *Sci. Rep.* **4,** 6560 (2014).
5. Hinton, G. E., Osindero, S. & Teh, Y. A fast learning algorithm for deep belief nets. *Neu. Comp.* **18(7),** 1527C1554 (2006).
6. Haykin, S. Neural networks: a comprehensive foundation 2nd edition. *Upper Saddle River* NJ, the US: Prentice Hall (1999).
7. Watts, D. J. & Strogatz, S. H. Collective dynamics of 'small-world' networks. *Nature* **393(6684),** 440–442 (1998).
8. Barabási, A. L. & Albert, R. Emergence of scaling in random networks. *Science* **286(5439),** 509–512 (1999).

9. Albert, R. & Barabási, A. L. Statistical mechanics of complex networks. *Rev. Mod. Phys.* **74(1),** 47 (2002).
10. Chen, P. & Redner, S. Community structure of the physical review citation network. *J. Informetr.* **4(3),** 278–290 (2010).
11. Lu, W., Janssen, J., Milios, E., Japkowicz, N. & Zhang, Y. Node similarity in the citation graph. *Knowl. Inf. Sys.* **11(1),** 105–129 (2007).
12. D'haeseleer, P., Liang, S. & Somogyi, R. Genetic network inference: from co-expression clustering to reverse engineering. *Bioinformatics* **16(8),** 707–726 (2000).
13. Sawa, T. & Ohno-Machado, L. A neural network-based similarity index for clustering DNA microarray data. *Comput. Biol. Med.* **33(1),** 1–15 (2003).
14. Zhang, B. & Horvath, S. A general framework for weighted gene co-expression network analysis. *Stat. Appl. Genet. Mol.* **4(1)** (2005).
15. Gibson, S. M. *et al.* Massive-scale gene co-expression network construction and robustness testing using random matrix theory. *PLoS ONE* **8(2),** e55871 (2013).
16. Clauset, A., Moore, C. & Newman, M. E. Hierarchical structure and the prediction of missing links in networks. *Nature* **453(7191),** 98–101 (2008).
17. Lü, L. & Zhou, T. Link prediction in complex networks: A survey. *Physica A* **390(6),** 1150–1170 (2011).
18. Stumpf, M. P. *et al.* Estimating the size of the human interactome. *Proc. Natl. Acad. Sci. USA* **105(19),** 6959–6964 (2008).
19. Yu, H. *et al.* High-quality binary protein interaction map of the yeast interactome network. *Science* **322(5898),** 104–110 (2008).
20. Sharan, R. *et al.* Conserved patterns of protein interaction in multiple species. *Proc. Natl. Acad. Sci. USA* **102(6),** 1974–1979 (2005).
21. Al Hasan, M., Salem, S., Pupacdi, B. & Zaki, M. J. Clustering with lower bound on similarity. *In Advances in Knowledge Discovery and Data Mining* (pp. 122–133). Springer Berlin Heidelberg (2009).
22. Pan, Y., Li, D. H., Liu, J. G. & Liang, J. Z. Detecting community structure in complex networks via node similarity. *Physica A* **389(14),** 2849–2857 (2010).
23. Xiang, B., Chen, E. H. & Zhou, T. Finding community structure based on subgraph similarity. In *Complex Networks* (pp. 73–81). Springer Berlin Heidelberg (2009).
24. Armano, G. & Javarone, M. A. Clustering datasets by complex networks analysis. *Complex Adaptive Systems Modeling* **1,** 5 (2013).
25. Frey, B. J. & Dueck, D. Clustering by passing messages between data points. *Science* **315(5814),** 972–976 (2007).
26. Elina, M. & Marika, J. Homogamy in socio-economic background and education, and the dissolution of cohabiting unions. *Demographic Research* **30,** 1769–1792 (2014).
27. Lü, L. Y. *et al.* Recommender systems. *Phys. Rep.* **519,** 1–49 (2012).
28. Guo, Q., Song, W. J. & Liu, J. G. Ultra-accurate collaborative information filtering via directed user similarity. *EPL* **107,** 18001 (2014).
