

Coupling diversity across human behavior spaces

CHAO FAN^{1,2}, JUNMING HUANG^{1,3,4}, ZHIHAI RONG^{1,5} and TAO ZHOU^{1,5}

¹ *Complex Lab, School of Computer Science and Engineering, University of Electronic Science and Technology of China - Chengdu 611731, PRC*

² *College of Arts and Sciences, Shanxi Agricultural University - Taiqu 030801, PRC*

³ *Paul and Marcia Wythes Center on Contemporary China, Princeton Institute for International and Regional Studies, Princeton University - Princeton, NJ 08540, USA*

⁴ *Center for Complex Network Research, Department of Physics, Northeastern University - Boston, MA 02115, USA*

⁵ *Big Data Research Center, University of Electronic Science and Technology of China - Chengdu 611731, PRC*

received 20 October 2018; accepted in final form 14 November 2018

published online 3 December 2018

PACS 89.65.-s – Social and economic systems

PACS 87.23.Ge – Dynamics of social systems

PACS 89.75.Fb – Structures and organization in complex systems

Abstract – The heterogeneous nature of human behaviors contributes to the complexity of human-activated systems. Empirical observations and theoretical models reveal the temporal and spatial heterogeneity of many aspects of human behaviors, including social connections and geographic movements, while little is known whether and how human individual's behavioral diversities are correlated across different aspects. With statistical analysis on large-scale data of aligned online and offline human behaviors, we show that behavior spaces are coupled, independent from the specific choice of measurements. The coupling further expands to individual's direct and indirect social contacts. This finding provides insight into understanding homophily in different social systems and further improving the predictability of human online and offline behaviors.

Copyright © EPLA, 2018

Introduction. – Human behaviors are naturally heterogeneous in terms of temporal intervals [1], spatial mobility [2] and social connections [3]. Understanding such diverse behaviors is crucial to model and predict organizational and dynamical characteristics of large-scale human-activated systems. Empirical observations have revealed the burstiness, heavy tails, regularity, periodicity and predictability of human behaviors [1,4–6], and a collection of theoretical models [1,2,7,8] has been proposed to explain the mechanisms underlying these features, with further applications to simulate and predict human behaviors in a wide range of scenarios [9–16].

Diversity is a ubiquitous phenomenon in human social systems [17–21] and manifested in many aspects of human mobility. People with different backgrounds and demographics show far different mobility patterns [22,23]. The heavy-tailed distributions of displacement distance and radius of gyration suggest the diversity of traveling patterns [2,4]. Various motif structures of mobility patterns reveal the diversity of travel sequences in daily movements [24]. People also show various motivations or

preferences of behavior, such as exploring new locations or returning to visited locations when choosing a destination [8,25]. Researchers have been trying to find universal laws of diversities [2,4,26] and the correlation between mobility diversity and economic phenomenon [27,28], while for now it remains challenging to analyze, predict and simulate human behaviors due to the nature of high diversity of such behaviors, especially at the individual level.

To overcome this difficulty, scientists have tried to make use of associated behaviors and social relationships to analyze mobility behavior. In the past, there were not adequate data to bridge the studies on human mobility and social relationships. Fortunately, the recent availability of fusion data combining online social network structure and offline spatio-temporal mobility trajectories provides the opportunity to investigate how human mobility is associated by social relations. Such data include Call Detail Records [29] and Location Based Social Networks [30], such as Twitter [31], Gowalla [32] and Jiepan [33]. Previous studies have revealed that human mobility is strongly influenced by social relationships [34–36]. Therefore, the

use of social relationships can significantly improve the accuracy of location prediction [32]. Besides, the mobility similarity between individuals is correlated with social proximity, revealing the coupling relations between human behaviors in online and offline spaces [37]. However, there is still not sufficient knowledge about the correlation of diversity of human behaviors in different spaces. Is the diversity of individuals' online and offline behavior related? Do those who have a variety of friends on social network also visit various places? Is the diversity similar or different between friends?

