# Modern Food Foraging Patterns: Geography and Cuisine Choices of Restaurant Patrons on Yelp

Qi Xuan, *Member, IEEE*, Mingming Zhou, Zhi-Yuan Zhang, Chenbo Fu, Yun Xiang, Zhefu Wu, and Vladimir Filkov

***Abstract*— Animals search for food based on certain optimal principles and over time form foraging patterns effective for survival in changing environments. Due to the many choices available in modern society, we also face a decision on where to get their food. We call this "modern human food foraging," since the Internet makes foraging much more convenient than before. People search online for food venues, or restaurants, through websites such as Yelp, and write reviews for the food they tasted, which in turn, facilitate others' searches in the future. These activities make the whole community of restaurant patrons wiser over time. Moreover, the archives of all these choices and evaluations are publicly available, and can help researchers better understand human foraging patterns in modern society. In this paper, we use a Yelp data set to study modern human food foraging patterns, with respect to both geography and cuisine. To understand spatial patterns, we cluster reviewed restaurants geographically and construct a taste similarity network, representing the topology of restaurant cuisine space. We find that people steadily expand their foraging domains from the nearest to them to the distant in geography and from the most familiar to the novel in cuisine. Using longitudinal data of restaurant reviews, we build a geographical foraging network and a taste foraging network for each patron based on which, we propose three kinds of entropies to characterize foraging patterns. We show that the modern foraging patterns of restaurant patrons in both geography and cuisine are of high regularity, indicating that their behaviors are rather predictable. The foraging patterns are also associated with individual social status in the community. Namely, people having a higher variety in the restaurant cuisines they have visited, but fewer actual locations they visited, tend to attract more followers.**

***Index Terms*— Geographical factor, human foraging pattern, layered network, sequence analysis, structural complexity.**

## I. Introduction

FORAGING behavior, or food selection, plays an important role in the ability of many organisms to survive and reproduce [1]. A series of economic models have been proposed to understand foraging behavior, in terms of optimizing a payoff from a foraging decision [2]. It was found that the organisms prefer to select and persist with the foraging behavior delivering the highest ratio of energetic gain to cost per unit time, where the energetic gain is associated with food resources and cost partly attributed to foraging distance. Restaurant site selection, e.g., is a typical application of the highest energetic gain principle in modern society, i.e., many restaurants of different cuisines are built near high population density areas and transportation hubs [3], to ensure that people can easily get there and find satisfactory food, so they can be profitable. It can be argued that while the geographical factor is a major determinant in calculating the cost of searching for food, people's food and taste preference can play a big role, especially because tastes can vary and different foods may be preferred at different times [4].

In modern societies, restaurants in many places provide tremendous variety of cuisines and food choices, such as seafood, Italian grill, American grill, pizza, and so on, whereas each restaurant may only provide few of them. For example, a restaurant on Yelp with a tag *seafood* may provide various kinds of fish, and a restaurant with a tag *bakeries* may provide pastries and croissants. Such cuisine-related tags are meant to improve the search efficiency and help patrons find what they need quickly [5]. Reviews of tagged restaurants are helpful to the Yelp community and also provide a good opportunity to document the taste of a person at a given time, i.e., the preference of a person for certain kinds of food. Together with the geographical information, these can facilitate the design of more personalized recommendation systems. Therefore, the term *modern human food foraging* means that, different from former human food foraging, such behavior is influenced, and generally facilitated, by modern social media and online recommendation systems.

Although there are many studies utilizing social media data, most of them simply focus on general human geographical foraging patterns and rarely on individual or social aspects. Social individuals learn from each other to maximize their fitness in a changing environment [6]. Nowadays, many social networks emerge in online communities to facilitate crowdsourcing coordination and information filtering. For instance, in Apache Software Foundation, developers communicate with others through email to better solve programming problems [7]. On Twitter, people follow others to get important information in time [8]. On Yelp, a patron can add friends (two-way selection) to form a friendship network or simply "follow" others to get their current interests. Since the "follow" relationship

Manuscript received March 18, 2017; revised October 28, 2017 and February 27, 2018; accepted March 10, 2018. Date of publication April 25, 2018; date of current version May 25, 2018. This work was supported in part by the National Natural Science Foundation of China under Grant 61572439, Grant 61273212, and Grant 11505153, and in part by the Natural Science Foundation of Zhejiang Province under Grant LY18F030021 and Grant LY18F010025. *(Corresponding author: Qi Xuan.)*

Q. Xuan, M. Zhou, C. Fu, Y. Xiang, and Z. Wu are with the College of Information Engineering, Zhejiang University of Technology, Hangzhou 310023, China (e-mail: xuanqi@zjut.edu.cn; zmmzjut@gmail.com; cbfu@zjut.edu.cn; xiangyun@zjut.edu.cn; wuzf@zjut.edu.cn).

