Mao et al. *EPJ Data Science* (2015) 4:15
DOI 10.1140/epjds/s13688-015-0053-1

EPJ Data Science
a SpringerOpen Journal

REGULAR ARTICLE Open Access

Quantifying socio-economic indicators in developing countries from mobile phone communication data: applications to Côte d'Ivoire

Huina Mao[1*], Xin Shuai[2], Yong-Yeol Ahn[3] and Johan Bollen[3]

*Correspondence: maoh@ornl.gov
[1]Oak Ridge National Laboratory, Oak Ridge, TN, USA
Full list of author information is available at the end of the article

Abstract

The widespread adoption of mobile devices that record the communications, social relations, and movements of billions of individuals in great detail presents unique opportunities for the study of social structures and human dynamics at very large scales. This is particularly the case for developing countries where social and economic data can be hard to obtain and is often too sparse for real-time analytics. Here we leverage mobile call log data from Côte d'Ivoire to analyze the relations between its nation-wide communications network and the socio-economic dynamics of its regional economies. We introduce the CallRank indicator to quantify the relative importance of an area on the basis of call records, and show that a region's ratio of in- and out-going calls can predict its income level. We detect a communication divide between rich and poor regions of Côte d'Ivoire, which corresponds to existing socio-economic data. Our results demonstrate the potential of mobile communication data to monitor the economic development and social dynamics of low-income developing countries in the absence of extensive econometric and social data. Our work may support efforts to stimulate sustainable economic development and to reduce poverty and inequality.

Keywords: mobile phone data; network analysis; big data analysis; socio-economic measurement; developing countries; economic development

1 Introduction

Accurate and timely information is a necessary condition for the implementation of policies that fosters socio-economic development. Hence, governments and private organizations invest significant resources in the construction of socio-economic indicators that are frequently derived from resource-intensive surveys and economic reports. For instance, the U.S. Department of Labor records unemployment insurance claims on a weekly basis to monitor changes in labor market conditions. The University of Michigan and Thomson Reuters publish U.S. Consumer Confidence index every month obtained from extensive consumer surveys. Gallup reports the daily U.S. Economic Confidence index, which is derived from comprehensive and large-scale economic surveys. The coverage of these indicators is not only significant from a temporal perspective, but also in spatial terms: the

© 2015 Mao et al. This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

American Community Survey of U.S. Bureau of the Census reports the household income data of more than 3,100 counties (or equivalent) in the country. The U.K. government measures the Index of Multiple Deprivation - a composite index of income, employment, education, health, crime, and housing - for 32,482 communities across the entire country.

Given the costs and resources involved with the construction of such socio-economic indicators, developing countries struggle to maintain similar levels of coverage. Côte d'Ivoire might be illustrative of this issue. It is a lower-middle income country in the Sub-Saharan Africa with an area that is 50% larger than the UK, comprising 19 regions and 81 departments, and a total population of more than 24 million individuals. However, the most recent economic data we were able to obtain for Côte d'Ivoire are the income and poverty rate measured at the development pole level,[a] reported more than seven years ago [1]. Developing countries have an urgent need for accurate, timely, and affordable social and economic indicators to implement effective policies to promote economic growth, reduce poverty, improve urban planning, and optimize resource allocation.

An unprecedented amount of real-time data is presently generated from a variety of consumer electronics, e.g. cell phones, social media, and a variety of electronic sensors, and they permeate most societies worldwide. This data has led to the development of real-time socio-economic indicators that may complement traditional ones, although most of studies focus on developed countries. Numerous investigations have demonstrated the ability of social media and web search data to measure socio-economic activities at a fine spatial and temporal resolution [2]. For example, social media data can be used to track U.S. job losses [3], consumer confidence [4], social mood [5, 6], and investor sentiment [7]. Certain web search queries may correlate with a country's GDP [8], unemployment rate [9], or even predict financial markets [10, 11]. However, the access to the web is very limited in low-income areas, causing many of the above mentioned indicators to not be feasible exactly for where they are needed the most. For example, only 16% of the African population use the Internet [12] and this number is even lower for Côte d'Ivoire (only 2.72%). By contrast, mobile communication network can be a promising platform, because the generation rate is high even in developing countries. For instance, Côte d'Ivoire's mobile penetration rate is 83% [12]. Mobile phone usage continues to increase in Sub-Saharan Africa, becoming the prime communication infrastructure, while landlines in recent years are used by less than 0.1% of population [13]. Therefore, in the foreseeable future mobile communication networks will remain the most viable infrastructure for socio-economic data collection in developing countries.

Mobile communication data has enabled a plethora of intriguing studies on social phenomena, such as disease transmission [14], human communications [15], human mobility [16], mobile phone virus outbreak [17], and response to disasters [18]. For instance, the correlation between social communication diversity of local communities and their socio-economic well-being index is found in a recent work [19], suggesting that social structure may influence economic development: heterogeneous or diverse social contacts [20–22] provide more opportunities or channels for information and innovation diffusion, thus advancing economic development. Of course, the observed correlations do not reveal the presence nor direction of causality, but at the very least these studies indicate a strong relation between communication patterns and socio-economic development, which can be leveraged for the construction of real-time indicators.

The plurality of the mentioned studies focus on developed countries, perhaps due to the availability of extensive economic data. It is thus unclear whether existing findings can be generalized to developing nations. Here we investigate whether mobile communication data can fill the gaps in economic data of low-income countries, which may allow researchers and governments to study regional economies using existing mobile infrastructure. Leveraging the mobile phone usage dataset provided by the France Telecom-Orange as a part of the Data for Development challenge in 2013, we study how the structure of a nation's social communication networks relates to its economic development, focusing on Côte d'Ivoire. Its challenges and opportunities are representative of those of many other low-income developing countries and we believe that similar methods can be implemented for other developing countries. The main contributions and findings of this paper are as follows:

- We develop a set of call activity based indicators and demonstrate that some of these correspond to vetted socio-economic statistics. We find that social centrality in mobile communication networks - PageRank - can identify economic centers at the national and city levels. We also find that the degree to which an area initiates mobile communications is strongly related to its annual income and low poverty rate. If this effect is universal to developing countries, it may allow the construction of high-resolution, real-time indicators of socio-economic development.
- We identify regional communities based on mobile communication activities, and compare them to administrative boundaries. At the country level, we find that adjacent regions form communities, as found in previous studies [23–25]. We observe a similar pattern within the capital city of Côte d'Ivoire, Abidjan.
- We measure the rich-club coefficient of Côte d'Ivoire national mobile communication networks to find that rich areas are much more likely to communicate with other rich ones than expected. A rich-club is distributed across the South and South-West areas of Côte d'Ivoire, which mainly communicate with each other, and not as much as expected with other poor areas of the North and West.