This study aims to answer these questions by investigating Location Based Social Networks data. When people use such services, they check in at locations in the physical space through mobile devices and share their experience with online friends. Utilizing their check-in trajectories and friendship information, this paper focuses on the correlation between diversity of individuals' online social connections and offline spatial mobility, as well as its relationship with topological distance. The results show that the diversity of individual social connection and spatial mobility is positively correlated. The diversity of behavior is also positively correlated between indirect friends, with correlation strength decaying with topological distance.

Diversity. – Diversity measures how an individual's behaviors, categorized into labels, differ from each other. For example, a person having lunch at the same bistro every day shows a low diversity, while a person with a high diversity never repeats [38]. In the scope of this study, we consider two types of behaviors: friending and visit, in online and offline spaces, respectively. In each space, we measure diversity of an individual's behaviors in two aspects, namely *variety* and *balance*, borrowed from the semantic interpretation of diversity [39] among a couple of definitions.

Variety measures the number of labels (*i.e.*, different behaviors) of an individual, for example, the total number of unique restaurants a person has ever visited, which is obviously a non-negative integer. Specifically, we measure an individual's online variety as k_i , the *number of friends* in online space, and offline variety as l_i , the *number of locations* visited in offline space.

Balance measures the uniformity of an individual's behaviors, for example, whether she/he regularly rotates among a couple of restaurants, or prefers a certain restaurant in most cases. Technically, we consider that the behaviors of every individual follow a probability distribution over online degree space or offline frequency space. If an individual has friends with various degrees, or visits different locations with similar frequencies, she/he is the most diverse. Among various measurements of diversity [40–42], we use Shannon entropy to formally describe an individual's *balance diversity* as follows:

$$H_i = - \sum_{k \in \{k_j, j \in N(i)\}} p(k) \log(p(k)), \quad (1)$$

which sums over all unique values of degree k_j among the individual's friends $N(i)$, and $p(k)$ is the non-zero fraction of the individual's friends with degree k .

Considering the large space of possible values k_j could take (usually varying from several to hundreds), it is less likely that two friends collide with the same degree, and, therefore, any possible value of friend degree gets an even probability, making balance diversity a trivial inverse of variety diversity. For example, say both users A and B have 5 friends, whose degrees are [6, 7, 8, 9, 10] and [1, 10, 20, 30, 40], respectively. Equation (1) gives $H_A = H_B = 2.322$, while user A is believed intuitively to be less diverse because his friends have similar degrees. To avoid the loss of such information, we coarse-grain k_j into larger bins to merge friends with similar degree: We replace k_j with $\tilde{k}_j = \lfloor k_j/5 \rfloor$, and rewrite eq. (1) as a 5-binning version [43]:

$$\tilde{H}_i = - \sum_{\tilde{k} \in \{\tilde{k}_j, j \in N(i)\}} p(\tilde{k}) \log(p(\tilde{k})), \quad (2)$$

which sums over all unique values of coarse-grained degree \tilde{k}_j among an individual's friends. In the above example, users A and B have friends with coarse-grained degrees [1, 1, 1, 1, 2] and [1, 2, 4, 6, 8], respectively, leading to $\tilde{H}_A = 0.722$ and $\tilde{H}_B = 2.322$, capturing the intuition that A's friends are more similar.

Applying eq. (2) in the online and offline spaces gives our final definitions of *topological diversity* (TD) and *spatial diversity* (SD), *i.e.*, the Shannon entropy within friend degrees and location frequencies, respectively,

$$TD_i = \frac{\tilde{H}_i}{\log(k_i)} = - \frac{\sum_{\tilde{k} \in \{\tilde{k}_j, j \in N(i)\}} p(\tilde{k}) \log(p(\tilde{k}))}{\log(k_i)}, \quad (3)$$

$$SD_i = - \frac{\sum_{d \in D(i)} p(n_{i,d}) \log(p(n_{i,d}))}{\log(l_i)}, \quad (4)$$

where $n_{i,d}$ represents the times users i visits location d , and $p(n_{i,d}) = n_{i,d} / \sum_{d=1}^{l_i} n_{i,d}$ measures the fraction of times she/he visits location d among all her/his locations $D(i)$. Both TD and SD are normalized to break the autocorrelation between variety and balance: an individual with a larger number of friends will automatically have a higher entropy [27]. After normalization, TD/SD of individuals with different size of $N(i)$ and $D(i)$ are fairly comparable.

