Research Article

Diagnosing and Predicting the Earth's Health via Ecological Network Analysis

Zi-Ke Zhang,^{1,2} Ye Sun,¹ Chu-Xu Zhang,^{1,3} Kuan Fang,³ Xiang Xu,¹ Chuang Liu,¹ Xueqi Wang,⁴ and Kui Zhang²

¹ Institute of Information Economy, Hangzhou Normal University, Hangzhou 310036, China

² College of Communication Engineering, Chongqing University, Chongqing 400044, China

³ Web Sciences Center, University of Electronic Science and Technology of China, Chengdu 610054, China

⁴ Division of Translational Medicine, Shanghai Changzheng Hospital, Second Military Medical University, Shanghai 200003, China

Correspondence should be addressed to Kui Zhang; zk@cqu.edu.cn

Received 29 October 2013; Accepted 6 December 2013

Academic Editor: Zhen Jin

Copyright © 2013 Zi-Ke Zhang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Ecological balance is one of the most attractive topics in biological, environmental, earth sciences, and so on. However, due to the complexity of ecosystems, it is not easy to find a perfect way to conclusively explain all the potential impacts. In this paper, by considering several important elements, we seek to build a dynamic network model to predict the Earth's health, trying to identify and explain how the human behavior and policies affect the model results. We firstly empirically analyze both the topological properties and time-dependent features of nodes and propose an Earth's health index based on Shannon Entropy. Secondly, we identify the importance of each element by a machine learning approach. Thirdly, we use a spreading model to predict the Earth's health. Finally, we integrate the topological property and the proposed health index to identify the influential nodes in the observed ecological network. Experimental results show that the oceans are the key nodes in affecting the Earth's health, and Big countries are also important nodes in influencing the Earth's health. In addition, the results suggest a possible solution that returning more living lands might be an effective way to solve the dilemma of ecological balance.

1. Introduction

Ecological balance is one of the most attractive topics in biological, environmental, Earth sciences and many other related disciplines [1], especially since the industrialization has been undergoing for about two hundred years. To better understand how biosphere responds to the increasing pressure (e.g., population explosion, water and air pollution, climate change), there is a vast class of researches devoted to discovering possible solutions in alleviating pains of the Earth. However, due to the complexity of ecosystems, it is not easy to find a perfect way to conclusively explain all the potential impacts [2] that is responsible for the ecological fragility [3, 4]. Among various studies, *Ecological Network Analysis* (ENA) [5, 6] is regarded as one promising methodology to assess the Earth's health [7].

Ecological networks can be extracted from various information, resulting in different kinds of networks, where each

node represents a nation, a continent, an ocean, a habitat, or a park, and an edge is present when two nodes are directed or mutually influenced, varying from economic impact, population flow to environmental pollution, economic level, and so forth [5, 8].

Since the network information is not explicitly provided, we start our research by constructing the global ecological network via extracting nodes and links from the world geography map (see Figure 1). We consider each nation or ocean as one node, and a link is present if the two countries/oceans (or country and ocean) are geographically neighboring. For example, the two nodes, China and Russia, are connected since they are neighboring countries. In addition, there is also a link between the Pacific Ocean and China since they are also mutually connected. Figure 2 shows the constructed network. Furthermore, we also consider five time-dependent factors of each node, including the population size, economic level,

FIGURE 2: The 145-nation (including 19 oceans) network, where edges represent the geographically neighboring relationship. The positions of the nodes are drawn according to their respective centrality index, regardless of the geographical positions. The size of each node is set according to the stationary value of (14).

where $L_{ij} = 1$ denotes the length of the shortest path between nodes i and j .

(iv) *k-Core Index*. *k-core* [23], denoted as $S_K(i)$, is the core number of a node which is the largest value k of a k -core containing that node. It is obtained as follows: (i) remove from the graph all nodes of degree less than k and (ii) then

remove these vertices repeatedly until no further removal is possible. The remaining result, if exists, is the k -core. Thus, a network is organized as a set of successively enclosed k -cores.