Z.-Y. Zhang and V. Filkov are with the Department of Computer Science, University of California at Davis, Davis, CA 95616 USA (e-mail: zyzh@ucdavis.edu; filkov@cs.ucdavis.edu).

Digital Object Identifier 10.1109/TCSS.2018.2819659

is mostly about information dissemination [9] and does not need mutual confirmation, the number of followers, rather than friends, can be considered as a more appropriate measurement to characterize the *prestige* [10], [11] of an individual in the community. We thus define the number of followers as individual *social status* or *online status*.

Based on these, to analyze the modern foraging behaviors, we propose the following four research questions.

First, considering that the geographical distance and taste preference could influence the individual food foraging patterns based on the principle of economics, we ask the following question:

**RQ#1:** How do Yelp patrons expand foraging domains with respect to geography and cuisine?

Second, it stands to reason that a person who has visited more restaurants will likely have experienced food of various cuisines. On the other hand, we may also argue that people with narrower tastes are more picky and thus would not mind exploring and visiting more places to find satisfactory food. We thus ask the following question:

**RQ#2:** Do Yelp patrons with narrower tastes tend to visit more places, i.e., more geographical clusters of restaurants?

Third, recent studies on human dynamics showed that, although people may visit a large number of places, their mobility patterns over time are rather regular and thus their behavior is highly predictable [12]. Here, we adopt similar methods to study the complexity of human foraging patterns in two dimensions: geography and cuisine, to answer the following question.

**RQ#3:** To what extent can the foraging behaviors be predicted? Do Yelp patrons tend to switch between more similar cuisines?

Fourth, generally speaking, people who focus on the restaurants in particular locations may have stronger influence in these regions and thus earn higher social status. On the other hand, people who have wider variety of tastes may provide more profound reviews so as to attract more followers of diverse tastes. It is interesting to gain insight into the following question.

**RQ#4:** What is the relationship between social status and foraging patterns with respect to geography and cuisine?

To answer the above research questions, we use a Yelp data set, with geographic and cuisine (food type) tags for restaurants, to understand the principles of modern human foraging. Specifically, we make a dual contribution. First, in methodology, from the data, we construct networks of three different types: geographical foraging networks (GFNs), taste foraging networks (TFNs), and taste similarity networks (TSNs). For each restaurant patron in the Yelp data, we create a GFN, with nodes denoting the clusters of restaurants that the person has ever visited, and links if locations were visited in succession. From the cuisine tags, similarly, we establish a TFN for each patron, where each node represents a cuisine tag that the patron has ever had, and two nodes are linked if those cuisines have been tried in successive times. For the TSN, each node also represents a cuisine tag and two nodes are linked if the corresponding tags are ever attached to the same restaurant. We then use information theory, orthogonal decomposition, and multiple linear regression (MLR) to reveal the correlation between modern foraging patterns and individual social status in the Yelp community.

Our second empirical contribution is in applying the quantitative framework to Yelp data set, with the following results.

1) We find that the people generally search for food by expanding their geographical and cuisine spaces over time, from the near to the distant in geography and from the familiar to the novel in cuisine.
2) We present that most often people revisit a small number of locations and tastes frequently, making their behaviors highly predictable.
3) Moreover, we show that, when controlled for a number of confounds, those people who prefer higher cuisine variety, review higher priced restaurants, and forage in fewer locations tend to have more followers.

## II. Related Work

More recently, large-scale check-in data were collected from social media, such as Foursquare, Twitter, Facebook, and Yelp, to understand urban human activity patterns. With the promulgation of information technology, trajectories of people in the real world and online can be more easily tracked than before, providing a good opportunity for researchers to quantitatively study human behavior. Studying modern foraging patterns in a quantitative way, thus, is beneficial at least in the following two aspects: 1) it can aid urban planning and development by providing more compatible restaurant matches to the surrounding communities, in order to better serve local residents [13]–[16], e.g., local governments may better understand the taste and location preferences of the residents on restaurants, based on which they can give policies to increase the agreement between the two and 2) it can help online crowd review systems, such as Yelp, to better understand the temporal behaviors of restaurant patrons and facilitate personalized recommendations of higher fidelity [17]–[20]. For instance, the systems can recommend new food to the patrons with diverse tastes, whereas they may recommend traditional food with good word of mouth to those with narrower tastes, since they may be more picky about food.