Both topological and spatial diversity are based on Shannon entropy and range within [0, 1]. The diversity equals 0 when all her/his friends share the same degree, or all her/his visits go to a single destination. Similarly, the diversity goes to 1 when all her/his friends have different degrees even after binning, or her/his visits are evenly distributed over locations.

The heterogeneous nature of the network structure brought by the power-law degree distribution (see fig. 1(a)) produces an inherent autocorrelation between k

and TD since the statistics of high-degree nodes will bring bias to the observation of correlation [44]. Therefore, we leverage the null model technique [45,46] to generate reference networks without bias. A network null model is obtained by randomly rewiring the edges throughout the whole original network, meanwhile it ensures the invariance of the degree of each node. Starting from the original network, we switch two randomly selected edges, namely, unlink (A, B) and (C, D) and then link (A, C) and (B, D), on the premise that (A, C) and (B, D) are not connected. After repeating such operation 10 times the number of edges in a real network, a null network is obtained without changing the degree sequence of the original network. A total of 50 equivalent null models are generated to reduce random errors. We calculate TD of every node on each rewired network and then use Φ to represent the ratio between the diversity value in real network and null models as

$$\Phi(TD) = \frac{TD^{real\ network}}{TD^{null\ models}}. \quad (5)$$

$\Phi > 1$ means that a user's real diversity is higher than its random expectation, and vice versa.

Data description. – The dataset used in this study is collected from a worldwide Location Based Social Networks application, Gowalla. There are nearly 200000 users contributing over 6 million check-in records. The dataset contains the information of social relations and mobility records of each user, including the latitude, longitude and time of each check-in. More detailed descriptions can be seen in [32] and the raw data can be downloaded from *SNAP* website¹.

As a pre-process, we clean the data by 1) merging two check-in records if their interval time is less than 30 minutes, and two locations within a $500\text{ m} \times 500\text{ m}$ area, to remove duplicate check-ins, and 2) filtering out inactive users with less than 10 friends or 5 check-in locations to avoid unreliable statistics on incomplete observations. After that, 26647 users are remaining with 2698029 check-in records from February 2009 to October 2010, forming an undirected network with 254823 edges. The details about the sampling method for active users and the cleaned data are available at the webpage given in footnote².

Results. – Firstly, the probability distributions of the diversity metrics defined above are illustrated in fig. 1. The two variety metrics, shown in the upper panels (*i.e.*, the number of friends and locations), follow power-law distributions, indicating that most users have only a small number of friends and visit a small number of places, while there are a few users who have many friends and have visited lots of places. The two balance metrics, shown in the lower panel (*i.e.*, the topological and spatial diversity) obey right-skewed distributions with mean value deviating from a standard normal distribution to the right. Such dis-

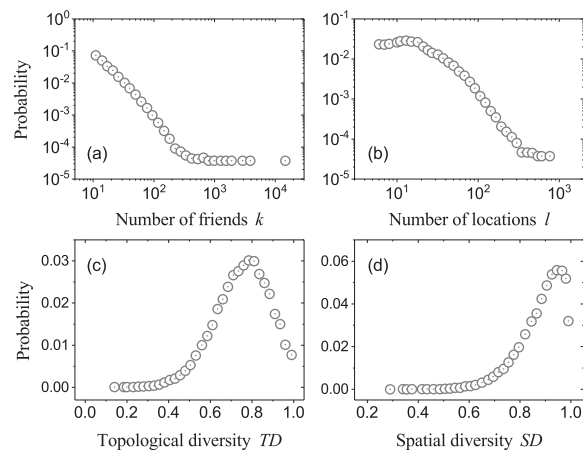


Fig. 1: The probability distributions of diversity metrics. The probabilities of number of friends (a) and number of locations (b) follow power-law distributions with exponent -2.28 and -2.22 , respectively. The probabilities of topological diversity (c) and spatial diversity (d) follow right-skewed distributions with peak at 0.78 and 0.97 , respectively.

tributions imply that most users have friends dissimilar in degrees on social networks and most users unevenly visit places in spatial mobility, namely a large variety of online or offline behavior patterns. While a few users have relatively simple social network structure and mobility pattern. In a word, heterogeneity can be observed through all diversity metrics.

