A LETTERS JOURNAL EXPLORING
THE FRONTIERS OF PHYSICS

March 2019

EPL, **125** (2019) 68002
www.epljournal.org
doi: 10.1209/0295-5075/125/68002

Regional economic status inference from information flow and talent mobility

JUN WANG[1], JIAN GAO[1,2(a)], JIN-HU LIU[1], DAN YANG[1] and TAO ZHOU[1,2]

[1] *Big Data Research Center, University of Electronic Science and Technology of China - Chengdu 611731, PRC*
[2] *Institution of New Economic Development - Chengdu 610049, PRC*

received 15 January 2019; accepted in final form 14 March 2019
published online 26 April 2019

PACS 89.65.-s – Social and economic systems
PACS 89.75.Fb – Structures and organization in complex systems
PACS 87.23.Ge – Dynamics of social systems

Abstract – Novel data has been leveraged to estimate the socioeconomic status in a timely manner, however, direct comparison on the use of social relations and talent movements remains rare. In this letter, we estimate the regional economic status based on the structural features of two networks. One is the online information flow network built on the following relations on social media, and the other is the offline talent mobility network built on the anonymized *résumé* data of job seekers with higher education. We find that while the structural features of both networks are relevant to the economic status, the talent mobility network in a relatively smaller size exhibits a stronger predictive power for the gross domestic product (GDP). In particular, a composite index of structural features can explain up to about 84% of the variance in GDP. The result suggests that future socioeconomic studies should pay more attention to the cost-effective talent mobility data.

Introduction. – Timely estimation of the social and economic status has important implications for addressing many development-related issues [1–4], such as developing policies to reduce poverty [5], forecasting the unemployment rate [6,7], and optimizing strategies for economic diversification [8,9]. Traditional socioeconomic status inference, however, usually follows a long-time delay due to the large consumption of resources in data collection. Thanks to the technological development, novel data sources are now increasingly available for estimating the socioeconomic status [10,11]. For example, Elvidge *et al.* [12] produced a global poverty map based on the brightness of night-time lights. Gao and Zhou [13] quantified the regional economic complexity by analyzing firm data. Dong *et al.* [14] measured the economic activity through mining mobile phone records. Liu *et al.* [15] inferred the city-level economic status from online activities. Blumenstock *et al.* [16] predicted district-level wealth distribution based on mobile phone usages. Sobolevsky *et al.* [17] estimated the individual socioeconomic status by analyzing bank card transactions. More related works are summarized in recent reviews [2,18].

Among these works, two streams of literature are of particular interest. One stream focuses on relations between social network structure and economic status [4,19]. For example, Eagle *et al.* [20] uncovered a strong correlation between social network diversity and socioeconomic indicators, Mao *et al.* [21] found that the ratio of in-going and out-going calls can predict a region's income level, and Holzbauer *et al.* [22] showed that cross-state long ties on social media are strongly correlated with GDP in the US. Recently, Jahani *et al.* [23] uncovered a strong correlation between ego-network structural diversity and individual income, and Luo *et al.* [24] found that individuals' influence in social network is predictive to their economic status. The other stream links the human mobility pattern to socioeconomic status and outcomes [25,26]. Individuals with different socioeconomic status have distinct mobility patterns [27,28], and the movement of talents is critical to the economic development [29]. For example, Frias-Martinez *et al.* [30] showed the predictive power of mobility patterns to socioeconomic status, Pappalardo *et al.* [31] found that movement diversity can well predict socioeconomic indicators, and Florez *et al.* [32] demonstrated that a group's income increases with the diversity of commuting trips.