### A. Human Behavior Pattern

With respect to foraging mobility, a recent study used GPS devices to study the movement pattern of Hadza individuals when they gathered food from the wild [21]. A scale-free

movement pattern was observed, which is considered optimal when foraging for heterogeneously located resources with little prior knowledge of distributional patterns [22]. Specifically, by collecting and analyzing the locations of people, from data of their mobile phone usage, it was found that human daily movement follows a truncated Lévy walk [23]. It was also shown that human daily trajectories exhibit a high degree of spatial regularity, i.e., each individual has a significant probability to return to a few much frequented locations, making their mobility highly predictable [12].

The individual activity patterns were characterized by the time distribution of visiting different places depending on activity category, which indicates a strong influence of urban contexts on people's activity participation and destination choices [24]. The location-based social network (LBSN) data were used to analyze the geo-temporal rhythms of users' checkins, informing the general consensus of user activity at a given time and place [13]. This can be used for recommendation or advertising application. On the field of urban computing/informatics, understanding the dynamics of the activities that take place in an urban area can significantly enhance functionalities such as resource and service allocation within a city [14]. By capturing urban dynamics to further provide innovative services for city inhabitants, Komninos *et al.* [15] discussed how this information can be used to guide visitors in the city or provide innovative services for city inhabitants using the cloud.

### B. Social Computing

Social computing is an area of computer science that is concerned with the intersection of social behavior and computational systems. A number of social computing technologies and methods are motivating an intense progress of online behavior analysis. Luca [19] studied the impact of consumer reviews on the changes in revenue, and found that a one-star increase in Yelp rating leads to a 5%–9% increase in revenue. Such effect is driven by independent restaurants, and ratings do not affect restaurants with chain affiliation. Luca and Zervas [20] further shed light on the economic incentives behind a decision of a business to leave fake reviews: independent restaurants are more likely to leave positive fake reviews for themselves, and fake negative reviews are more likely to occur when a business has an independent competitor. Chang and Sun [17] developed models that predict where users will check in, how their friends will respond, and whether their actions infer friendship. Hu *et al.* [18] studied business rating prediction, and observed that there exists weak positive correlation between businesses' ratings and its neighbors' ratings, regardless of the categories of the businesses. This finding enables better handling of a cold-start situation, for rating prediction of newly established businesses based on both their geographical neighbors and business categories.

Moreover, Mejova *et al.* [16] paid attention to health-related topics with people's culinary experiences. Their study would be instrumental to a successful understanding of public health using the vast amount of data in social media. Akbari *et al.* [25] focused on personal wellness events by harvesting social media data toward a healthier lifestyle. They proposed to automatically extract wellness events from users' published social contents for inferring individuals' lifestyle and wellness. Specifically, some works were concerned with health topic, such as investigating the relationship between fast food and chain restaurants and obesity [26]; studying the dietary and health activities for both native and foreign populations using Instagram posts [27]; building model to predict county-wide obesity and diabetes statistics based on a combination of demographic variables and food names mentioned on Twitter [28]; using machine learning model to predict repeatedly being over or under self-set daily calories goals [29]; and proposing multisource individual user profile learning framework to infer personal wellness attributes [30].

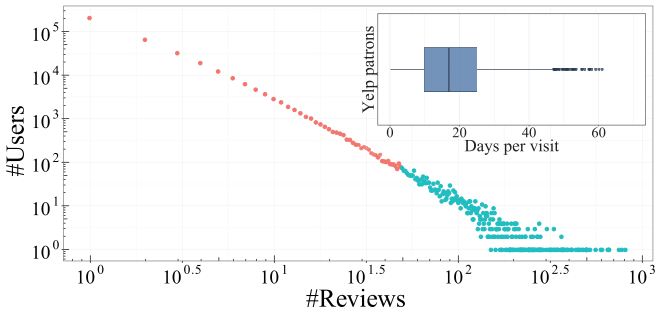



Fig. 1. Distribution of the counts of Yelp patrons who wrote certain number of reviews for restaurants. Inset: distribution of the average time interval between subsequent venue visits about Yelp patrons in green part.