Next, we observe how an individual's multiple diversity metrics correlate, referred to as self-correlation thereafter. We first inspect the correlation within a single space, starting by analyzing the correlation between two online diversity metrics, *i.e.*, the number of friends k and the topological diversity TD . In fig. 2(a), the horizontal and vertical axes represent degree k and ratio $\Phi(TD)$, respectively. A positive correlation between the number of friends and topological diversity is observed after ruling out the autorelation due to the degree of a node. That is to say, popular individuals usually have diverse friends, and vice versa. Moreover, we also find that topological diversity is greater than 1 when an individual has more than 40 friends, namely, the diversity of large-degree users is higher than the randomized value while that of small-degree individuals is lower than the randomized value. This indicates that the large-degree individuals are more willing to make various friends than small-degree individuals.

For offline diversity metrics, *i.e.*, the number of locations l and spatial diversity SD , we also utilize the null model technique to remove the bias brought by the inhomogeneity of visiting patterns (see fig. 1(b)). Specifically, we randomly assign an individual's total check-ins to his/her visited locations, ensuring that each location has been visited at least once and the total number of visits remains unchanged. For example, the visiting pattern of an individual with 100 check-ins at 5 locations, say $[20, 30, 15, 15, 20]$, may be shuffled to $[37, 31, 14, 5, 13]$ or

¹<https://snap.stanford.edu/data/loc-gowalla.html>.

²<http://www.pkbigdata.com/common/share/71.html>.

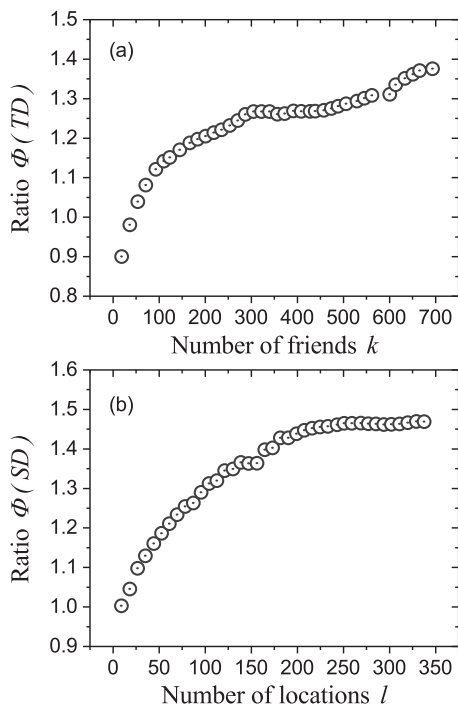


Fig. 2: The self-correlation between diversity metrics in the same space. (a) The correlation between the number of friends k and topological diversity TD . The y -axis represents the ratio of the diversity value between real and null networks. (b) The correlation between the number of locations l and spatial diversity SD . The y -axis represents the ratio of diversity value between real and null trajectories.

[63, 10, 22, 1, 4]. Then SD is recalculated for each randomized sequence. 50 experiments are performed to reduce the statistical errors. Similar to the previous analysis on TD , two groups of results of the correlation between l and SD are obtained, which are based on the real and randomized mobility trajectories, respectively. We use $\Phi(SD)$, which is similar to eq. (5) to represent the ratio between them and show the results in fig. 2(b). The monotonically increasing curve implies positive correlation between l and SD . That is to say, actively moving individuals show more diverging preference over locations. People who have visited more locations show more homogeneous visiting patterns. Furthermore, most part of the curve is higher than 1, indicating that the diversity of real human mobility behavior is higher than the randomized case.

Considering that the number of friends and locations monotonically increases with time, fig. 2 can be interpreted as the fact that both social and spatial diversities grow over time. In other words, people's behaviors become more and more diverse. When people use a social network service or live in a place for a long time, they tend to make more diverse friends or visit more different places, which brings a higher diversity.