2.2. Dynamical Factors. Besides the static network properties, we collect various data from World Bank

(<http://data.worldbank.org/>) to investigate the dynamical factors of all the nodes in the ecological network. The dataset includes population size, per capita GDP, area of land and marine, energy consumption per unit of GDP, and carbon dioxide emission of each country from year 1962 to 2011. Specifically, we observe the five following factors for each node.

- (i) *Population size*, denoted as PS_t^i , is the total population of node i in the year t .
- (ii) *Economic level*, denoted as EL_t^i , is the per capita GDP of node i in the year t .
- (iii) *Habitats area*, denoted as HA_t^i , is the total area of land and marine of node i in the year t .
- (iv) *Energy consumption*, denoted as EC_t^i , is energy consumption per unit of GDP of node i in the year t .
- (v) *Air pollution*, denoted as AP_t^i , is the total amount of carbon dioxide emission of node i in the year t .

Figure 3 shows how the five factors change from the year 1961 (HA starts from 1990 and EC starts from 1980 due to the data absence) to 2011. It can be seen that, generally, the values of all the factors increase year by year. It also shows that the population size has highly positive relationship with economic level. For example, the population size of China (CHN) and USA are both in the top five nations, and their average GDP also have high ranks among all the 126 nations. Meanwhile, their air pollution is ranked in the worst five nations, which might suggest that the development of economy would have negative impact on the environment. Furthermore, we list the top 20 nodes for both static properties and dynamical factors in Table 2. It can be seen that the Atlantic Ocean (ATO) holds the most significant role in maintaining the robustness of the ecological network because it connects the largest number of nodes (Table 3). Russia (RUS) has high network property, with degree rank number 5, Betweenness rank number 2, and Closeness rank number 2, but simultaneously has a relatively bad air quality (rank number 3). Other nations, such as China (CHN), have the similar situation. Comparatively, USA is not ranked in the top network structure list but has a large population (rank number 3) and a high economic level (rank number 6), which might promote its impact in affecting the global Earth's health.

2.3. Definition of the Earth's Health Index. Inspired by previous analyses, we consider that the Earth's health is not just affected by a single factor, but a joint influence resulting from many complicated factors. In this paper, we use the *Shannon Entropy* [24–26] to integrate the impact of all possible factors to characterize the Earth's health, denoted by EH, as

$$EH = -\sum_{i=1}^{|V|} \sum_{j=1}^{|F|} p_{f_i^j} \log p_{f_i^j}, \quad (4)$$

where $f_i^j = (f_i^j - f_{\min}^j) / (f_{\max}^j - f_{\min}^j)$ is the normalized fraction of factor j for node i , $|F|$ denotes the set of all the factors defined in Section 2.2, and the final Earth's health value, EH, runs over the sum of all nodes. According to the original definition of *Shannon Entropy*, the larger the entropy value is, the more equal the distribution will be. Therefore, a large value of EH suggests a good situation of ecological balance both among nations and factors and hence indicates good Earth's health and vice versa (Table 4).

2.4. Identifying the Influential Factor via Machine Learning Approach

2.4.1. Random Forest. We use *Random Forest* [27] to evaluate the importance of factors. *Random Forest* is an ensemble regressor/classifier that consists of many decision trees and then outputs the value that is the mode of the values/classes output by individual trees. In this scenario, we apply this method in a regression way and use it to evaluate each feature's importance. Compared with other regression models, we choose the *Random Forest* model because of its following advantages for our solution:

- (i) it can tackle high-order variable interactions or correlated predictor variables;
- (ii) it can be used not only for prediction, but also to assess variable importance;
- (iii) it can partially overcome the overfitting problem.