## III. Methodology

This paper is based on the data from Yelp Data Set Challenge (Round 8) that is publicly available.[1] It contains 25 071 restaurants across 242 cities. Each restaurant is tagged with one or more cuisine tags, indicating the kinds of food it provides, which we use to characterize people's taste preferences. There are also latitude and longitude numbers, used to determine a restaurant's precise location. There is other information also, such as price level, star rating, review count, and so on.

Here, we use the review data for temporal information to define a person's foraging pattern. Note that a patron may visit a restaurant many times but review it once. In this context, the restaurant in a review can be considered as a point of interest that the patron thinks worth mentioning. In the following, when we say that a patron visited a restaurant, place, or cluster for simplicity, we just mean the patron reviewed the restaurant or a restaurant in the cluster, since we cannot have the real visit data. This data set also contains personal information for 388 612 patrons who wrote at least one restaurant review on the Yelp website. The distribution of the count of patrons writing certain number of reviews is shown in Fig. 1. Each patron, additionally, has a list of friends and followers, and we use the latter to measure their social status in the community. In addition, we have the date a patron joined Yelp, the time they wrote each restaurant review, and their rating. Yelp users can vote a review as *useful*, *funny*, and *cool*, which can be used to measure the quality of the review.

[1] https://www.yelp.com/data_set_challenge

In this paper, we focus on a sample of 2287 most active Yelp reviewers who submitted at least 50 reviews each (the green part in Fig. 1), whose reviews cover 235 out of total 242 cities. We do this in order to get statistically meaningful results. As the inset of Fig. 1 shows, most of these patrons take around 20 days per subsequent venue visit, making our conclusions about the cuisine and geographical foraging patterns less biased.

### *A. Geographical Foraging Network*

For those restaurants close to each other, they induce very similar foraging cost on a patron. Thus, it is reasonable to cluster restaurants initially based on their locations and then study the foraging pattern between these geographical restaurant clusters.

Considering there are a large number of geographical regions in Yelp data set, and the place density is the driving force of urban movement [31], we would like to adopt the density-based spatial clustering of applications with noise (DBSCAN) [32] to cluster the restaurants in each city [33]. Note that the density in [31] is defined as the number of places per square kilometer km$^2$ averaged across the grid. Noulas *et al.* [31] found that the average distance of human movements is inversely proportional to the city's density. And thus, the density there could be considered as a distance metric. In DBSCAN, however, the density is only used to design the clustering approach. We choose DBSCAN due to its following advantages: compared with grid method [34], it can find arbitrarily shaped clusters; compared with K-means [35], it does not require specifying the number of clusters *a priori*. Thus, it is suitable for our task. When applying DBSCAN, we specify that each restaurant must belong to a single cluster, and adopt a method to automatically determine the clustering radius for each city [36]. Finally, we get a total of 3049 clusters with the mean size equal to 8.2, i.e., on average, there are about eight restaurants in each cluster.

We establish a GFN for each patron, as the weighted directed graph whose nodes are the clusters $c_i$ of restaurants that the patron has ever visited. In the graph, $c_i$ and $c_j$ are connected by an edge if they have been visited in successive times, with the edge pointing from the earlier visited cluster to the later one. We call those edges geographical foraging links. The weight $n_{ij}$ of the directed link pointing from cluster $c_i$ to $c_j$ is the total number of times that the person visited first $c_i$ and then $c_j$, in succession.

Fig. 2, for instance, shows a clustering of the restaurants in North Las Vegas using DBSCAN. The GFN of a patron who visited 24 restaurants in this city (grouped into 10 clusters) is shown in Fig. 3. That patron visited these restaurants a total of 41 times from June 7, 2011 to December 29, 2015. Seventy-five percent of the visits are to four main clusters: 12 times in Cluster #2 (take up 29%), nine times in Cluster #9 (22%), and five times each in Cluster #8 (12%) and #10 (12%).