Figure 3 reports the self-correlation between diversity metrics in different spaces. Both panels illustrate clearly a consistent positive correlation between online and offline

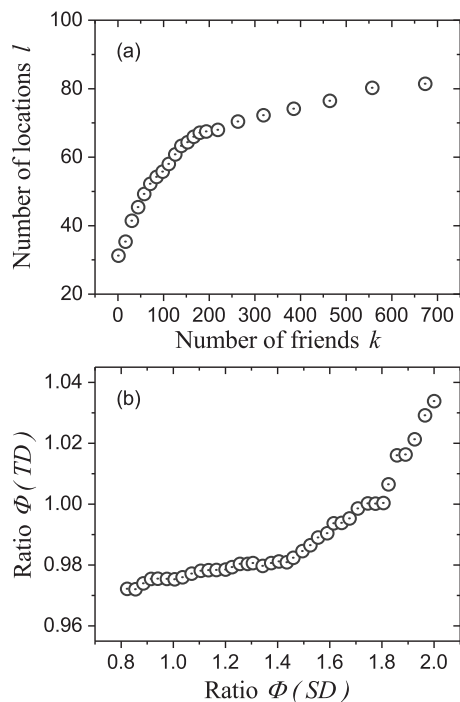


Fig. 3: The self-correlation between diversity metrics across different spaces. (a) The correlation between the number of friends k and number of locations l . (b) The correlation between the ratio of spatial diversity $\Phi(SD)$ and ratio of topological diversity $\Phi(TD)$.

diversity, although with different growing patterns. That is to say, those users who have more friends in online space tend to visit more places in offline space, and those who have more diverse visiting patterns in spatial mobility tend to have more diverse neighborhood structures on social networks. Such a universal pattern indicates the in-depth consistency of people's online and offline behavior.

On the basis of understanding the self-correlation between different diversity metrics, we then inspect the correlation of diversity metrics between neighbors, referred to as social correlation. Since it is known from previous analysis that the heterogeneous network structure will bias the statistic, the user-friend pairs are sampled without overlapping to avoid auto-coupling. Specifically, we use the sampling without replacement method to pick out the pairs to ensure that every individual appears in the sample only once to avoid the influence from the hub nodes. In the sampling process, once an edge is chosen, the two individuals connected by it as well as all the edges attached to them are removed from the sample pool. Fifty parallel experiments are performed to reduce the statistical errors and only 63080 edges on average are sampled in each experiment. We observe the correlation of a diversity metric between neighbors, and then average it over 50 experiments. The results are shown in fig. 4. The four curves all show positive correlations, indicating an assortative mixing pattern between individuals and their friends regardless of the diversity metrics. This tells that it is common to

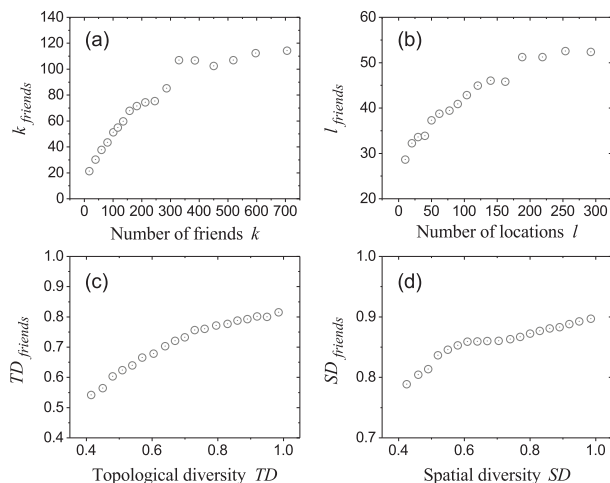


Fig. 4: The social correlation of diversity metrics including the number of friends (a), number of locations (b), topological diversity (c) and spatial diversity (d). The x -axis and y -axis represent the diversity metrics of individuals and their direct friends, respectively.

observe friends with similar behavior patterns, in terms of making diverse friends or visiting different locations. This similarity is consistent with the homophily phenomenon observed in sociological studies [47–50].