2 Materials and methods

2.1 Dataset

Our call log data is provided by France Telecom-Orange through the Data for Development (D4D) Challenge[b] in 2013. Orange has about five million customers in Côte d'Ivoire, which is about one third of its mobile subscriptions, a significant sample of mobile users in Côte d'Ivoire. The data consists of anonymous metadata pertaining to 2.5 billion calls and SMS exchanges that took place from December 1, 2011 to April 28, 2012. In this study, we use two mobile phone datasets including the hourly antenna-to-antenna traffic and individual trajectories for 50,000 customers for two weeks with a spatial resolution down to the individual antenna. In addition, we use the geographical information (latitudes and longitudes) of the antennae and sub-prefectures.

The dataset pertains to calls between 1,231 cell towers that are distributed across 50 departments (see Figure 1 (left); the color in the graph has no special meaning). Based on regional economic levels, the entire nation is divided into 10 development poles: Centre-North, Centre-West, North-East, North, West, South, South-West, Centre, Centre-East, and North-West [1]. Figure 1 (right) shows a poverty map at the level of development poles, with darker colors indicating higher levels of poverty. As shown, economic development

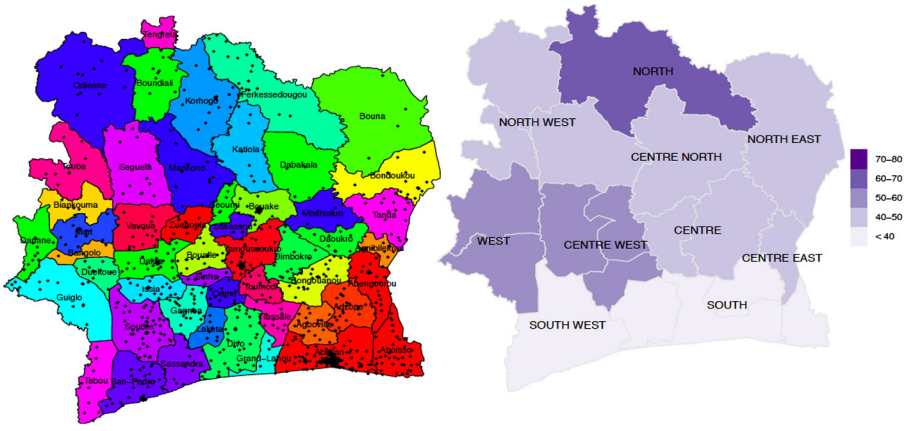

Figure 1 The distribution of cell towers and poverty map of Côte d'Ivoire. (a) The distribution of the Orange cell phone towers over 50 departments. **(b)** Poverty map of ten development poles.

in Côte d'Ivoire is distributed very unevenly with Southern regions generally being richer than Northern regions. The economic capital, Abidjan, further distinguishes itself from other areas: its annual income per capita is about three times higher than that of the north [1]. Strong income inequality also exists inside Abidjan: Cocody is the wealthiest commune; Plateau is the business district and central government area, and most of its residents are Caucasian; many slums are distributed in Adjamé, and Marcory and Treichville are also poor areas.

Figure 1 shows that the cellphone tower distribution is strongly skewed by the population and wealth: there are many more towers in the South than the North. Specifically, the South and South-West have 695 towers (i.e. 56% of all the Orange antennae), while the North only has 46 antennae. The Bafing region in the North West (a poor area) has nine towers, which is the least number of towers among all regions. Likewise, among 396 towers in Abidjan, Cocody - the wealthiest commune - has the largest number of towers (96), while Adjamé - the poorest commune - has only 20 towers. Among these three subprefectures outside Abidjan-Ville, there are only nine towers. Given the important role of Abidjan and the rich mobile phone data generated, we conduct our analysis not only at the country level, but also at the commune[c] level in Abidjan.

It is not surprising to find an unequal distribution of mobile phone antenna towers among rich and poor areas. The long-standing gap in information access and communication technology between different socio-economic groups is often referred to as the 'digital divide'. A strong digital divide may further exacerbate economic inequality and poverty. Wide adoption of telecommunication technologies, particularly mobile communication system, has therefore been considered a promising tool for poverty reduction since they provide easier and cheaper access to information.

In the following sections, we examine the mobile phone communication data (1) to develop socio-economic indicators to track the economic development within Côte d'Ivoire; and (2) to understand the relation between social communication patterns and economic development.

2.2 Network construction and terminology

We construct mobile communication networks between locations. Let $A = \{a_1, a_2, \ldots, a_n\}$ denotes all antennae, $R = \{r_1, r_2, \ldots, r_m\}$ denotes all regions, $D = \{d_1, d_2, \ldots, d_l\}$ denotes all departments, and $P = \{p_1, p_2, \ldots, p_k\}$ denotes all development poles in Côte d'Ivoire. As we mentioned before, $|A| = 1{,}231$, $|R| = 19$, $|D| = 50$ and $|P| = 10$. We construct two types of networks based on calls and trajectories on four different scales: antennae, departments, regions, and poles.

Call networks: The Dataset 1 contains the number (and duration) of calls between any pair of antennae for every hour during a five month period. We construct three types of networks: (1) $G_a^n = (A, N_a)$ is a directed weighted network where $N_a = \{n(a_i, a_j)\}$ denotes the total number of calls between a_i and a_j; (2) $G_a^d = (A, D_a)$ is also a directed weighted network where $D_a = \{d(a_i, a_j)\}$ denotes the total duration of calls between a_i and a_j. We further map antennae into different levels of administrative areas based on their geolocation and aggregate the mobile communication flow (in terms of numbers or duration of calls) between pairs of administrative areas. For instance, $G_d^n = (D, N_d)$ is a directed weighted network of department-to-department communication, where $N_d = \{n(d_i, d_j) = \sum_{(a_u, a_v): a_u \in d_i \wedge a_v \in d_j} n(a_u, a_v)\}$ denotes the total number of calls between department d_i and d_j, aggregated from their antenna records. Here mobile communications within the same department are ignored. Similarly, we generate call networks G_r^n (region-to-region number of calls), G_p^n (pole-to-pole call volume), G_d^d (department-to-department duration of calls), G_r^d (region-to-region duration of calls), and G_p^d (pole-to-pole duration of calls), respectively.