To measure the complexity of a person's geographical foraging pattern, we introduce three types of network entropy measures on her GFN [12], [37]. The first, and simplest, is the uniform *random entropy*

$$E_R = \log_2 N \tag{1}$$

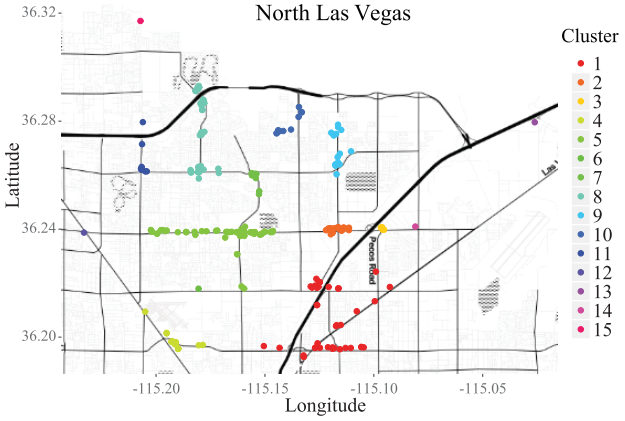



Fig. 2. Clustering of the restaurants in North Las Vegas with DBSCAN. There are 15 clusters, represented by different colors.

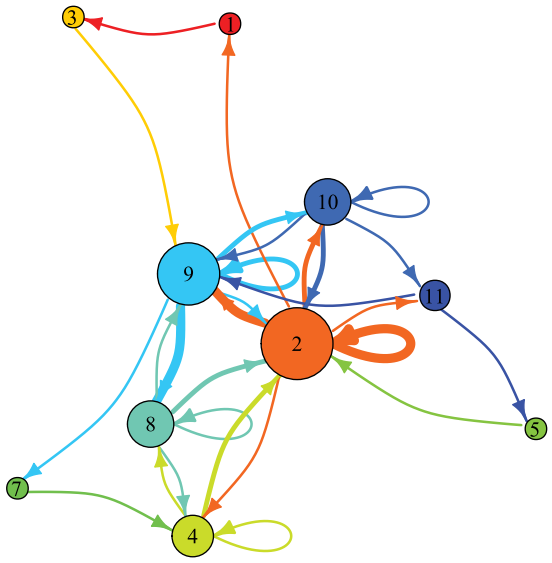



Fig. 3. GFN for a patron who visited 24 restaurants (belonging to 10 clusters) in North Las Vegas. The number attached to each node corresponds to a cluster in Fig. 2. The node size is proportional to the *logarithm* of frequency that the person visited the cluster, and the link width is proportional to the weight of the link.

which is used to measure the uncertainty that a patron randomly select one of $N$ restaurant clusters of restaurant's visit.

In reality, however, people do not choose uniformly at random. They are more likely to visit some clusters more frequently than others, as shown in Fig. 3. One reason is that restaurants themselves are never distributed uniformly in geography: we can always find more restaurants in places with higher traffic, near commercial centers, or close to universities, since these places have relatively large flux of people [3]. People's mobility patterns are also not uniform [12], as they prefer to visit places near their homes or workplaces. Therefore, the next restaurant chosen by a person will have a lower uncertainty than choosing by chance. To model the entropy in this case, suppose a patron has visited cluster $c_i$ for $n_i$ times, then the probability that the patron will visit this cluster next time is

$$p_i = \frac{n_i}{\sum_{j=1}^{N} n_j} \tag{2}$$

which forms the basis for measuring the *uncorrelated entropy*

$$E_U = -\sum_{i=1}^{N} p_i \log_2 p_i. \tag{3}$$

But the GFN provides even more information about foraging behavior, likely leading to better prediction. Namely, if $n_{ij}$ is the weight of the link from cluster $c_i$ to $c_j$, then, the conditional probability that a person visits $c_i$ before $c_j$ can be estimated by

$$p(j|i) = \frac{n_{ij}}{\sum_{k\in\pi_i} n_{ik}} \tag{4}$$

where $\pi_i$ is the outgoing neighbor set of cluster $c_i$ in the GFN. Then, considering foraging behavior as a Markov process [38], we can define the *Markov entropy* as

$$E_M = -\sum_{i=1}^{N}\left[p_i \sum_{j\in\pi_i} p(j|i)\log_2 p(j|i)\right]. \tag{5}$$

### B. Taste Foraging Network

In addition to geography, we want to study the preferences and switching patterns of patrons between different cuisines over time. To do that, we define a TFN for each patron, represented by a weighted directed graph whose nodes are cuisines $t_i$ that the patron has ever tasted. Cuisines $t_i$ and $t_j$ are connected by an edge if they have been tried in successive times, with the edge pointing from the earlier to the later. We call those edges taste-foraging links. Each link in a TFN also has a weight, signifying the total number of times that the patron has switched from one cuisine to the other successively. We then calculate the three entropy measures of TFN, similar to GFN.