2.4.2. Base Learner: Classification and Regression Tree (CART). Given a training vector $x_i \in R^n$, $i = 1, \dots, l$ and a label vector $y_i \in R^l$, the decision tree recursively partitions the space such that the samples with the same labels can be classified together. Let the data at node m be represented by Q , for each candidate split $\theta = (j, t_m)$ consisting of a feature j and threshold t_m ; partition data into two parts:

$$Q_{\text{right}(\theta)} = \frac{Q}{Q_{\text{left}(\theta)}}, \quad (5)$$

$$Q_{\text{left}(\theta)} = (x, y) \mid x_j \leq t_m. \quad (6)$$

The impurity at m is computed using an impurity function H , the choice of which depends on the task being solved:

$$G(Q, \theta) = \frac{n_{\text{left}}}{N_m} H(Q_{\text{left}(\theta)}) + \frac{n_{\text{right}}}{N_m} H(Q_{\text{right}(\theta)}). \quad (7)$$

Then, select the parameters that minimize the impurity:

$$\theta^* = \arg \min_{\theta} G(Q, \theta). \quad (8)$$

After this recurse for subsets $Q_{\text{left}(\theta^*)}$ and $Q_{\text{right}(\theta^*)}$ until the maximum allowable depth is reached, $N_m < \text{minSamples}$. In regression problem, for a node m , representing a region

FIGURE 3: Five factors of evolution of all nodes versus time. (a) Population; (b) economic level; (c) habitats area; (d) energy consumption per GDP; (e) air pollution. In each subgraph, five nations with top amount of corresponding property are highlighted.

TABLE 2: Top 20 nodes of the static and dynamical properties.

Rank	S_D	S_B	S_C	S_K	PS	EL	HA	EC	AP
1	ATO	ATO	ATO	ATO	CHN	NOR	VEN	PER	USA
2	PCO	RUS	RUS	PAK	IND	CHE	DEU	COL	CHN
3	CHN	CHN	FRA	TKM	USA	DNK	ECU	CHE	RUS
4	MSO	PCO	NHO	UZB	IDN	ISL	ZMB	BWA	IND
5	RUS	MSO	ZAF	KAZ	BRA	IRL	NIC	IRL	JPN
6	INO	INO	MRT	AFG	PAK	USA	BWA	PAN	DEU
7	DEU	FRA	CHN	AZE	RUS	SWE	SAU	CRI	CAN
8	IRN	ABO	MDG	IRN	BGD	JPN	GTM	GAB	GBR
9	COD	UKR	JPO	TUR	NGA	NLD	TZA	ITA	KOR
10	ABO	JPO	PCO	FRA	JPN	FIN	ZWE	DNK	IRN
11	CBO	ZAF	CAN	FIN	MEX	AUT	DOM	ALB	ITA
12	KAZ	IRN	USA	SWE	PHL	GBR	CHE	GRC	MEX
13	UKR	MRT	ARG	AUT	DEU	BEL	SEN	TUR	ZAF
14	SDN	ERI	MEX	CZE	VNM	CAN	AUT	PRT	FRA
15	FRA	IDN	HND	DEU	EGY	DEU	SVK	GBR	SAU
16	HUN	NHO	CRI	HUN	ETH	FRA	BRA	ESP	AUS
17	POL	TUR	PAN	POL	IRN	AUS	EST	AUT	BRA
18	AUT	DEU	INO	SVK	TUR	ITA	KHM	NAM	IDN
19	BSO	BBO	NAM	CHE	THA	SGP	CIV	ECU	UKR
20	AFG	SDN	DZA	BLR	FRA	ESP	POL	SGP	ESP

S_D denotes the degree index, S_B denotes the Betweenness index, S_C denotes the Closeness index, and S_K denotes the k -core index, respectively. PS, EL, HA, EC, and AQ, respectively, denote the population size, economic level, habitats area size, energy consumption, and air pollution in the year of 2011.

TABLE 3: Abbreviations of ocean nodes.