Mobility networks: The Dataset 2 contains individual movement trajectories, approximated by the geographic location of the cell phone antennae during calls. We define a movement trajectory of a user u_i during a period of time as a sequence of antennae that his/her mobile phone connected to: $S(u_i) = \{a_{s_1}, a_{s_2}, \ldots, a_{s_n}\}$. Then for each pair $(a_{s_i}, a_{s_{i+1}})$ we build an edge from a_{s_i} to $a_{s_{i+1}}$ if $a_{s_i} \neq a_{s_{i+1}}$ and use $w(a_{s_i}, a_{s_{i+1}})$ to count the frequency of such movements in $S(u_i)$. Finally, we obtain a weighted directed trajectory record network $G_a^t = (A, T_a)$, where $T_a = \sum_{u \in U} w(a_{s_i}, a_{s_{i+1}})$ denotes the collective trajectories of all users $U = \{u_1, u_2, \ldots, u_z\}$. There may be overlapping users, but it does not likely affect our collective trajectory, since 50,000 users are randomly sampled for each fortnight period during the whole five month period. Similar to analyzing the call record network as above, we can aggregate antennae into different levels of administrative areas, and construct mobility networks - G_d^t, G_r^t, and G_p^t - respectively.

In summary, we construct twelve networks based on four levels (antennae, departments, regions, and poles) and three measures (number and duration of calls, and movement trajectory). In addition, for a case study we separately analyze the mobile phone communication network of the capital city, Abidjan, G_c^*, which aggregates the flow of antenna-to-antenna in its communes. Among all constructed call and mobility networks, we choose proper ones for the analysis at different levels.

3 Results

3.1 Visualization of mobile phone calls and human mobility

As a basic exploration of data, we visualize the call and mobility graphs on the map of Côte d'Ivoire.

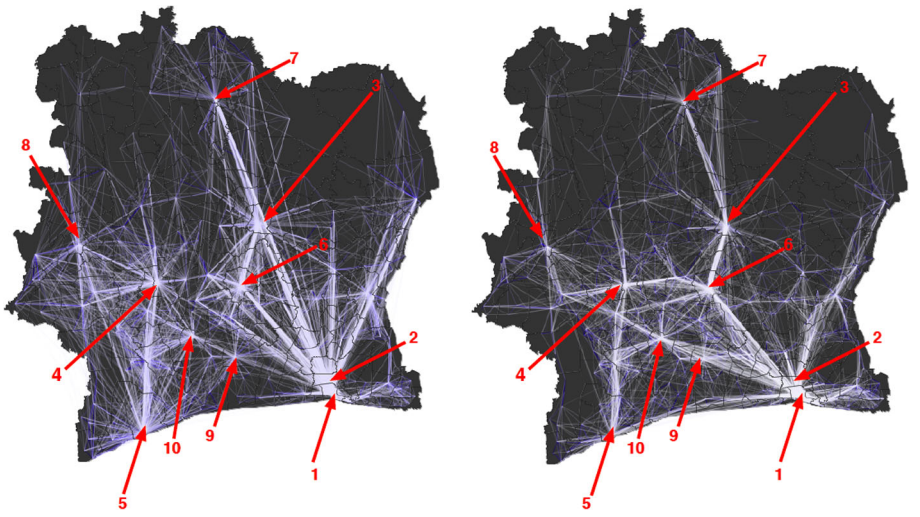

Figure 2 Network traffic visualization at the antenna level. (Left) Mobile phone call. **(Right)** Human mobility. The brightness of the edges indicates the weight defined by the number or duration of phone calls in G_a^n, G_a^d, as well as the flux of people between connected nodes in G_a^t. To improve clarity, we remove low-weighted edges. Ten largest cities in Côte d'Ivoire are 1: Abidjan; 2: Abobo; 3: Bouaké; 4: Daloa; 5: San-Pédro; 6: Yamoussoukro; 7: Korhogo; 8: Man; 9: Divo; 10: Gagnoa.

First, we normalize the weight of each edge by:

$$W_n = \frac{W - W_{\min}}{W_{\max} - W_{\min}}, \tag{1}$$

where W_n is the normalized weight, W is the raw weight, $W_{\max}$ and $W_{\min}$ are the maximum and minimum weights. We set up a threshold ξ and filter all edges with $W_n < \xi$. We set $\xi = 0.001$ for G_a^n, $\xi = 0.01$ for G_a^d and $\xi = 0.0003$ for G_a^t, which are manually tuned for clear visualization of networks. Note that this thresholding is used solely for the visualization, not for the other analyses.

The call network, G_a^n, is shown in Figure 2 (left). The normalized weight attached each edge is proportional to the brightness of the corresponding edge. It is clear that mobile communication activities reflect the economic development: the brightest part is Abidjan, the economic capital of Côte d'Ivoire. The southern part is much brighter than the northern part, which is mostly dark except for the capital city of the North - Korhogo. The North-West and North-East are the poorest development poles and the darkest in the map. Most of the bright hubs are the ten largest cities based on population [26]. We mark these ten cities in Figure 2 (left) using numbers from 1 to 10. The call duration network, G_a^d, shows almost identical patterns as above, and is not shown.

The mobility network, G_a^t, is shown in Figure 2 (right). Comparing Figure 2 (left) with Figure 2 (right), the most evident pattern is that mobile phones facilitate communication between the North and South regions that are far from each other. Specifically, there are a large number of calls between Abidjan and Korhogo. Abidjan is the major economic and trade center, and Korhogo is an important producer of agricultural goods. Hence, the frequency of mobile communications between regions may be reflective of their economic ties. In the following sections we further analyze call traffic, and study its correlation to socio-economic indicators.

3.2 Correlation between mobile call activities and economic indicators

3.2.1 CallRank and economic activity

PageRank, which captures the relative importance of web pages based on random walk process [28], is one of the most widely used centrality measures in network analysis. We adopt PageRank to define an indicator - CallRank - for measuring the importance of an area based on mobile communication networks. Although the algorithm definition is exactly that of PageRank, hereafter we refer to it as CallRank to stress that we are calculating the metric over a network of mobile phone calls and the resulting rankings pertain to the mobile phone communication network. We first measure CallRank for 19 regions using region-to-region number of calls weighted network, G_r^n. Table 1 shows the names of 19 regions of Côte d'Ivoire, the development pole that they belong to, CallRank scores of each region, and the annual average per capita income of each development pole.