(a) E-mail: gaojian08@hotmail.com

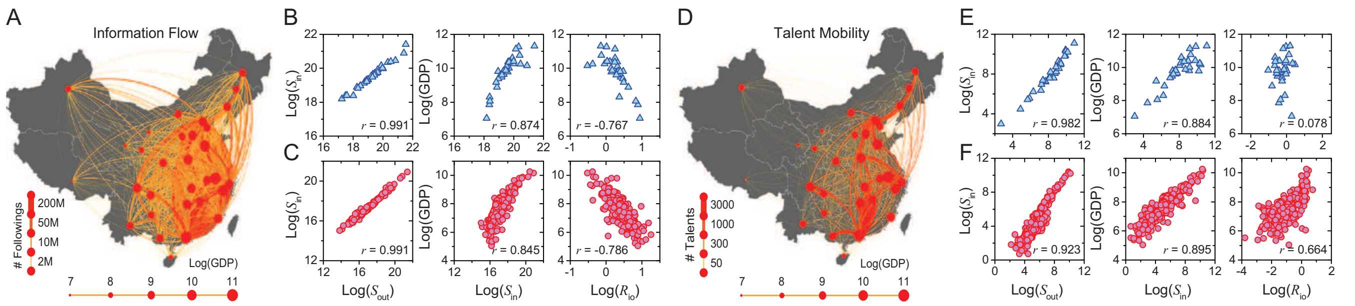

Fig. 1: Relations between network structural features and economic status. Nodes represent provinces with size showing their GDP in natural logarithmic form. (A) presents the online information flow (OIF) network with link weight being the number of followings. (B) and (C) present the relations regarding OIF at the province and city level, respectively. (D) presents the offline talent mobility (OTM) network with link weight being the number of moved talents. (E) and (F) present the relations regarding OTM at the province and city levels, respectively. The Pearson correlation coefficient r is also presented.

Most of the previous works focus on either social network structures or human behavioral patterns [33]. Yet, the direct comparison between the predictive power of online social network structure and offline human mobility pattern to the regional and individual socioeconomic status remains insufficient. One challenge that hinders studies towards this direction is the lack of large-scale and high-resolution online information and offline mobility data. Recently, the increasing availability of large-scale social and economic data with high spatial and temporal resolutions, such as mobile phone records [14], behavioral data [34], web-based ratings [35], public profiles [36], has made it possible to estimate the socioeconomic status in a timely manner and with a relatively low cost [2,9]. This provides us with a chance to compare the capability of information flow and talent mobility on speculating the economic status.

In this letter, we infer the regional economic status from the following relations on social media and the talent movements recorded by anonymized *résumé* data. We first build two directed and weighted networks, named online information flow network and offline talent mobility network. Then, we calculate several network structural features and link them to GDP. The results show that some features exhibit strong correlations with GDP such as the loops and outgoing spatial diversity of the information flow network and the out-strength and ingoing topological diversity of the information flow network. Overall, the talent mobility network features perform better in predicting the economic status. After performing the regression analysis for robustness checks, we further construct a composite index of both network structural features, which can explain up to about 84% of the variance in GDP.

Data and methods. – In this section, we first introduce two large-scale online-crawled datasets for network construction, then present some measures to quantify network structures, and lastly introduce the methods applied for correlation and regression analyses.

Data description. The online information flow (OIF) network is built based on the public profiles and following relations among about 433 million users of the China's social network Weibo, which provides similar functions to Twitter. Specifically, from profiles we extract users' locations covering 336 prefecture-level cities aggregated into 31 provinces (see ref. [15] for details). Then, based on the users' following relations we build the OIF network G^I among regions (cities or provinces dependent on the resolution) and represent it by a weighted adjacent matrix A^I, whose element $a^I_{i,j}$ is the volume of information from region i to j, which is roughly estimated by the number of followings from region j to i. As users within the same region can follow each other, G^I contains loops, *i.e.*, $a^I_{i,i} \neq 0$ in A^I. Figure 1(A) presents the visualization of the provincial-level OIF network and table 1 summarizes basic statistics.

Table 1: Statistics of the online information flow (OIF) and the offline talent mobility (OTM) networks. The average link weight is calculated by $\langle a \rangle = \sum_{i,j} a_{i,j} / \sum_{i,j} \delta_{i,j}$, where $\delta_{i,j} = 1$ if there is a link from node i to j, and $\delta_{i,j} = 0$ if otherwise.

Network	Resolution	# Regions	# Links	$\langle a \rangle$
OIF	Province	31	961	1.277×10^7
	City	336	112,896	1.087×10^5
OTM	Province	31	818	347.7
	City	287	9,746	29.18

The offline talent mobility (OTM) network is built based on the self-reported *résumé* data of about 142 thousand anonymized Chinese job seekers with higher education (see ref. [36] for details). Specifically, we roughly estimate the flow of talents among regions based on the movements of job seekers from the birth city to the living city in career development and from living city to expected city in job hunting. The *résumé* data covers 287 prefecture-level cities aggregated into 31 provinces. Notice that, some cities are isolated due to sparsity, and only cities remaining in the giant connected network are counted. The directed and weighted OTM network G^T can also be represented by a weighted adjacent matrix A^T, whose element $a^T_{i,j}$ is the number of talents from region i to j. Similarly, G^T contains loops. Figure 1(D) visualizes the provincial-level OTM network and table 1 summarizes basic statistics.