However, calculating a TFN from our data may not be straightforward, as a restaurant may have more than one cuisine tag. We introduce a Gaussian denoise algorithm (GDA) [39] in Algorithm 1 to estimate the cuisine that a patron is most likely to have had in their visit to the restaurant.

Fig. 4 shows the TFN for a person who tried 12 different cuisines. The person went to 39 restaurants a total of 55 times from May 3, 2011 to December 5, 2015. Seventy-six percent of the visits associate with only five cuisines: 12 times for *Italian* food (take up 22%), 11 times for *New American* food (20%), seven times for *Steak* (13%), six times for *French* food (11%), and five times for *Pizza* (10%). Based on the TFN for each patron, as above for GFNs, we can define random, uncorrelated, and Markov entropies, which will be used to characterize the cuisine foraging patterns for that patron.

### C. Taste Similarity Network

In addition, to easily compare cuisines across restaurants, we construct a TSN based on restaurants' cuisine tags. In the TSN, each node represents a cuisine tag $t_i$, and two nodes $t_i$ and $t_j$ are linked if they are ever attached to the same restaurant. We would expect that the tags attached to the same restaurants would tend to be closer to each other than those attached to different restaurants. To define similarity

**Algorithm 1** GDA for Cuisines

**Input:** A list of restaurants that a patron visited in sequence, denoted by $R_1, R_2, \ldots, R_T$, where restaurants may repeat. Restaurant $R_i$ has a set of cuisine tags $\vartheta_i = \{t_1^i, t_2^i, \ldots, t_{k_i}^i\}$, which all get the same score $\mu^i$ that the patron has given the restaurant. Denote by $\vartheta = \bigcup_{i=1}^{T} \vartheta_i$ the full cuisine list of the patron.

**Output:** A list of cuisines that the patron will most likely choose from the restaurants this patron visited at different times, denoted by $t_1, t_2, \ldots, t_T$.

**Algorithm:** For each cuisine $\zeta_f \in \vartheta$, estimate its Gaussian distribution $N(\hat{\mu}_f, \hat{\sigma}_f^2)$ of scores, and do:

```
for i = 1 : T do
    for j = 1 : k_i do
        find a cuisine ζ_f ∈ ϑ satisfying ζ_f ≡ t_j^i, then
        calculate ε_j^i = |μ^i − μ̂_f|/σ̂_f
    end for
    if ε_g^i = min{ε_j^i | ∀t_j^i ∈ ϑ_i}
        set t_i = t_g^i
    end if
end for
Return list of cuisines t_1, t_2, ..., t_T.
```

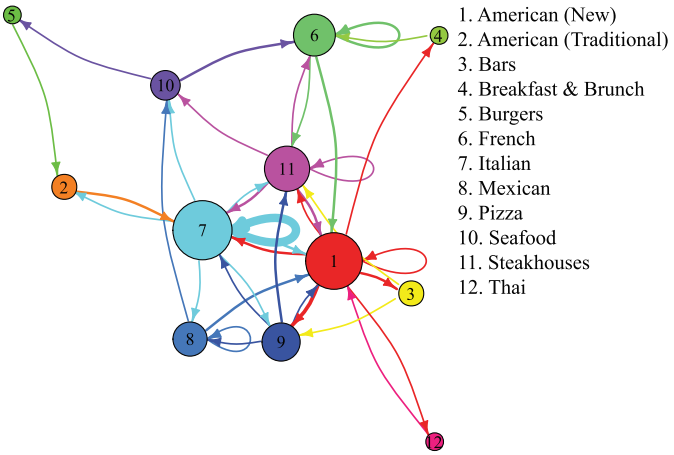



Fig. 4. TFN for a patron who tried 12 cuisines in 39 restaurants. The node size is proportional to the frequency with which that cuisine was tried, and the link width is proportional to the weight of the link.