The CallRank score of Lagune is 0.2658, which is much larger than that of any other regions. Lagune is located in the South development pole that is one of the richest areas in Côte d'Ivoire, and plays a leading role in the country's economy, containing both the first and second largest cities, Abidjan and Abobo. Bas-Sassandra has the second largest CallRank, which is located in the South-West with the highest annual average per capita income. The capital of Bas-Sassandra - San-Pédro - is one of the five largest cities in the nation [26] and the second largest port after Abidjan [27].

Given the high income and importance of Lagunes and Bas-Sassandra, it is not surprising to see their CallRank scores are the highest among all regions. However, CallRank is not determined by the level of income. For instance, the region Haut-Sassandra (in bold in Table 1), which is located in the Center-West that is one of the poorest poles in the nation, has the 3rd largest CallRank right next to Bas-Sassandra, the richest region in the country. Another poorest region, Savanes, with the lowest annual income, has a relatively high CallRank, i.e. 6th out of 19, which is even higher than CallRank of rich regions such as Sud-Bandama and Sud-Comoé. Then, what does CallRank capture? Although Haut-Sassandra and Savanes are poor regions, they play an important role in the

Table 1 CallRank of 19 regions of Côte d'Ivoire

Regions/poles	CallRank	Annual income* (levels)
Lagunes/South	0.2658	334,147 (Rich)
Bas-Sassandra/South-West	0.0996	348,257 (Rich)
Haut-Sassandra/Center-West	**0.0717**	**243,236 (Poor)**
Lacs/Center	0.0547	287,080 (Middle)
Vallée du Bandama/Center-North	0.0531	281,660 (Poor)
Savanes/North	**0.0459**	**191,540 (Poor)**
Moyen-Comoé/Center-East	0.0430	289,126 (Middle)
N'zi-Comoé /Center	0.0427	287,080 (Middle)
Agnéby/South	0.0366	334,147 (Rich)
Dix-Huit Montagnes/West	0.0362	256,319 (Poor)
Zanzan/North-East	0.0352	301,966 (Middle)
Marahoué/Center-West	0.0348	243,236 (Poor)
Sud-Bandama/South	0.0347	334,147 (Rich)
Fromager/Center-West	0.0346	243,236 (Poor)
Sud-Comoé/South	0.0334	334,147 (Rich)
Moyen-Cavally/West	0.0238	256,319 (Poor)
Worodougou/North-West	0.0231	284,393 (Poor)
Denguélé/North-West	0.0171	284,393 (Poor)
Bafing/North-West	0.0141	284,393 (Poor)

*Annual income is in Central African CFA frac (CFAF). 1 CFAF = 0.002 US Dollar.

Table 2 CallRank of 13 communities of capital city, Abidjan in Côte d'Ivoire

Communes	CallRank
Cocody	0.0827
Yopougon	0.0807
Abobo	0.0792
Adjame	0.0773
Plateau	0.0755
Attecoube	0.0727
Treichville	0.0727
Port-Bouët	0.0705
Marcory	0.0704
Koumassi	0.0698
Anyama	0.0658
Songon-Agban	0.0589
Bingerville	0.0530

country's economy: Haut-Sassandra contributes to a large proportion of cocoa production that is the main export and income of Côte d'Ivoire. Haut-Sassandra's capital, Daloa, is an important trading center, and is responsible for a quarter of the Côte d'Ivoire's national output. Savanes produces mostly cotton, cashew trees, fruit trees, and contains 66% of cattle from the country [29]. Savanes's capital, Korhogo, is one of ten biggest city of Côte d'Ivoire. On the other hand, regions that are in low annual income levels and play inactive roles in the economy have lowest CallRank scores, i.e. Worodougou (17/19), Denguélé (18/19) and Bafing (19/19). Therefore, CallRank seems to reflect economic importance of a region rather than its actual economic development level, which is in line with the original meaning of PageRank for quantifying the importance of nodes in networks.

Similarly, we calculate CallRank for ten communes in Abidjan, which are shown in Table 2. Cocody, the richest commune, has the highest CallRank; Yopougon, the most populous and rich commune, has the second highest CallRank. By contrast, the three subprefectures (i.e. Anyama, Songon-Agban, and Bingerville), which exercise few economic activities and have small population, score the lowest in terms of CallRank. However, the lack of socio-economic statistics at finer scales prohibits further quantitative investigations.

3.2.2 Correlation between mobile phone indicators and socioeconomic statistics

In addition to CallRank, we quantify other aspects of our communication and mobility graphs, and test the correlation between these network features and economic indicators. Significant correlations may suggest that mobile phone communication data can be used for developing countries to monitor and react to regional economic development swiftly without paying the high cost to execute high-quality national census or large-scale surveys.

Our economic data for Côte d'Ivoire is rather sparse, since it is aggregated at the level of ten development poles [1]. In addition to the total average annual per capita income and poverty rate, other economic indicators include Gini index and the ratio of average income in urban areas to average income in rural areas (i.e. U/R ratio). These economic data are collected from the International Monetary Fund country report [1]. We define and extract network features from G_p^n, G_p^d and G_p^t, and list them and their descriptions in Table 3.

Figure 3 shows Spearman rank correlation coefficients between mobile network indicators and economic statistics. Since we only have ten data points of development poles, it is

Table 3 Mobile network indicators from call and mobility networks at the pole level

Mobile phone indicators	Description
inFlow	The weighted in-degree
outFlow	The weighted out-degree
selfFlow	The sum of weights from antenna-to-antenna edges within the pole
externalFlow	inFlow + outFlow
totalFlow	selfFlow + externalFlow
exRatio	externalFlow/totalFlow
outRatio	outFlow/externalFlow
towernum	Number of towers
CallRank	PageRank
diversity	A measure of normalized entropy of communication adopted from [19]