Some macroeconomic data at the province and city levels are collected respectively from the official books entitled "China Statistical Yearbook (2017)" and "China City Statistical Yearbook (2017)" released by the National Bureau of Statistics of China. Due to the time-consuming statistics, these books provide data with one year-lag, namely, for the year 2016. We have successfully collected GDP of 31 provinces and 290 prefecture-level cities while we failed for the remaining 46 cities due to the missing data. The unit of GDP data is 10000 RMB (about 1500 USD).

Structural features. Considering a network G with a weighted adjacency matrix A, we first calculate three direct structural features, namely, S_{out}, S_{in}, and S_{loop} [37]. Specifically, for a region i, $S_{out}(i) = \sum_j a_{i,j}$ sums the weights of outgoing links, $S_{in}(i) = \sum_j a_{j,i}$ sums the weights of ingoing links, and $S_{loop}(i) = a_{i,i}$ is the weight of the self-loop link. Then, we calculate three relative structural features, namely, R_{io}, R_{lo} and R_{li}. Specifically, $R_{io}(i) = S_{in}(i)/S_{out}(i)$ measures the rates of local information/talent retention. $R_{lo}(i) = S_{loop}(i)/S_{out}(i)$ measures information/talent drain, where $R_{lo} = 0$ and $R_{lo} = 1$ mean all information/talents are drained and kept, respectively. $R_{li}(i) = S_{loop}(i)/S_{in}(i)$ measures information/talent gain, where $R_{li} = 0$ means that new information/talents are gained and $R_{li} = 1$ means that previous information/talents are kept.

Moreover, we quantify diversity by calculating four network structural features: two topological diversity measures (H_{out} and H_{in}) and two spatial diversity measures (D_{out} and D_{in}) [20]. Specifically, the ingoing and outgoing topological diversity of a region is defined by the Shannon entropy associated with the information/talent flow into and out of the region, respectively. Formally, the outgoing topological diversity $H_{out}(i)$ for region i is given by

$$H_{out}(i) = -\sum_{j\neq i} p_{i,j}\log(p_{i,j}), \qquad (1)$$

where $p_{i,j} = a_{i,j}/\sum_j a_{i,j}$. The outgoing spatial diversity $D_{out}(i)$ for region i is calculated by normalizing $H_{out}(i)$ using the number of involved regions. Mathematically,

$$D_{out}(i) = \frac{H_{out}(i)}{\log(k_{out}(i))}, \qquad (2)$$

where $k_{out}(i)$ is the out-degree of region i. Analogously, the ingoing topological diversity $H_{in}(i)$ for region i is defined in the similar manner, by

$$H_{in}(i) = -\sum_{j\neq i} p_{j,i}\log(p_{j,i}), \qquad (3)$$

where $p_{j,i} = a_{j,i}/\sum_j a_{j,i}$. The ingoing spatial diversity $D_{in}(i)$ for region i is calculated by normalizing $H_{in}(i)$ using the number of involved regions, as

$$D_{in}(i) = \frac{H_{in}(i)}{\log(k_{in}(i))}, \qquad (4)$$

where $k_{in}(i)$ is the in-degree of region i.

Analytical methods. To exploit the relations between structural features and GDP, we perform both correlation analysis and regression analysis. The Pearson correlation coefficient r is used to quantify the linear correlation between two variables. The value r is in the range $[-1, 1]$, from negative to positive correlation. The ordinary least squares (OLS) model is employed to regress GDP against structural features. The estimated equation is given by

$$\begin{aligned}\log(\mathrm{GDP}) = {} & \beta_0 + \beta_1 S_{out} + \beta_2 S_{in} + \beta_3 R_{io} \\ & + \beta_4 S_{loop} + \beta_5 R_{lo} + \beta_6 R_{li} + \beta_7 H_{out} \\ & + \beta_8 H_{in} + \beta_9 D_{out} + \beta_{10} D_{in} + \varepsilon, \end{aligned} \qquad (5)$$

where the structural variables are in the logarithmic form expect for the diversity measures, $\{\beta_0, \beta_1, \ldots, \beta_{10}\}$ are regression coefficients of variables, and ε is the error term.