Abbr.	Full name
ATO	Atlantic Ocean
PCO	Pacific Ocean
INO	Indian Ocean
ARO	Arctic Ocean
MSO	Mediterranean Sea
CBO	Caribbean Sea
BSO	Black Sea
ABO	Arabian Sea
BLO	Baltic Sea
NOO	Norwegian Sea
CPO	Caspian Sea
ACO	Antarctic Ocean
JPO	Sea of Japan
NSO	North Sea
RSO	Red Sea
CSO	South China Sea
BBO	Bay of Bengal
GMO	Gulf of Mexico
GTO	Gulf of Thailand

R_m with N_m observations, we choose the criterion of Mean Squared Error (MSE) as impurity function H :

$$c_m = \frac{1}{N_m} \sum_{i \in N_m} y_i, \quad (9)$$

$$H(X_m) = \frac{1}{N_m} \sum_{i \in N_m} (y_i - c)^2.$$

2.4.3. Construction of Random Forest. Let the number of training cases be N and let the number of variables in the regressor be M . We are told the number m of input variables to be used to determine the decision at a node of the tree; m should be much less than M .

- (i) Choose a training set for this tree by choosing n times with replacement from all N available training cases (i.e., take a bootstrap sample). Use the rest of the cases to estimate the error of the tree.
- (ii) For each node of the tree, randomly choose m variables on which to base the decision at that node. Calculate the best split based on these m variables in the training set.
- (iii) Each tree is fully grown and not pruned (as may be done in constructing a normal tree classifier). For prediction a new sample is pushed down the tree. It is assigned the value of the training sample in the terminal node it ends up in.

2.4.4. Out-of-Bag (OOB) Evaluation. To evaluate the *Random Forest* model, we use 2/3 of the data as training set and remaining 1/3 (saying *Out-of-Bag*) are regarded as the test set when constructing the base learner. We calculate the average result from 50-round simulations to alleviate the random fluctuations. For each tree t , the *Out-of-Bag* (OOB) simulation is tested in detail as follows.

- (i) Consider the associated OOB_t sample.
- (ii) Denote by err_{OOB_t} the error of a single tree t on this OOB_t sample.

TABLE 4: Abbreviations of nation nodes.

Abbr.	Full name	Abbr.	Full name
AFG	Afghanistan	AGO	Angola
ALB	Albania	ARE	United Arab Emirates
ARG	Argentina	AUS	Australia
AUT	Austria	AZE	Azerbaijan
BEL	Belgium	BFA	Burkina Faso
BGR	Bulgaria	BGD	Bangladesh
BIH	Bosnia and Herzegovina	BLR	Belarus
BOL	Bolivia	BRA	Brazil
BWA	Botswana	CAF	Central African Republic
CAN	Canada	CHE	Switzerland
CHL	Chile	CHN	China
CIV	Cote D'Ivoire	CMR	Cameroon
COD	Congo	COL	Colombia
CRI	Costa Rica	CUB	Cuba
CZE	Czech Republic	DEU	Germany
DNK	Denmark	DOM	Dominican Republic
DZA	Algeria	ECU	Ecuador
ERI	Eritrea	ESP	Spain
EST	Estonia	ETH	Ethiopia
FIN	Finland	FRA	France
GAB	Gabon	GBR	United Kingdom
GHA	Ghana	GIN	Guinea
ALB	Albania	GTM	Guatemala
HND	Honduras	HRV	Croatia
HTI	Haiti	HUN	Hungary
IDN	Indonesia	IND	India
IRL	Ireland	IRN	The Islamic Republic of Iran
IRQ	Iraq	ISL	Iceland
ITA	Italy	JAM	Jamaica
JOR	Jordan	JPN	Japan
KAZ	Kazakhstan	KEN	Kenya
KGZ	Kyrgyzstan	KHM	Cambodia
KOR	Korea	LAO	Laos
LBR	Liberia	LBY	Libya
LKA	Sri Lanka	LTU	Lithuania
LVA	Latvia	MDA	Moldova
MDG	Madagascar	MEX	Mexico
MKD	Macedonia	MLI	Mali
MMR	Myanmar	MNG	Mongolia
MOZ	Mozambique	MRT	Mauritania
MYS	Malaysia	NAM	Namibia
NER	Niger	NGA	Nigeria
NIC	Nicaragua	NLD	Netherlands
NOR	Norway	NPL	Nepal
NZL	New Zealand	OMN	Oman
PAK	Pakistan	PAN	Panama
PER	Peru	PHL	Philippines
POL	Poland	PRK	North Korea
PRT	Portugal	ROU	Romania
RUS	Russian Federation	SAU	Saudi Arabia

TABLE 4: Continued.