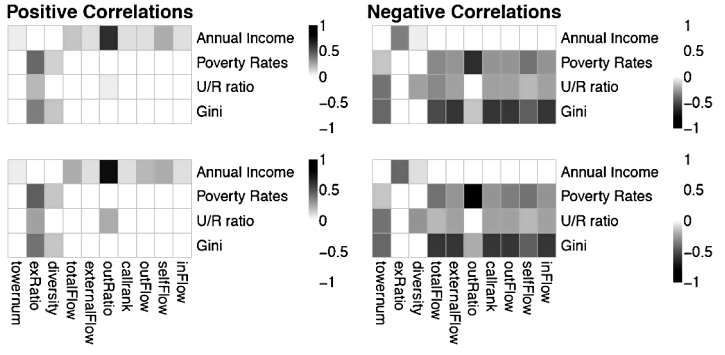

Figure 3 Correlation between economic indicators and mobile network measures at the pole level. **(Top)** Call volume weighted network vs. economy. **(Bottom)** Call duration weighted network vs. economy.

difficult to determine statistical significance. Yet, we observe a consistent signal between outRatio and the measures of average annual income and poverty. For G_p^n, the call volume weighted pole-to-pole network, the correlation coefficients are 0.63 and −0.66, respectively, and both correlations have p-values smaller than 0.05. The outRatio calculated from G_p^d, call duration weighted pole-to-pole network, shows an even stronger correlation with annual income (0.80) and poverty rate (−0.83) with p-values smaller than 0.01. In order to test the validity of our finding, we compare our results with randomly permuted networks in which we keep the same edge weight distribution but randomly permute the sequence of edges. We randomly shuffle the edges of the original network for 100 times, compute the outRatios of all nodes in the shuffled network, and then correlate them with the corresponding economic indicators, such as poverty rates. As a result, we obtain a distribution of 100 Spearman rank correlation coefficients, which follows a normal distribution with a standard deviation of 0.32 and a 95% confidence interval of mean at $[0.04 \pm 1.96 \times \frac{0.32}{\sqrt{100}}]$ (i.e. $[-0.02, 0.10]$). From the distribution, we conclude that the probability that the Spearman rank correlation coefficient is less or equal to −0.83 is only 0.0031. This indicates that our finding of the significant negative correlation (−0.83) between poverty rate and outRatio is highly unlikely to be random. Similarly, we find the probability that the correlation between outRatio and income and outgoing call ratio is greater than or equal to 0.80 is only 0.0061, which again confirms that their positive correlation is statistically significant. Therefore, the directionality of calls seems to have a strong predictive power of the economic level. This may indicate that rich areas have greater opportunities or means to initiate calls to other areas, or that the directionality mirrors their commanding eco-

nomic positions. Our finding alludes the studies on *pecking order* [30], a term originated from the dominance and hierarchy in chicken, which possibly allows the detection of social hierarchy from online social network topology [31]. In addition, we find that the Gini coefficient is correlated with various flow measures: in particular on the left of Figure 3 (i.e. negative correlation), we observe that the Gini index exhibits significant correlations with mobile phone indicators, such as *inFlow*, *outFlow*, and *CallRank*. Further investigations may reveal intriguing patterns about economy and communication structure.

We also correlate indicators from G_p^t (i.e. pole-to-pole mobility network) with the same regional economic statistics, but did not find any significant correlations between these indicators and economic statistics. We thus suggest that human mobility data may possess less predictive power for economic indicators than mobile phone communication data.

3.3 Communication-induced communities based on mobile phone data

Communication data allows us to study community boundaries based on social interactions rather than those based on historical or administrative divisions. In line with the previous work on United Kingdom [23], Belgium [24], and United States [25], we adopt the Louvain method [32] to perform community detection on three networks G_a^n (antenna-to-antenna number of calls), G_a^d (antenna-to-antenna duration of calls), and G_a^t (antenna-to-antenna movement trajectories).

The detected communities from three networks, G_a^n, G_a^d, and G_a^t, are shown in the left, middle, and right panels in Figure 4, respectively. Colors represent communities, the black borders represent administrative region boundaries, and the white lines represent subprefecture boundaries. Each prefecture is assigned to the community to which the majority of its antennae belong. The prefectures without any antenna are left blank.

We highlight three observations: (1) the number of communities detected from G_a^t (28) is larger than that from the communication networks, G_a^n (18) and G_a^d (7). This may suggest that mobile phones facilitate the communications across regions, thus merging more adjacent areas into one community; (2) all communities are geographically localized, a finding is consistent with the previous studies [23–25] that have shown that communication-based communities are well mapped into geographic space; (3) rich areas, the South and South-West, tend to split into smaller communities, while poor areas tend to merge into

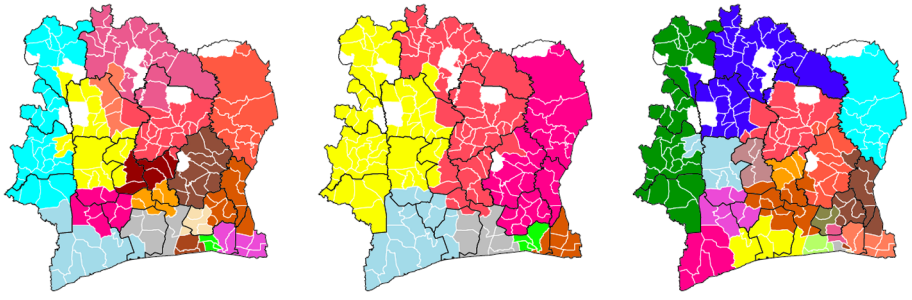

Figure 4 Network community detection in Côte d'Ivoire. (Left) 18 communities detected based on the call volume weighted network. **(Middle)** 7 communities detected based on the call duration weighted network. **(Right)** 28 communities detected based on the human mobility network. Communities are geographically close. There are fewer number of communities based on mobile phone call networks than the human mobility network, suggesting that mobile phones may facilitate the communications between adjacent areas.

a large community. This may be simply due to their larger number of antennae than poor areas. In a relevant work [33], the authors use the average size of airtime credit purchases as a proxy of the relative wealth of an individual based on the assumption that rich people can make larger purchases than poor people. They apply the Louvain method for community detection, and show that the communities within some cities (e.g. Abidjan, Bouake, and San Pedro) are diverse, i.e. people within the same community have diverging purchasing behavior with some people making small purchases while others making larger ones. This study may indicate that different socioeconomic groups can exist within the same community.