Results. – In this section, we first analyze correlations between simple structural features and GDP, then summarize correlations between diversity-related features and GDP, and finally perform some robustness checks using regression models, based on which a composite index is further constructed to explore the prediction accuracy.

Correlation between simple features and GDP. The visualizations of province-level online information flow (OIF) and offline talent mobility (OTM) networks are presented in figs. 1(A) and (D), in which the direct link weights are the numbers of followings and talents from origin to target provinces, respectively. For OIF, figs. 1(B) and (C) (left) present the relations between S_{out} and S_{in} at the province and city levels, respectively. We find that S_{out} and S_{in} are perfectly correlated with each other, as suggested by $r \approx 0.99$ at both resolutions. In contrast, as shown in figs. 1(E) and (F) (left), the correlations between S_{out} and S_{in} for OTM are relatively weaker, suggesting the unbalance of talent flows into and out of regions.

The volume of information and talent flows can be relevant to a region's economic status. For OIF, figs. 1(B) and (C) (middle) present the relations between S_{in} and GDP at the province and city levels, respectively. We notice that S_{in} exhibits a high correlation ($r \approx 0.86$) with GDP. Figures 1(E) and (F) (middle) present a similar trend for OTM, while the correlations ($r \approx 0.89$) are stronger at both resolutions. The ratio of ingoing and outgoing flows can also be linked to economic status. For OIF, figs. 1(B) and (C) (right) present the relations between R_{io} and GDP, where we find negative correlations ($r \approx -0.78$) at both resolutions. This suggests that developed regions spread information better. As presented by figs. 1(E) and (F) (right) for OTM, however, we find a positive correlation ($r \approx 0.66$) only at the province level.

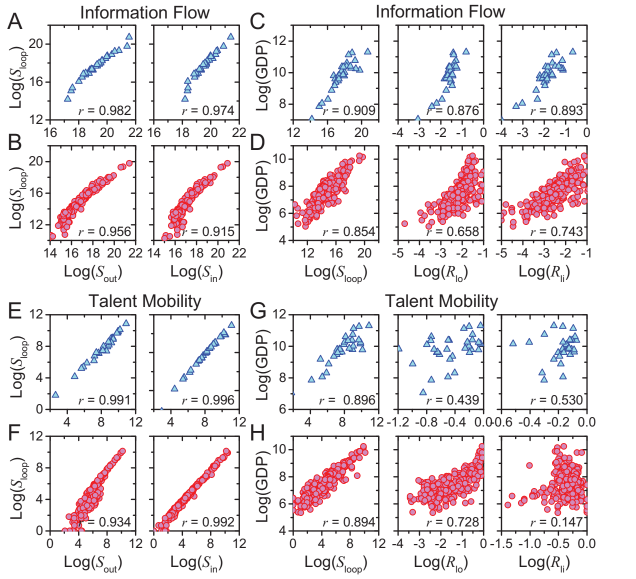

Fig. 2: Relations between network loops and economic status. (A) and (B) present relations between loops and strengths of the online information flow (OIF) network at the province and city levels, respectively. Analogously, (E) and (F) present relations for the offline talent mobility (OTM) network. (C) and (D) present relations between loop-related measures of OIF and GDP at the province and city levels, respectively. Analogously, (G) and (H) present relations for OTM. The Pearson correlation coefficient r is also presented.

These results suggest that attractiveness for talents in fine-grained regions reflects the economic status better. This observation may be originated from the inequality of regional economic development. For instance, China faces seriously unbalanced regional economic development, where more developed cities usually have talent gain, while less developed cities may have talent drain. This unbalanced talent mobility and economic development at the city level may result in the positive correlations. However, such correlation can be diminished at the aggregated province level as a province can have multiple cities with different social and economic status, and talents can move among cities located in the same province.

The strength of loops (S_{loop}) in the OIF and OTM networks suggests the retention of local information and talents, respectively. For OIF, figs. 2(A) and (B) present how S_{loop} is related to S_{out} (left) and S_{in} (right) at province and city levels, respectively. Similarly, figs. 2(E) and (F) present the relations for OTM. Overall, we find that loops are perfectly correlated ($r \approx 0.99$) with strengths. Further, we explore how information and talent retentions are linked to economic status by calculating correlations between GDP and three loop-related features, namely, S_{loop}, R_{lo}, and R_{li}. For OIF, we find from figs. 2(C) and (D) that GDP is positively correlated with all the three features, and S_{loop} exhibits the strongest correlation ($r \approx 0.91$) at the province level. Similar results hold for OTM as shown in figs. 2(G) and (H), and S_{loop} has a high correlation ($r \approx 0.90$) with GDP at both resolutions. These results suggest the predictive power of local information and talent retentions for regional economic status.