Upon the positive social correlation between diversity metrics, we extend the interest to a more generalized form: How does an individual’s behavior correlate with his/her friend’s friend, and even further? We measure the *topological distance* between an individual and an indirect friend with the length of the shortest path between them on a social network, denoted as *hop*. We use the Pearson correlation coefficient r to quantify the correlation between diversity metrics of two individuals with different topological distance from 1 to 6. The logarithm form of diversity metrics is used to calculate Pearson’s r . As shown in fig. 5, the average r for various diversity metrics consistently illustrate that direct neighbors have the strongest correlation, and increased topological distance weakens the correlation. When the distance reaches 3 hops, only the correlation of topological diversity is greater than 0.1. When it goes to 4 and beyond, all correlation coefficients converge to 0. That is to say, the correlation of behavior between individuals is persistent within 3 hops. This phenomenon also confirms the “three degrees of influence” theory in the realm of social network [51–53]. Furthermore, it is noticeable that the points of online diversity (squares and up-triangles) are significantly higher than those of offline diversity (circles and down-triangles) when the topological distance is 1. This means that the online connection between friends is much higher than offline connection, indicating that people are more likely to cluster on social networks while less synchronized when traveling in real life.

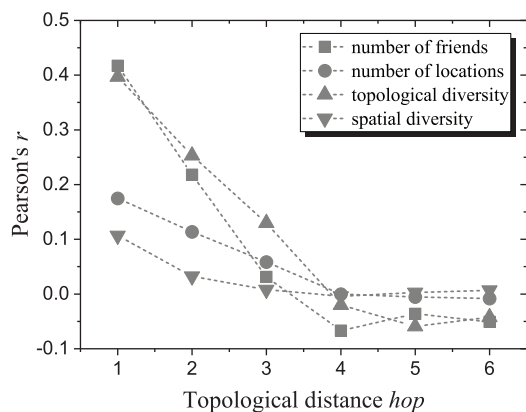


Fig. 5: The relationship between Pearson coefficient and topological distance. Pearson’s r measure the correlation of diversity metrics including the number of friends (squares), number of locations (circles), topological diversity (up-triangles) and spatial diversity (down-triangles) between pairs who are several hops away.

Discussions. – The popularity of the mobile Internet applications provides opportunities for the study on the coupling of people’s online and offline behaviors. By analyzing data of Location Based Social Networks, this study investigated the correlation between diversity of people’s online social network and offline mobility pattern as well as its spread on topology. The comprehensive experiments demonstrate a universal positive and metric-independent correlation between the diversity of human behaviors, which exists not only between online and offline behavior of an individual, but also between neighboring individuals. Specifically, people with many friends have a greater chance to visit a large number of places, and those who have diverse visiting pattern also have diverse online social network structure. The positive correlation between connected individuals implies that the social network is assortatively constructed. Such a positive correlation decays with the topological distance between individuals, and will disappear after 3 hops.

Our study provides a clear picture of the relationship between human behaviors in different spaces in the perspective of diversity. The coupling between behaviors could reduce the complexity of the associated system, further reduce the dimension of features when analyzing online and offline behaviors. Besides, if an offline co-occurrence network, which is obtained by connecting those people who appear in the same place at the same time, is coupled with the online social network, the so-called multiplex networks [54] could be used to analyze the characters of some dynamical processes, such as spreading [55]. Furthermore, our methodology for analyzing correlation could also be used for other researches on human behaviors. It should be noted that both the normalization and null model are designed to eliminate autocorrelation when analyzing diversity measurements. Such autocorrelation in self-correlation analysis (see figs. 2 and 3) can be eliminated

by either method, but that in social correlation analysis (see figs. 4 and 5) can only be eliminated by normalization. Therefore, we keep both of the two methods.

Finally, diversity can be understood from the perspective of predictability as entropy is negatively associated with predictability [6,15]. Specifically, a user's high topological diversity means that the degrees of his friends are widely distributed, so it is difficult to guess whether a friend is of a high or a low degree. Similarly, a high spatial diversity indicates that the user evenly visits various locations, leading to the difficulty to guess the next location she/he will visit. Thus, we hope that this study could give insights into the predictability of human behaviors, especially the location prediction with the help of social relations.