We now turn our focus to a single city, Abidjan, the capital of the country. Abidjan has 396 cell towers, more than 30% of all towers in the country. Approximately 3.6 million individuals (1/6 of the total population) live in Abidjan and strong disparities exist across its communes. The District of Abidjan consists of Abidjan-Ville and three external subprefectures: Anyama, Bingerville, and Songon. Abidjan-Ville is divided into two halves: southern Abidjan and northern Abidjan, which are connected by two bridges (Houphouët-Boigny and Charles de Gaulle). Southern Abidjan has six communes: Cocody (the wealthiest residential area), Plateau (the business district and central government area), Adjamé (the slum area), Yopougon (the most populous area), Abobo, and Attécoubé (shopping complex, the national park). Northern Abidjan include four communes: Marcory and Treichville (poor areas), Port-Bouët (home to the airport), and Koumassi (an important industrial area).

In order to study communication patterns among these ten communes of Abidjan-Ville, we compare the call volume between every two communes against the sum of their independent call volume at the log scale, i.e. $\log(V_{12})$ vs. $\log V_1 + \log V_2$, where V_1 and V_2 represents the call volume of commune 1 and commune 2, respectively; V_{12} represents the total call volume between these two communes. Note that here we only consider calls within Abidjan. The result is shown in Figure 5.

We find that Southern Abidjan communes - Marcory (ID: 5), Port-Bouët (ID: 6), Treichville (ID: 8), and Koumassi (ID: 9) - stand out, having a lot of calls between each other, given the independent call volume (see the blue triangle cluster on the left of Figure 5). Between the southern and northern communes as well as within the northern communes, we can see a clear scaling relation, i.e. their mutual call volume scales as a function of their individual call volume. Two possible factors may influence the communication within Abid-

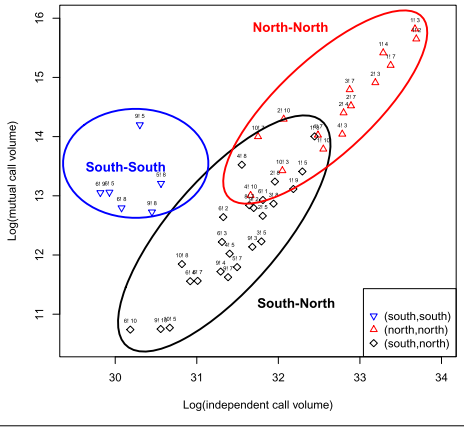

Figure 5 Mutual call volume vs. independent call volume of ten communes in Abidjan. Ten communes are 1: Cocody; 2: Yopougon; 3: Abobo; 4: Plateau; 5: Marcory; 6: Port-Bouët; 7: Adjamé; 8: Treichville; 9: Koumassi; 10: Attécoubé. The mutual calls between the south and the north as well as between the north and itself increase as the number of their individual call volume increase. In contrast, there are much more mutual calls made within southern communes compared to their independent call volume.

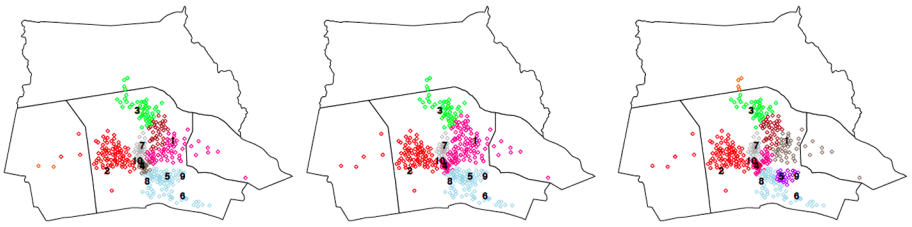

Figure 6 Network community detection in Abidjan. (Left) Eight communities detected based on the call volume weighted network. **(Middle)** Five communities detected based on the call duration weighted network. **(Right)** Nine communities detected based on the human mobility network. 1: Cocody; 2: Yopougon; 3: Abobo; 4: Plateau; 5: Marcory; 6: Port-Bouët; 7: Adjamé; 8: Treichville; 9: Koumassi; 10: Attécoubé. Adjacent areas tend to be in one community. The southern communes (including 5, 6, 8, and 9) form one community in the call networks (see the left and middle). In all the three networks, it can be seen that commune 7 (Adjamé), where major slums are located, stands as one single community separated from nearby communes.

jan: (1) Geography: Abidjan is divided into northern and southern parts by the lagoon. This geographical division may influence the communication patterns. (2) Socioeconomic differences: Northern Abidjan is the central part of the city, having the business center, administrative area, and upscale residential districts, while two of four communes in Southern Abidjan - Marcory and Treichville - are poor areas and the other two (i.e. Port-Bouët and Koumassi) have an international airport and industrial area. Besides the geographic factor and economic differences of Northern and Southern Abidjan, there could be cultural and ethical differences between them. The homophily principle, namely that people with similar characteristics (e.g. culture, religions, ethnicities, economic status) are more likely to form connections [34] may be at play, where the southern part of Abidjan may consist of more homogeneous population.

We further examine three networks of Abidjan: two call networks weighted by call volume (G_c^n) and duration time (G_c^d), and human mobility network (G_c^t). We show the community structure in these networks in Figure 6. Each tower is drawn as a circle based on its latitude and longitude. We denote the center of each commune by an ID ranging from 1 to 10, colors representing communities.

From the call frequency weighted network G_c^n, we detect eight communities as shown in Figure 6 (left), whereas only five communities are detected from call duration weighted networks with several adjacent communities merged further as shown in Figure 6 (middle). Nine communities are found based on human trajectory records (Figure 6 (right)), which are similar to the communities detected from the call frequency weighted network, with the exception of certain big communities that are further divided due to the geographical distance restrictions.

In sum, Figure 6 shows that (1) adjacent areas are more likely to be in the same community, which is consistent with earlier findings at the regional level; (2) the most populous communes, Yopougon and Abobo, form a single large community themselves (in red and green respectively); Cocody (the wealthiest commune) dominates the community in both call networks. The southern Abidjan communes with ID: 5, 6, 8, 9, merge into one big community in both Figure 6 (left) and Figure 6 (middle). This is consistent with our finding from Figure 5, from which we show that the southern communes strongly communicate among themselves; (3) Adjamé (ID: 7), where major slums are located, forms a small community (in gray at the center) consistently across all networks. This finding may reflect the isolation of the poorest region, although Adjamé does not stand out in Figure 5. Note that

the community boundaries may change if different community detection methods are applied. To further test the socioeconomic segregation phenomena, we analyze the rich-club effect of the national mobile phone communication network in the next section.