between two cuisines $t_i$ and $t_j$, or the weight $w_{ij}$ on the link between them, we use *cosine similarity*, which is quite popular in the area of network science and is adopted in many applications, such as link prediction [40] and recommendation system design [41]. If we denote $G_i$ as the set of restaurants having cuisines $t_i$, then the weight of the link between cuisines $t_i$ and $t_j$ is defined as ($|\cdot|$ is set cardinality)

$$w_{ij} = \frac{|G_i \cap G_j|}{\sqrt{|G_i|}\sqrt{|G_j|}}. \tag{6}$$

We integrate the TFNs of all patrons as one by summing all the weights of directed links between the same pairs of nodes, and the overall TFN and TSN can be considered as layers in a two-layer network, as they are over the same set of cuisines. In this layered network, each pair of cuisines can thus

be either connected by both taste-foraging and taste-similarity links, only one of them, or disconnected. Given the overall TFN and TSN, we next define two congruences between them to measure the extent of patrons' cuisine shifting in agreement with the cuisine similarity among restaurants. Similar metrics were also used in software engineering [37], [42], [43]. Let $W_\alpha$ be the set of weights of the TFN edges that are also in the TSN, regardless of edge directions, $W_\beta$ be the set of weights of the corresponding TSN edges, and $W_\gamma$ be the set of weights of TFN edges that are not in the TSN. Then, the first congruence $C_1$ of TFN on TSN is defined as the fraction of the average weights on edges that are in agreement in both networks

$$C_1 = \frac{\langle W_\alpha \rangle}{\langle W_\alpha \rangle + \langle W_\gamma \rangle}. \tag{7}$$

And the second congruence $C_2$ is the *Spearman* correlation between $W_\alpha$ and $W_\beta$. Both congruences fall into the range $[0, 1]$, and have higher values when the networks are in better agreement.

TABLE I
SUMMARY OF ABBREVIATIONS

| Abbreviation | Definition |
|---|---|
| GFN | Geographical Foraging Network |
| TFN | Taste Foraging Network |
| TSN | Taste Similarity Network |
| GSD | Geographical Searching Distance |
| TSS | Taste Searching Similarity |

TABLE II
SUMMARY OF DESIGNATIONS

| Designation | Definition |
|---|---|
| $N_S$ | The number of restaurants |
| $N_R$ | The number of reviews |
| $N_V$ | The number of votes |
| $PL$ | The average price levels across reviews |
| $E_R(G)$ | Random entropy of GFN |
| $P(G)$ | Distributional complexity of GFN |
| $Q(G)$ | Structural complexity of GFN |
| $E_R(T)$ | Random entropy of TFN |
| $P(T)$ | Distributional complexity of TFN |
| $Q(T)$ | Structural complexity of TFN |

### *D. MLR Model and Orthogonal Decomposition*

To model the social status of a person in Yelp against the above measures of complexity of foraging patterns in both geography and cuisine, we use MLR [44]. In MLR model, explanatory variables should be linearly independent; however, the three types of entropy are all strongly correlated with the network size $N$ and, thus, are not suitable to be considered together as predictors in the same MLR model. We then use *orthogonal decomposition* [45] to first transform them by eliminating their dependence on $N$.

Considering all Yelp reviewers together, we decorrelate the vectors $\mathbf{E_U}$ and $\mathbf{E_M}$ from $\mathbf{E_R}$ to get the network-size uncorrelated complexities $P$ and $Q$, respectively. That is, the inner products $\langle \mathbf{E_R}, \mathbf{P} \rangle$, $\langle \mathbf{E_R}, \mathbf{Q} \rangle$, and $\langle \mathbf{P}, \mathbf{Q} \rangle$ all equal to zero. We centralize these vectors first, and obtain $P$ and $Q$ from the following equations [37]:

$$\mathbf{P} = \mathbf{E_U} - \frac{\langle \mathbf{E_U}, \mathbf{E_R} \rangle}{\|\mathbf{E_R}\|^2} \mathbf{E_R} \tag{8}$$

$$\mathbf{Q} = \mathbf{E_M} - \frac{\langle \mathbf{E_M}, \mathbf{E_R} \rangle}{\|\mathbf{E_R}\|^2} \mathbf{E_R} - \frac{\langle \mathbf{E_M}, \mathbf{P} \rangle}{\|\mathbf{P}\|^2} \mathbf{P}. \tag{9}$$

We refer to $P$ as the *distributional complexity*, capturing the global heterogeneity of the foraging distribution in geography (GFN) or cuisine (TFN); it is maximized when a person has visited all restaurant clusters or has tried all cuisines uniformly, and is minimized if they mainly visited or tasted just one of them. We refer to $Q$ as the *structural complexity*, which in turn, captures the local heterogeneity; it is minimized if the patron's next location or, respectively, cuisine can be exactly predicted given her current one.