* * *

This work was partially supported by the National Natural Science Foundation (Nos. 61473060, 61433014, 61603074, 61673085, 71731004 and 61803245), the Open Project Program of State Key Laboratory of Theoretical Physics, Institute of Theoretical Physics, Chinese Academy of Sciences, China (No. Y5KF201CJ1), Major Project of the National Social Science Grant (No. 12-ZD218), the Fundamental Research Funds for the Central Universities (No. ZYGX2016J192) and the Science Promotion Programme of UESTC (No. Y03111023901014006). Research reported in this publication was partly supported by The Eunice Kennedy Shriver National Institute of Child Health & Human Development of the National Institutes of Health under Award No. P2CHD047879. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

REFERENCES

- [1] BARABÁSI A.-L., *Nature*, **435** (2005) 207.
- [2] BROCKMANN, D., HUFNAGEL, L. and GEISEL, T., *Nature*, **439** (2006) 462.
- [3] SZELL M., LAMBIOTTE R. and THURNER S., *Proc. Natl. Acad. Sci. U.S.A.*, **107** (2010) 13636.
- [4] GONZALEZ M. C., HIDALGO C. A. and BARABÁSI A.-L., *Nature*, **453** (2008) 779.
- [5] RYBSKI D., BULDYREV S. V., HAVLIN S., LILJEROS F. and MAKSE H. A., *Proc. Natl. Acad. Sci. U.S.A.*, **106** (2009) 12640.
- [6] SONG C., QU Z., BLUMM N. and BARABÁSI A. L., *Science*, **327** (2010) 1018.
- [7] MALMGREN R. D., STOUFFER D. B., MOTTER A. E. and AMARAL L. A., *Proc. Natl. Acad. Sci. U.S.A.*, **105** (2008) 18153.
- [8] SONG C., KOREN T., WANG P. and BARABÁSI A.-L., *Nat. Phys.*, **6** (2010) 818.
- [9] MAKSE H. A., HAVLIN S. and STANLEY H. E., *Nature*, **377** (1995) 779.
- [10] JUNG W.-S., WANG F. and STANLEY H. E., *EPL*, **81** (2008) 48005.
- [11] VESPIGNANI A., *Science*, **325** (2009) 425.
- [12] BARTHÉLEMY M., *Phys. Rep.*, **499** (2011) 1.
- [13] LÜ L., MEDO M., YEUNG C. H., ZHANG Y. C., ZHANG Z. K. and ZHOU T., *Phys. Rep.*, **519** (2012) 1.
- [14] WESOLOWSKI A., QURESHI T., BONI M. F. *et al.*, *Proc. Natl. Acad. Sci. U.S.A.*, **112** (2015) 11887.
- [15] XU P., YIN L., YUE Z. and ZHOU T., Arxiv preprint, arXiv:1806.03876 (2018).
- [16] FAN C. and GUO J. L., *Chin. Phys. B*, **21** (2012) 070209.
- [17] PERC M. and SZOLNOKI A., *Phys. Rev. E*, **77** (2008) 011904.
- [18] CHEN D. B., XIAO R., ZENG A. and ZHANG Y. C., *EPL*, **104** (2014) 68006.
- [19] ZHOU T., KUSCSIK Z., LIU J. G., MEDO M., WAKELING J. R. and ZHANG Y. C., *Proc. Natl. Acad. Sci. U.S.A.*, **107** (2010) 4511.
- [20] GUO Q., JI L., LIU J. G. and HAN J., *Physica A*, **468** (2017) 698.
- [21] NGUYEN V. X., XIAO G., XU X. J., LI G. and WANG Z., *EPL*, **121** (2018) 26002.
- [22] YAN X.-Y., HAN X.-P., WANG B.-H. and ZHOU T., *Sci. Rep.*, **3** (2013) 2678.
- [23] LENORMAND M., LOUAIL T., CANTÚ-ROS O. G. *et al.*, *Sci. Rep.*, **5** (2015) 10075.
- [24] SCHNEIDER C. M., BELIK V., COURONNÉ, T., SMOREDA Z. and GONZÁLEZ M. C., *J. R. Soc. Interface*, **10** (2013) 20130246.
- [25] PAPPALARDO L., SIMINI F., RINZIVILLO S. *et al.*, *Nat. Commun.*, **6** (2015) 8166.
- [26] YAN X. Y., WANG W. X., GAO Z. Y. and LAI Y. C., *Nat. Commun.*, **8** (2017) 1639.
- [27] EAGLE N., MACY M. and CLAXTON R., *Science*, **328** (2010) 109.
- [28] ŠĆEPANOVIĆ, S., MISHKOVSKI I., HUI P., NURMINEN J. K. and YLÄ-JÄÄSKI A., *PLoS ONE*, **10** (2015) e0124160.
- [29] BLONDEL V. D., DECUYPER A. and KRINGS G., *EPJ Data Science*, **4** (2015) 10.
- [30] LIAN D., XIE X., ZHANG F. *et al.*, *IEEE Data Eng. Bull.*, **38** (2015) 35.
- [31] JURDAK R., ZHAO K., LIU J., ABOUJAOUDE M., CAMERON M. and NEWTH D., *PLoS ONE*, **10** (2015) e0131469.
- [32] CHO E., MYERS S. A. and LESKOVEC J., in *Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (ACM Press) 2011, pp. 1082–1090.
- [33] YANG Z., LIAN D., YUAN N. J., XIE X., RUI Y. and ZHOU T., *Physica A*, **469** (2017) 232.
- [34] WANG D., PEDRESCHI D., SONG C., GIANNOTTI F. and BARABÁSI A.-L., in *Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (ACM Press) 2011, pp. 1100–1108.
- [35] TOOLE J. L., HERRERA-YAQUÉ C., SCHNEIDER C. M. and GONZÁLEZ M. C., *J. R. Soc. Interface*, **12** (2015) 20141128.
- [36] LU X., BENGTTSSON L. and HOLME P., *Proc. Natl. Acad. Sci. U.S.A.*, **109** (2012) 11576.
- [37] FAN C., LIU Y., HUANG J., RONG Z. and ZHOU T., *Sci. Rep.*, **7** (2017) 11975.
- [38] CAO Y., GAO J., LIAN D. *et al.*, *J. R. Soc. Interface*, **15** (2018) 20180210.