3.4 Rich-club analysis

The rich-club coefficient quantifies the degree of connectivity among *rich* nodes in a network, quantifying the strength of the 'rich club' effect [35, 36]. The *richness* of a node can take various definitions, such as degree, centrality, or other measures. Here we define richness of a node (i.e. a development pole) as the average annual income of the region and use the weighted rich-club coefficient proposed by [37].

Every node has a richness parameter r. For each value of r, a club that consists of all nodes with richness larger than r is formed. For each of these clubs, $E_{>r}$, the number of links connecting the members, and $W_{>r}$, the sum of the weights attached to these links are measured. Then we calculate the ratio of $W_{>r}$ to the sum of the weights attached to the $E_{>r}$ strongest links within the whole network as follows:

$$\phi^w(r) = \frac{W_{>r}}{\sum_{l=1}^{E_{>r}} w_l^{\text{rank}}}, \tag{2}$$

where $w_l^{\text{rank}} \geq w_{l+1}^{\text{rank}}$ with $l = 1, 2, \ldots, E$ are the ranked weights of links on the network and E is the total number of links. To account for the factor that even random networks can exhibit a baseline degree of rich-club effect, the null model is generated by randomizing the original network while preserving its degree distribution. The rich-club coefficient is thus defined as:

$$\rho^w(r) = \frac{\phi^w(r)}{\phi^w_{\text{null}}(r)}, \tag{3}$$

where $\rho^w(r)$ refers to the weighted rich-club effect as assessed vs. the appropriate null model. When $\rho^w(r)$ is larger than one, the observed rich-club coefficient in original network is larger rather than expected from the random null-model.

We calculate the rich-club coefficient for G_p^n, G_p^d, and G_p^t as a function of richness levels measured by the annual income. The results of G_p^n and G_p^d are similar, so we only show G_p^n in Figure 7 (left) and G_p^t in Figure 7 (right). Figure 7 shows a sudden increase in rich-club coefficient when the income level is above CFAF 300,000. From Table 1, we find that only three poles have entered the CFAF 300,000 level 'rich-club', i.e. South, South-West, and North-East. This result demonstrates that rich areas form the 'rich club' in mobile communication and mobility networks; they mainly communicate with each other and separate themselves from poor areas.

4 Discussion

We analyze the mobile phone call logs and human mobility traces from a large-scale mobile dataset collected in Côte d'Ivoire with the aim to further understand the country's economic development. First, we develop several network indicators from call and mobility networks, and compare them to economic indicators. We found that CallRank informs us about the importance of regions, and the relative frequency of initiating calls to other

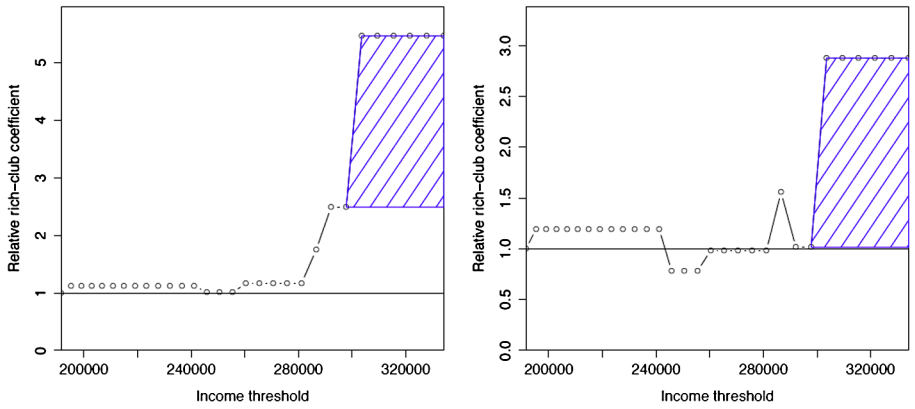

Figure 7 Rich-club effect analysis. The shaded area show that the 'rich-club' consists of poles with higher than CFAF 300,000 annual average per capita income. **(Left)** Rich-club in mobile communication network. **(Right)** Rich-club in human mobility network.

areas (i.e. outgoing call ratio) consistently correlates with local economic statistics like low poverty rate and high annual income. Our research implies that features derived from mobile phone data may be useful in measuring and predicting economic development, thereby complementing scarce economic statistics in developing countries. Second, we identify regional communities from mobile communication and mobility graphs. We confirm previous results that have shown that, although mobile phones facilitate communications across regions, people are inclined to communicate with others who are geographically close. Socioeconomic segregation is suggested by the results of a rich-club analysis at the country level, where we find that rich areas communicate more frequently with other rich areas than poor areas, thus forming a 'rich-club'. Social connectivity can improve information transfer, technical innovation, and economic development [38], whereas social segregation can aggravate economic inequality, social instability, and impede economic growth in the long-term. Future work may study the degree of social connectivity across different countries (e.g. developed and least developed countries), and its causal relation to their economic development with the aim to find better social structure for sustainable development.

In summary, our work demonstrates the promising possibility of leveraging mobile phone traces to monitor economic activities in developing countries, which frequently lack advanced information infrastructure and resources. Continued adoption of mobile phones may underpin efforts to more accurately observe fundamental social and economic dynamics in low-income countries and support efforts to foster social and economic development.

Competing interests
The authors declare that they have no competing interests.

Authors' contributions
HM, XS, YA, and JB conceived the research ideas. HM and XS conducted data analysis and prepared figures. All the authors participated in data analysis and manuscript preparation.

Author details
[1]Oak Ridge National Laboratory, Oak Ridge, TN, USA. [2]Thomson Reuters, St Paul, MN, USA. [3]School of Informatics and Computing, Indiana University, Bloomington, IN, USA.

Acknowledgements
At least one or more of the authors of this manuscript are employees of UT-Battelle, LLC, under contract DE-AC05-00OR22725 with the U.S. Department of Energy. Accordingly, the United States Government retains and the publisher, by accepting the article for publication, acknowledges that the United States Government retains a non-exclusive, paid-up, irrevocable, world-wide license to publish or reproduce the published form of this manuscript, or allow others to do so, for United States Government purposes. The work was conducted while Huina Mao and Xin Shuai were with Indiana University.

Endnotes
[a] There are ten development poles in Côte d'Ivoire. Each development pole includes one or more administrative regions with similar income levels.
[b] http://www.d4d.orange.com/home.
[c] Defined as a French administrative division, roughly equivalent to townships.