29. Ghoshal, G. & Barabási, A. L. Ranking stability and super-stable nodes in complex networks. *Nat. Comm.* **2,** 394 (2011).
30. Hamers, L. *et al.* Similarity measures in scientometric research: The Jaccard index versus Salton's cosine formula. *Information Processing and Management* **25(3),** 315–318 (1989).
31. Jaccard, P. Étude comparative de la distribution florale dans une portion des Alpes et des Jura. *Bull. Soc. Vaudoise Sci. Nat.* **37,** 547 (1901).
32. Sørensen, T. A method of establishing groups of equal amplitude in plant sociology based on similarity of species content and its application to analysis of the vegetation on Danish commons. *Biol. Skr.* **5,** 1 (1948).
33. Ravasz, E., Somera, A.L., Mongru, D.A., Oltvai, Z.N. & Barabási, A.-L. Hierarchical organization of modularity in metabolic networks. *Science* **297,** 1551 (2002).
34. Leicht, E.A., Holme, P. & Newman, M.E.J. Vertex similarity in networks. *Phys. Rev. E* **73,** 026120 (2006).
35. Adamic, L.A. & Adar, E. Friends and neighbors on the web. *Social Networks* **25,** 211 (2003).
36. Zhou, T., Lü, L. & Zhang, Y.-C. Predicting missing links via local information. *Eur. Phys. J. B* **71,** 623 (2009).
37. Zhou, T., Ren, J., Medo, M. & Zhang, Y. C. Bipartite network projection and personal recommendation. *Phys. Rev. E* **76,** 046115 (2007)
38. Zhang, Y. C., Blattner, M. & Yu, Y. K. Heat conduction process on community networks as a recommendation model. *Phys. Rev. Lett.* **99,** 154301 (2007).
39. Liu, J. G., Zhou, T. & Guo, Q. Information filtering via biased heat conduction. *Phys. Rev. E* **84,** 037101 (2011).
40. Guo, Q., Leng, R., Shi, K. & Liu, J. G. Heat conduction information filtering via local information of bipartite networks. *Eur. Phys. J. B* **85,** 1 (2012).
41. Grujić, J. Movies recommendation networks as bipartite graphs. *Lect. Notes Comput. Sc.* **5102,** 576C583 (2008).
42. Liu, J. G., Hou, L., Zhang, Y. L., Song, W. J. & Pan, X. Empirical analysis of the clustering coefficient in the user-object bipartite networks. *Int. J. Mod. Phys. C* **24,** 1350055 (2013).
43. Ni, J. *et al.* Ceiling effect of online user interests for the movies. *Physica A* **402,** 134 (2014).
44. Deshpande, M. & Karypis, G. Item-based top-n recommendation algorithms. *ACM Transactions on Information Systems (TOIS)* **22(1),** 143–177 (2004).
45. Cremonesi, P., Koren, Y. & Turrin, R. Performance of recommender algorithms on top-n recommendation tasks. In *Proce. of the fourth ACM Conf. on Recomm. Sys.* (39–46). ACM New York, NY, USA (2010).
46. Ji, L., Liu, J. G., Hou, L. & Guo, Q. Identifying the Role of Common Interests in Online User Trust Formation. *PloS ONE* **10(7),** e0121105 (2015).
47. Zhao, Z. D. *et al.* Emergence of scaling in human-interest dynamics. *Sci. Rep.* **3,** 3472 (2013).

Acknowledgements

Author Contributions

L.H. and J.G.L. conceived the idea. L.H. and X.P. performed the calculation. J.G.L., L.H., X.P, Q.G. and T.Z. analyzed the data. L.H. and J.G.L. wrote the manuscript.

Additional Information

Supplementary information accompanies this paper at http://www.nature.com/srep

Competing financial interests: The authors declare no competing financial interests.

How to cite this article: Liu, J.-G. *et al.* Stability of similarity measurements for bipartite networks. *Sci. Rep.* **6,** 18653; doi: 10.1038/srep18653 (2016).