Abbr.	Full name	Abbr.	Full name
SDN	Sudan	SEN	Senegal
SGP	Singapore	SLV	El Salvador
SOM	Somalia	SRB	Serbia
SSD	Southern Sudan	SVK	Slovak Republic
SVN	Slovenia	SWE	Sweden
TCD	Chad	THA	Thailand
TJK	Tajikistan	TKM	Turkmenistan
TUR	Turkey	TZA	Tanzania
UGA	Uganda	UKR	Ukraine
USA	United States	UZB	Uzbekistan
VEN	Bolivarian Republic of Venezuela	VNM	Vietnam
YEM	Republic of Yemen	ZAF	South Africa
ZMB	Zambia	ZWE	Zimbabwe

(iii) Randomly permute the values of X^j in OOB_t to get a permuted sample denoted by OOB_t^j , the error of predictor t on the perturbed sample:

$$\text{err}(X^j) = \frac{1}{N_{\text{tree}}} \sum_t (\text{err} OOB_t^j - \text{err} OOB_t), \quad (10)$$

where X^j represents j th variable of each data. Then we can get $\text{err}(X^j)$ for each j and normalize it. The bigger the $\text{err}(X^j)$ is, the more important the variable is. Figure 4 shows the $\text{err}(X)$ result for all the five features. It can be seen that, among all the features, the $\text{err}(X^j)$ value of habitats area size is the biggest, hence the most important factor for the Earth's health, which is in the agreement with reports by public media that people, especially humans living cities, now occupy smaller and smaller space than before, resulting in comparatively a much worse living condition.

2.5. Predicting the Earth's Health in Ecological Networks. We use a dynamic spreading model [28] to predict the Earth's health by considering the observed ecological network structure. The model runs as follows.

- (i) At the initial step, each node i is set a health value by averaging the Earth's health index (EH_i^0 defined by (4)) from the most recent five years.
- (ii) We then choose 10% of the nodes as "seed" nodes and add ΔEH_i^0 to each seed node. The ΔEH_i^0 is calculated by averaging over the incremental EH values of the most recent five years.
- (iii) At each time step $t + 1$, each node i will affect all its neighbouring nodes' Earth's health index by

$$\Delta EH_{i \rightarrow j}^{t+1} = \alpha * \Delta EH_i^t * (EH_i^t * EH_j^t)^\beta, \quad (11)$$

where $\Delta EH_{i \rightarrow j}^{t+1}$ is the Earth's health influence from node i to j and α and β are tunable parameters.

- (iv) Then, the node j 's Earth's health index at time step $t+1$ is summed over all j 's neighbours as

$$EH_j^{t+1} = EH_j^t + \sum_{e_{ij} \in E} \Delta EH_{i \rightarrow j}^{t+1}, \quad (12)$$

where $e_{ij} = 1$ if there is a link between node i and node j , and $e_{ij} = 0$ otherwise.

- (v) Finally, the global Earth's health, at time step $t+1$, is summarized over all the nodes:

$$EH^{t+1} = \sum_{i=1}^{|V|} EH_i^{t+1}. \quad (13)$$

3. Results

3.1. Performance Comparison. To test the performance of our spreading model, we set $\alpha = 4$ and $\beta = 0.65$ to predict the global Earth's health in the computer simulation. In addition, we also use *Gaussian Fitting* to compare with our model. Figure 5 shows the comparison results. It can be seen that our model can better fit the real data, comparing with *Gaussian Fitting*. In addition, the proposed EH index shows that the Earth's health is getting worse from 2008, which gives us the warning that we should put much more attention to our environment. Correspondingly, results from Section 2.4.4 suggest for us a possible solution that returning more living lands might be the most effective way to solve this dilemma.

3.2. Sensitivity Analysis. We then perform the sensitivity analysis to test the robustness of our model. We randomly delete P fraction of the links and see whether our model is reliable or not. Figure 6 reports different prediction results of model for various values of P . It can be seen that the prediction result of model is quite robust that even a large fraction of links, 60% for instance, is removed. Therefore, it can be concluded that our model is reliable for predicting the Earth's health.