-
- [39] STIRLING A., *J. R. Soc. Interface*, **4** (2007) 707.
- [40] GINI C., *Studi Economico-Giuridici della R. Università di Cagliari*, **3** (1912) 3.
- [41] SIMPSON A., *Nature*, **163** (1949) 4148.
- [42] SHANNON C. E., *Mob. Comput. Commun. Rev.*, **5** (2001) 3.
- [43] BRADLEY K. S. and SCHATZ G. C., *J. Phys. Chem.*, **98** (1994) 3788.
- [44] YANG D., PAN L. and ZHOU T., *Chaos*, **27** (2017) 033113.
- [45] MASLOV S. and SNEPPEN K., *Science*, **296** (2002) 910.
- [46] MILO R., ITZKOVITZ S., KASHTAN N. *et al.*, *Science*, **303** (2004) 1538.
- [47] KANDEL D. B., *Am. J. Sociol.*, **84** (1978) 427.
- [48] IBARRA H., *Admin. Sci. Q.*, **37** (1992) 422.
- [49] MCPHERSON M., SMITH-LOVIN L. and COOK J. M., *Annu. Rev. Sociol.*, **27** (2001) 415.
- [50] ZHANG Q. M., XU X. K., ZHU Y. X. and ZHOU T., *Sci. Rep.*, **5** (2015) 10350.
- [51] CHRISTAKIS N. A. and FOWLER J. H., *N. Engl. J. Med.*, **357** (2007) 370.
- [52] CHRISTAKIS N. A. and FOWLER J. H., *N. Engl. J. Med.*, **358** (2008) 2249.
- [53] FOWLER J. H. and CHRISTAKIS N. A., *Br. Med. J.*, **337** (2008) a2338.
- [54] BOCCALETTI S., BIANCONI G., CRIADO R. *et al.*, *Phys. Rep.*, **544** (2014) 1.
- [55] SORIANO-PAÑOS D., LOTERO L., ARENAS A. and GÓMEZ-GARDEÑES J., *Phys. Rev. X*, **8** (2018) 031039.