Received: 12 August 2015 Accepted: 5 October 2015 Published online: 13 October 2015

References
1. IMF (2009) Cote d'Ivoire: poverty reduction strategy paper. Country report 09/156, IMF
2. Einav L, Levin JD (2013) The data revolution and economic analysis. Technical report, National Bureau of Economic Research
3. Antenucci D, Cafarella M, Levenstein MC, Ré C, Shapiro MD (2014) Using social media to measure labor market flows. Technical report, National Bureau of Economic Research
4. O'Connor B, Balasubramanyan R, Routledge BR, Smith NA (2010) From tweets to polls: linking text sentiment to public opinion time series. In: ICWSM, pp 122-129
5. Bollen J, Mao H, Zeng X-J (2011) Twitter mood predict the stock market. J Comput Sci 2(1):1-8
6. Gilbert E, Karahalios K (2010) Widespread worry and the stock market. In: ICWSM, pp 59-65
7. Mao Y, Wei W, Wang B, Liu B (2012) Correlating S&P 500 stocks with Twitter data. In: Proceedings of the first ACM international workshop on hot topics on interdisciplinary social networks research. ACM, New York, pp 69-72
8. Preis T, Moat HS, Stanley HE, Bishop SR (2012) Quantifying the advantage of looking forward. Sci Rep 2:350
9. Ettredge M, Gerdes J, Karuga G (2005) Using web-based search data to prediction macroeconomic statistics. Commun ACM 48(11):87-92
10. Da Z, Engelberand J, Gao P (2015) The sum of all fears: investor sentiment and asset prices. Rev Financ Stud 28:1-32
11. Da Z, Engelberand J, Gao P (2011) In search of attention. J Finance 66(5):1461-1499
12. ITU (2013) The world in 2013: ICT facts and figures
13. ITU (2009) Information society statistical profiles 2009: Africa
14. Wesolowski A, Eagle N, Tatem AJ, Smith DL, Noor AM, Snow RW, Buckee CO (2012) Quantifying the impact of human mobility on malaria. Science 338(6104):267-270
15. Onnela J-P, Saramäki J, Hyvönen J, Szabó G, Lazer D, Kaski K, Kertész J, Barabási A-L (2007) Structure and tie strengths in mobile communication networks. Proc Natl Acad Sci USA 104(18):7332-7336
16. Gonzalez MC, Hidalgo CA, Barabasi A-L (2008) Understanding individual human mobility patterns. Nature 453(7196):779-782
17. Wang P, González MC, Hidalgo CA, Barabási A-L (2009) Understanding the spreading patterns of mobile phone viruses. Science 324(5930):1071-1076
18. Bengtsson L, Lu X, Thorson A, Garfield R, von Schreeb J (2011) Improved response to disasters and outbreaks by tracking population movements with mobile phone network data: a post-earthquake geospatial study in Haiti. PLoS Med 8(8):1001083
19. Eagle N, Macy M, Claxton R (2010) Network diversity and economic development. Science 328(5981):1029-1031
20. Page SE (2008) The difference: how the power of diversity creates better groups, firms, schools, and societies. Princeton University Press, Princeton
21. Newman MEJ (2003) The structure and function of complex networks. SIAM Rev 45(2):167-256
22. Granovetter M (1973) The strength of weak ties. Am J Sociol 78(6):1360-1380
23. Ratti C, Sobolevsky S, Calabrese F, Andris C, Reades J, Martino M, Claxton R, Strogatz S (2010) Redrawing the map of Great Britain from a network of human interactions. PLoS ONE 5(12):14248
24. Blondel V, Krings G, Thomas I (2010) Regions and borders of mobile telephony in Belgium and in the Brussels metropolitan zone. Brussels Stud: Article ID 42
25. Thiemann C, Theis F, Grady D, Brune R, Brockmann D (2010) The structure of borders in a small world. PLoS ONE 5(11):15422
26. Ivory Coast - largest cities. http://www.geonames.org/CI/largest-cities-in-ivory-coast.html. Accessed 22 Sept 2015
27. San-Pédro, Ivory Coast. Wikipedia, the free encyclopedia. https://en.wikipedia.org/wiki/San-P%C3%A9dro,_Ivory_Coast. Accessed 22 Sept 2015
28. Page L, Brin S, Motwani R, Winograd T (1999) The pagerank citation ranking: bringing order to the web. Technical report, Stanford University, Stanford, CA
29. Schmitte E (2012) The importance of social networks to inform and support farmers about adaptation strategies regarding climate change in Côte d'Ivoire. Master's thesis, Federal Institute of Technology, Zürich, Switzerland
30. Schjelderup-Ebbe T (1922) Contributions to the social psychology of the domestic chicken. Z Psychol 88:225-252
31. Gupte M, Shankar P, Li J, Muthukrishnan S, Iftode L (2011) Finding hierarchy in directed online social networks. In: Proceedings of the 20th international conference on world wide web. ACM, New York, pp 557-566
32. Blondel VD, Guillaume JL, Lambiotte R, Mech ELJS (2008) Fast unfolding of communities in large networks. J Stat Mech Theory Exp 2008(10):10008
33. Gutierrez T, Krings G, Blondel VD (2013) Evaluating socio-economic state of a country analyzing airtime credit and mobile phone datasets. http://arxiv.org/abs/1309.4496

34. McPherson M, Smith-Lovin L, Cook JM (2001) Birds of a feather: homophily in social networks. Annu Rev Sociol 27:415-444
35. Colizza V, Flammini A, Serrano MA, Vespignani A (2006) Detecting rich-club ordering in complex networks. Nat Phys 2(2):110-115
36. Zhou S, Mondragón RJ (2004) The rich-club phenomenon in the Internet topology. IEEE Commun Lett 8(3):180-182
37. Opsahl T, Colizza V, Panzarasa P, Ramasco JJ (2008) Prominence and control: the weighted rich-club effect. Phys Rev Lett 101(16):168702
38. Andris C, Bettencourt LM (2013) Development, information and social connectivity in Côte d'Ivoire. Working paper 13-06-023, Santa Fe Institute

Submit your manuscript to a SpringerOpen journal and benefit from:

► Convenient online submission
► Rigorous peer review
► Immediate publication on acceptance
► Open access: articles freely available online
► High visibility within the field
► Retaining the copyright to your article

Submit your next manuscript at ► springeropen.com