3.3. Identifying the Influential Node in Ecological Networks. Our Earth's health index tries to diagnose and predict the global Earth's health status. In addition, in order to find which node (saying nation or ocean) plays the most important role in affecting the Earth's health, we additionally perform analysis to rank the node importance. We integrate the k -core value, $S_K(i)$ (see Section 2.1), and the Earth's health index EH_i to evaluate the node i 's importance, I_i , which is consequently defined as

$$I_i = EH_i * S_K(i). \quad (14)$$

Figure 7 illustrates the node importance in affecting the Earth's health by I_i versus the corresponding rank, where some typical nodes are marked. The oceans (with the highest rank) indeed are the key nodes in affecting the Earth's health; USA, Russia, and China are also important nations in influencing the Earth's health. Some small nations, such as Madagascar (MDG) and Iraq (IRQ), play less important roles for the Earth's health.

FIGURE 4: $err(X^j)$ values of five representative factors versus time.

FIGURE 5: Earth health (EH) tendency comparisons for (1) empirical data, (2) our model, and (3) for *Gaussian Fitting* (GF).

4. Conclusion and Discussion

In this paper, we collect various data and construct a 145-nation (including 126 nations, 19 oceans/seas, and 403 edges) world ecological network, with each node representing a nation or an ocean and each edge representing geographical neighboring relationship of the corresponding two nodes. Firstly, we analyze both the topological properties and time-dependant features of nodes. Secondly, we propose an Earth's health index based on *Shannon Entropy*. Thirdly, we identify the importance of elements by a machine learning approach (*Random Forest*). Fourthly, we design a spreading model to predict the Earth's health and perform sensitive analysis to test its robustness. Finally, we integrate the topological

FIGURE 6: Comparisons with the original network for various fractions of links are removed from the network for $P = 0.2, 0.4, 0.6$, respectively.

FIGURE 7: Node importance versus corresponding rank.

property (k -core index) and the health index to identify the influential nations in the observed ecological network.

The model results indicated that the oceans (with the highest rank) indeed are the key nodes in affecting the Earth's health. The Big countries, for example, USA, Russia, and China, are also important nations in influencing the Earth's health. Correspondingly, it suggests for us a possible solution that returning more living lands might be the most effective way to solve this dilemma. The combination of topological properties and local factors leads to good performance in both predicting the good and bad trends of the Earth's health. The model can be easily extended by considering more factors. However, our model needs empirical support from more sufficient data. Also, the incremental mechanism may hinder long-term prediction.

Acknowledgments

This work was partially supported by the National Natural Science Foundation of China (Grant nos. 11105024, 1147015, 11305043, and 11205040), the EU FP7 Grant 611272 (project GROWTHCOM), and the start-up foundation and Pandeng project of Hangzhou Normal University.