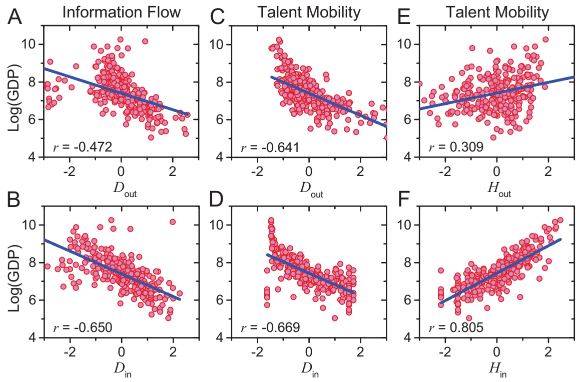

Fig. 3: Relations between diversity-related features and economic status. (A) and (B) present how GDP is related to D_{out} and D_{in} of the online information flow (OIF) network, respectively. Analogously, (C) and (D) present relations for the offline talent mobility (OTM) network. (E) and (F) present how GDP is related to H_{out} and H_{in} of OTM, respectively. Lines present the linear fits, and r is the Pearson correlation coefficient.

Correlation between diversity features and GDP. We explore relations between GDP and two diversity-related features, namely, spatial diversity (D) and topological diversity (H). For OIF, figs. 3(A) and (B) present how GDP is related to the outgoing (D_{out}) and ingoing (D_{in}) spatial diversities, respectively. We observe strong negative correlations for both cases, specifically, $r = -0.472$ for D_{out} and $r = -0.650$ for D_{in}. As shown in figs. 3(C) and (D), while similar observations hold for OTM, the correlations ($r \approx -0.66$) are stronger. In particular, we notice that D_{in} has a stronger correlation with GDP for both networks. Previous study based on the UK communications showed that social network spatial diversity is positively correlated with community-level development [20], however, our results based on both the OIT and OTM networks in China suggest spatial diversities as negative predictors of regional economic status.

The topological diversity is equal to the spatial diversity for OIF as it is fully connected. Thereby, only for OTM we present how GDP is related to topological diversities H_{out} and H_{in} in figs. 3(E) and (F), respectively. We find that the correlation $r = 0.805$ between H_{in} and GDP is significantly larger than the correlation $r = 0.309$ between H_{out} and GDP, showing that H_{in} is a more relevant feature to economic status. In summary, we find that D and H of OIF and D of OTM are negative predictors of GDP, while H of OTM is positively correlated with GDP.

Regression analysis and composite index. The Pearson correlations between structural features and economic development (GDP) are summarized in fig. 4. As shown in fig. 4(A) for OIF, simple structural features except R_{io} have strongly positive correlations with GDP, while

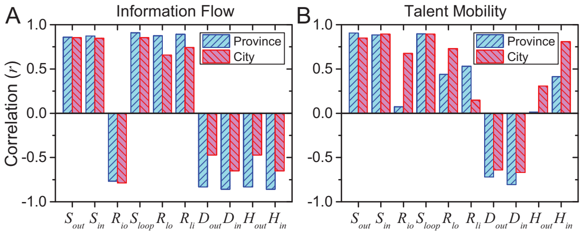

Fig. 4: Pearson correlations between structural features and GDP. (A) and (B) present correlations between GDP and structural features of the online information flow network and the offline talent mobility network, respectively.

diversity-related features exhibit strongly negative correlations. Moreover, the network structural features are more relevant to economic status at the province level than at the city level. In particular, the most relevant features are loops and diversities at the province level as well as strengths and loops at the city level. As presented in fig. 4(B) for OTM, the most relevant features are S_{out}, S_{in} and S_{loop} at both the province and city levels.