References

- [1] L. V. Berkner and L. C. Marshall, "The history of oxygenic concentration in the Earth's atmosphere," *Discussions of the Faraday Society*, vol. 37, pp. 122–141, 1964.
- [2] D. B. Anthony, A. H. Elizabeth, B. Jordi et al., "Approaching a state shift in Earth's biosphere," *Nature*, vol. 486, pp. 52–58, 2012.
- [3] R. V. Solé and J. M. Montoya, "Complexity and fragility in ecological networks," *Proceedings of the Royal Society B*, vol. 268, no. 1480, pp. 2039–2045, 2001.
- [4] J. M. Montoya, S. L. Pimm, and R. V. Solé, "Ecological networks and their fragility," *Nature*, vol. 442, no. 7100, pp. 259–264, 2006.
- [5] R. E. Ulanowicz, "Quantitative methods for ecological network analysis," *Computational Biology and Chemistry*, vol. 28, no. 5–6, pp. 321–339, 2004.
- [6] B. D. Fath, U. M. Scharler, R. E. Ulanowicz, and B. Hannon, "Ecological network analysis: network construction," *Ecological Modelling*, vol. 208, no. 1, pp. 49–55, 2007.
- [7] M. Keller, D. S. Schimel, W. W. Hargrove, and F. M. Hoffman, "A continental strategy for the national ecological observatory network," *Frontiers in Ecology and the Environment*, vol. 6, no. 5, pp. 282–284, 2008.
- [8] C. Kazanci and Q. Ma, "Extending ecological network analysis measures to dynamic ecosystem models," *Ecological Modelling*, vol. 242, pp. 180–188, 2012.
- [9] M. Kitsak, L. K. Gallos, S. Havlin et al., "Identification of influential spreaders in complex networks," *Nature Physics*, vol. 6, no. 11, pp. 888–893, 2010.
- [10] J. Hill, *Influence: a theory of the structural organization of systems [Doctoral thesis]*, University of Georgia, Athens, Ga, USA, 1981.
- [11] J. T. Finn, "Measures of ecosystem structure and function derived from analysis of flows," *Journal of Theoretical Biology*, vol. 56, no. 2, pp. 363–380, 1976.
- [12] D. J. Watts and S. H. Strogatz, "Collective dynamics of 'small-world' networks," *Nature*, vol. 393, no. 6684, pp. 440–442, 1998.
- [13] A.-L. Barabási and R. Albert, "Emergence of scaling in random networks," *Science*, vol. 286, no. 5439, pp. 509–512, 1999.
- [14] R. Albert and A.-L. Barabási, "Statistical mechanics of complex networks," *Reviews of Modern Physics*, vol. 74, no. 1, pp. 47–97, 2002.
- [15] S. N. Dorogovtsev and J. F. F. Mendes, "Evolution of networks," *Advances in Physics*, vol. 51, no. 4, pp. 1079–1187, 2002.
- [16] M. E. J. Newman, "The structure and function of complex networks," *SIAM Review*, vol. 45, no. 2, pp. 167–256, 2003.
- [17] S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, and D.-U. Hwang, "Complex networks: structure and dynamics," *Physics Reports*, vol. 424, no. 4–5, pp. 175–308, 2006.
- [18] L. F. Costa, F. A. Rodrigues, G. Travieso, and P. R. U. Boas, "Characterization of complex networks: a survey of measurements," *Advances in Physics*, vol. 56, no. 1, pp. 167–242, 2007.
- [19] J. Kleinberg, "The convergence of social and technological networks," *Communications of the ACM*, vol. 51, no. 11, pp. 66–72, 2008.

- [20] Z. K. Zhang and C. Liu, "A hypergraph model of social tagging networks," *Journal of Statistical Mechanics: Theory and Experiment*, vol. 2010, Article ID P10005, 2010.
- [21] L. L. Lü and T. Zhou, "Link prediction in complex networks: a survey," *Physica A*, vol. 390, no. 6, pp. 1150–1170, 2011.
- [22] L. C. Freeman, "Centrality in social networks conceptual clarification," *Social Networks*, vol. 1, no. 3, pp. 215–239, 1979.
- [23] S. N. Dorogovtsev, A. V. Goltsev, and J. F. F. Mendes, "K-core organization of complex networks," *Physical Review Letters*, vol. 96, no. 4, Article ID 040601, 2006.
- [24] C. E. Shannon, "A mathematical theory of communication," *The Bell System Technical Journal*, vol. 27, pp. 379–423, 1948.
- [25] N. Eagle, M. Macy, and R. Claxton, "Network diversity and economic development," *Science*, vol. 328, no. 5981, pp. 1029–1031, 2010.
- [26] Z. Q. Jiang, W. J. Xie, M. X. Li, B. Podobnik, W. X. Zhou, and H. E. Stanley, "Calling patterns in human communication dynamics," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 110, no. 5, pp. 1600–1605, 2013.
- [27] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.
- [28] C. Liu and Z.-K. Zhang, "Information spreading on dynamic social networks," *Communications in Nonlinear Science and Numerical Simulation*, vol. 19, no. 4, pp. 896–904, 2014.