We further perform some robustness checks by employing the ordinary least squares (OLS) model to regress GDP against structural features at the city level. Table 2 summaries the regression results. As shown in columns (1) and (2), the model including the OIF and the OTM network structural features can explain up to 76.2% and 80.2% of the variance in GDP, respectively. In particular, we notice that D_{out} and D_{in} of OIF are respectively significantly positive and negative predictors of GDP, while only H_{out} of OTM is a positive predictor of GDP. Including both network structural features in column (3), where column (3-1) and column (3-2) correspond respectively to OIF and OTM, the variance in GDP can be explained by up to 83.2%. We additionally find that loops are the features that can best explain the variance in GDP with the adjusted $R^2 = 74.3\%$ for OIF and $R^2 = 79.8\%$ for OTM. These results confirm that the OTM network structural features are more predictive to regional economic development.

Based on the regression analysis, we construct a composite index of network structural features for the best prediction of regional economic status. Specifically, the composite index is calculated through weighting structural features by their regression coefficients. Formally, the composite index $CI(i)$ for region i is given by

$$CI(i) = \sum_{j=1}^{10} \beta_j^I M_{j,i}^I + \sum_{j=1}^{10} \beta_j^T M_{j,i}^T. \tag{6}$$

$M = \{\vec{S}_{out}, \vec{S}_{in}, \vec{R}_{io}, \vec{S}_{loop}, \vec{R}_{lo}, \vec{R}_{li}, \vec{D}_{out}, \vec{D}_{in}, \vec{H}_{out}, \vec{H}_{in}\}$ represents the ten vectors of network structural features, and $\vec{\beta} = \{\beta_1, \ldots, \beta_{10}\}$ is the vector of corresponding regression coefficients as shown in table 2. Specifically, M^I and β^I are for OIF, and M^T and β^T are for OTM. All network structural features are standardized by the z-score [38] before constructing the composite index.

Table 2: The predictive power of network structural features for economic development. The OLS model is used to regress log(GDP) against network structural features. All variables are standardized by the z-score before being included. The symbol "$--$" marks the omitted variables due to high collinearity that are automatically identified by the regression model. Significant level: $^{*}p < 0.1$, $^{**}p < 0.05$, and $^{***}p < 0.01$.

	OLS Model			
Variables	(1)	(2)	(3-1)	(3-2)
S_{out}	0.823^{***}	0.587^{***}	$--$	0.266^{***}
S_{in}	$--$	$--$	0.363^{***}	$--$
R_{io}	0.217^{**}	0.300^{***}	0.051	0.128^{**}
S_{loop}	$--$	$--$	$--$	$--$
R_{lo}	0.216^{***}	$--$	0.087^{*}	$--$
R_{li}	$--$	0.041	$--$	0.059^{**}
D_{out}	0.208^{***}	-0.010	0.203^{***}	-0.096
D_{in}	-0.291^{***}	0.013	-0.192^{***}	0.010
H_{out}	$--$	0.067^{**}	$--$	0.016
H_{in}	$--$	0.103	$--$	0.120^{*}
Obs.	290	280	280	
Adj. R^2	0.762	0.802	0.832	

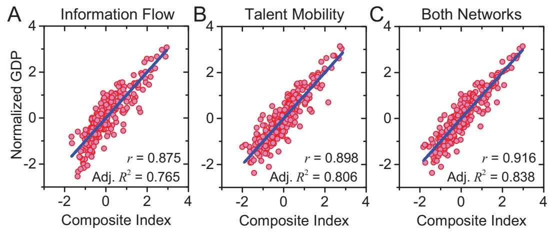

Fig. 5: Predictive power of the composite index for economic development. (A) and (B) present how GDP is related to the composite index of online information flow and offline talent mobility network structural features. (C) presents the relation between GDP and the composite index of both networks. The Pearson correlation coefficient r is presented. The line is the linear fit, and the adjusted R^2 shows its explanatory power.

The correlations between the composite index and normalized GDP at the city level are presented in figs. 5(A) and (B) for OIF and OTM, respectively. For both networks, we find that GDP is strongly and positively correlated with the composite index. In particular, the composite index of OTM exhibits a slightly larger correlation ($r = 0.898$) with GDP than the one of OIF ($r = 0.875$). The composite index of OIF and OTM can explain 76.5% and 80.6% of the variance in GDP, respectively. These observations suggest strong predictive powers of information and talent flows for regional economic development. Further, we construct a composite index using the structural features of both networks. As shown in fig. 5(C), the composite index has the largest correlation ($r = 0.916$) with GDP, and it can explain up to 83.8% the variance in GDP. The result shows that combining network features of information flow and talent mobility can enhance the performance of the economic status inference.

Conclusion and discussions. – In summary, we have explored the inference of the regional economic status from the online information flow network and the offline talent mobility network. The former was built on the following relations among about 433 million social media users, and the latter was built on the self-reported *résumé* data of over 142 thousand job seekers with higher education. After performing the correlation analysis, we found that strengths of both networks have strongly positive correlations with GDP, and the loop-related network features are the most relevant. Moreover, we uncovered the negative correlations between GDP and the spatial diversities for both networks, while the topological diversities of the talent mobility network are positively correlated with GDP. Interestingly, we found that the talent mobility network features exhibit a stronger predictive power for GDP although it covers only about 1/3000 people in comparison with the information flow network. This suggests a more cost-effective way to infer the economic status by leveraging some relative small-scale offline talent mobility data.

The correlations between GDP and the information flow network structural features diminish at the fine-grained resolution. In particular, we observed negative correlations between spatial diversities and GDP, which is different from the previous finding [20]. Whether this inconsistency is originated from the inequality and complexity of China's regional development [13] remains an open issue. Through the regression analysis, we found that the significant predictors of GDP are out-strength, ratios of loops and spatial diversities of the information flow network as well as the out-strength, loops and outgoing topological diversities of the talent mobility network. Based on the regression results, we further constructed a composite index of both network structural features that can explain up to about 84% of the variance in GDP. The result suggests a way of improving the economic status inference through combining different network information.

The presented results should be interpreted in the light of some limitations on the data and analytical methods, which ask for further explorations. The estimation of information flow was solely based on social media, where taking into account other information exchange channels such as online chats [39] and mobile communications [40,41] would help. The *résumé* data covers a relatively small sample, where adding other large-scale data from human resource services [36], academic publishers [42] and formal talent markets [43] will be an improvement. Recent available large-scale and high spatio-temporal data would advance studies on comparing the predictive power of different data sources on inferring the socioeconomic status. Moreover, a limited number of structural features were considered, where many network ranking indicators [44] can also be considered. In addition, it would be interesting to apply some variant models to predict and validate regional and temporal change of GDP based on time-windowed past GDP and network data, and we leave this for future work when data are available. Keeping these aforementioned limitations in mind, we hope our work will spark further studies on the economic status inference from the aspects of both information flow and talent mobility.

* * *

The authors acknowledge Hao Chen, Jing-Yi Liao, Zhong-Zheng Peng, Zhi-Hai Rong and Jun-Ming Shao for helpful discussions and Rui-Tong Wang for processing the raw data files. This work was partially supported by the National Natural Science Foundation of China (Grant Nos. 61433014, 61603074, 61673086, and 61703074).

REFERENCES

[1] Schweitzer F., Fagiolo G., Sornette D., Vega-Redondo F., Vespignani A. and White D. R., *Science*, **325** (2009) 422.
[2] Einav L. and Levin J., *Science*, **346** (2014) 1243089.
[3] Perc M., *J. R. Soc. Interface*, **11** (2014) 20140378.
[4] Zhang X., Shao S., Stanley H. E. and Havlin S., *EPL*, **108** (2014) 58001.
[5] Birdsall N. and Londoño J. L., *Am. Econ. Rev.*, **87** (1997) 32.
[6] Llorente A., Garcia-Herranz M., Cebrian M. and Moro E., *PLoS ONE*, **10** (2015) e0128692.
[7] Yuan J., Zhang Q.-M., Gao J., Zhang L., Wan X.-S., Yu X.-J. and Zhou T., *Physica A*, **444** (2016) 442.
[8] Alshamsi A., Pinheiro F. L. and Hidalgo C. A., *Nat. Commun.*, **9** (2018) 1328.
[9] Gao J., in *2017 14th International Computer Conference on Wavelet Active Media Technology and Information Processing* (IEEE Press) 2017, pp. 337–341.
[10] Blumenstock J. and Eagle N., in *Proceedings of the 4th ACM/IEEE International Conference on Information and Communication Technologies and Development* (ACM Press) 2010, p. 6.
[11] Jean N., Burke M., Xie M., Davis W. M., Lobell D. B. and Ermon S., *Science*, **353** (2016) 790.
[12] Elvidge C. D., Sutton P. C., Ghosh T., Tuttle B. T., Baugh K. E., Bhaduri B. and Bright E., *Comput. Geosci.*, **35** (2009) 1652.
[13] Gao J. and Zhou T., *Physica A*, **492** (2018) 1591.
[14] Dong L., Chen S., Cheng Y., Wu Z., Li C. and Wu H., *EPJ Data Sci.*, **6** (2017) 29.
[15] Liu J.-H., Wang J., Shao J. and Zhou T., *Physica A*, **457** (2016) 581.
[16] Blumenstock J., Cadamuro G. and On R., *Science*, **350** (2015) 1073.
[17] Sobolevsky S., Massaro E., Bojic I., Arias J. M. and Ratti C., in *2017 IEEE International Conference on Big Data* (IEEE Press) 2017, pp. 1313–1318.
[18] Gao J. and Zhou T., *Journal of the University of Electronic Science and Technology of China*, **45** (2016) 625.
[19] Granovetter M., *J. Econ. Perspect.*, **19** (2005) 33.

[20] Eagle N., Macy M. and Claxton R., *Science*, **328** (2010) 1029.
[21] Mao H., Shuai X., Ahn Y.-Y. and Bollen J., *EPJ Data Sci.*, **4** (2015) 15.
[22] Holzbauer B. O., Szymanski B. K., Nguyen T. and Pentland A. S., in *12th International Conference and School on Network Science* (Springer) 2016, pp. 178–185.
[23] Jahani E., Saint-Jacques G., Sundsøy P., Bjelland J., Moro E. and Pentland A. S., in *Proceedings of the 9th International Conference on Social Informatics* (Springer) 2017, pp. 41–50.
[24] Luo S., Morone F., Sarraute C., Travizano M. and Makse H. A., *Nat. Commun.*, **8** (2017) 15227.
[25] Carra G., Mulalic I., Fosgerau M. and Barthelemy M., *J. R. Soc. Interface*, **13** (2016) 20160306.
[26] Fan C., Huang J., Rong Z. and Zhou T., *EPL*, **124** (2018) 48001.
[27] Yan X.-Y., Han X.-P., Wang B.-H. and Zhou T., *Sci. Rep.*, **3** (2012) 2678.
[28] Lotero L., Cardillo A., Hurtado R. and Gómez-Gardeñes J., *Interconnected Networks* (Springer) 2016, pp. 149–164.
[29] Kerr S. P., Kerr W., Özden Ç. and Parsons C., *J. Econ. Perspect.*, **30** (2016) 83.
[30] Frias-Martinez V., Soguero-Ruiz C., Frias-Martinez E. and Josephidou M., in *Proceedings of the 3rd ACM Symposium on Computing for Development* (ACM Press) 2013, p. 15.
[31] Pappalardo L., Vanhoof M., Gabrielli L., Smoreda Z., Pedreschi D. and Giannotti F., *Int. J. Data Sci. Analyt.*, **2** (2016) 75.
[32] Florez M., Jiang S., Li R., Mojica C. H., Transmilenio S. A., Rios R. A. and González M. C., in *Proceedings of the Transportation Research Board 96th Annual Meeting* (TRB) 2018, pp. 1–19.
[33] Zhao Z.-D., Cai S.-M. and Lu Y., *Chaos*, **25** (2015) 063106.
[34] Cao Y. *et al.*, *J. R. Soc. Interface*, **15** (2018) 20180210.
[35] Chen L.-J. and Gao J., *Physica A*, **506** (2018) 679.
[36] Yang X., Gao J., Liu J.-H. and Zhou T., *Physica A*, **501** (2018) 86.
[37] Barabási A.-L. and Márton P., *Network Science* (Cambridge University Press, New York) 2016.
[38] Kreyszig E., *Advanced Engineering Mathematics* (John Wiley & Sons, New York) 1979.
[39] Guille A., Hacid H., Favre C. and Zighed D. A., *SIGMOD Rec.*, **42** (2013) 17.
[40] Onnela J.-P. *et al.*, *Proc. Natl. Acad. Sci. U.S.A.*, **104** (2007) 7332.
[41] Wang Q., Gao J., Zhou T., Hu Z. and Tian H., *EPL*, **114** (2016) 58004.
[42] Deville P., Wang D., Sinatra R., Song C., Blondel V. D. and Barabási A.-L., *Sci. Rep.*, **4** (2014) 4770.
[43] Murphy K. M., Shleifer A. and Vishny R. W., *Q. J. Econ.*, **106** (1991) 503.
[44] Lü L., Chen D.-B., Ren X.-L., Zhang Q.-M., Zhang Y.-C. and Zhou T., *Phys. Rep.*, **650** (2016) 1.