## IV. RESULTS AND DISCUSSION

For convenience, we summarize all the abbreviations and designations in Tables I and II, respectively. Here, we first study how Yelp restaurant patrons expand their geographical and cuisine spaces; then, we focus on the regularity of their foraging patterns to see to what extent such behaviors can be predicted; finally, we establish an MLR model for individual social status based on their foraging patterns.

### *A. Expansion of Foraging Over Geography and Cuisine*

In order to reduce the costs of foraging, people tend to search for food nearby. With time, however, they may get tired with the same restaurants and cuisines, and opt to seek new ones, within a relatively larger distance. Therefore, we may find a significant pattern that people expand their foraging domains over time, with respect to both geography and cuisine.

For each patron, we denote by $C = \{c_1, c_2, \ldots, c_n\}$ the set of clusters of restaurants in the geography they visited, ordered by their first visited time, e.g., the patron visited $c_j$ prior to visiting $c_i$ if $j < i$. Supposing the patron had visited $i-1$ clusters, and searched for the $i$th cluster for the first time, we then define the *Geographical Searching Distance* (GSD), $d_i$, as the average distance between cluster $c_i$ and all the other visited clusters $c_j$ $(j < i)$. To define the distance between two clusters, we first get the center of each cluster by averaging the locations of all the restaurants in it and then use the distance between the centers of the clusters. If a patron visited the clusters randomly, we will expect a nearly constant GSD for different clusters. However, we find that $d_i$ is an increasing function of the index $i$ in reality, considering only the 1946 patrons that have visited at least 10 clusters, as shown in Fig. 5(a), and there is no significant correlation between the two curves for real data and random case. This indicates that people are indeed more likely to steadily expand their foraging domains in geography over time, answering RQ#1.

Next, we denote by $T = \{t_1, t_2, \ldots, t_m\}$ the set of cuisines that a patron has tried. Suppose a patron has tried $i-1$ cuisines, and searches for the next, $i$th cuisine to try for the first time, then the *taste searching similarity* (TSS) $s_i$ at that time is defined as the average similarity between cuisine $t_i$

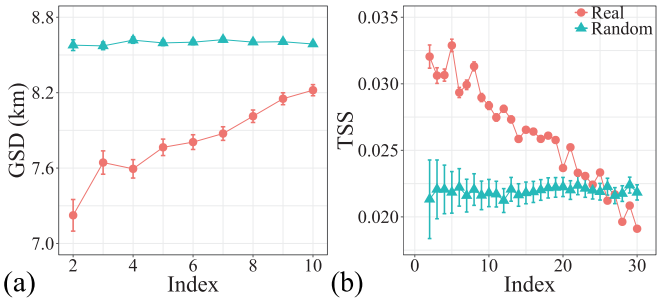



Fig. 5. (a) GSD $d_i$ as a function of index $i$ and (b) TSS $s_i$ as a function of index $i$, for empirical data and random cases, by considering all patrons together. Here, only the mean values and error bars are presented.

and all the other tried cuisines $t_j$ ($j < i$), calculated by (6). Again, we find that people tend to steadily expand their cuisine foraging domains also, from the familiar to the novel, rather than at random, i.e., $s_i$ is a decreasing function of the index $i$, as shown in Fig. 5(b), and there is again, no significant correlation between the two curves for real data and random case, for another answer to RQ#1. In this case, a total of 722 patrons having had at least 30 cuisines are considered.

Note that the results in Fig. 5(a) and (b) are across all cities. When we consider the cities in each state independently, we get the exactly same trend, but with slightly different slopes for different states. Interestingly, for the seven states covering 2272 Yelp patrons in consideration (2287 in total), including Arizona, Nevada, Pennsylvania, North Carolina, Wisconsin, Quebec (Canada), and Edinburgh (Scotland), we find a significantly negative correlation (with Spearman correlation coefficient $-0.893$ and the significance $p = 0.012$) between the slope for geographical expansion and the restaurant density (the number of restaurants divided by the area of state), which is consistent with the results in [31].

To answer whether people with narrower tastes tend to visit more places, i.e., more geographical clusters of restaurants, we establish an MLR model for the number of restaurant clusters in terms of the number of cuisines for the 2287 most active Yelp reviewers, while controlling the number of restaurants, denoted by $N_S$. We logged these values first in order to stabilize the variance and improve the model fit. Note that when we logged the number of cuisines for a patron, we get the random entropy of the TFN of that patron, i.e., $E_R(T)$. The model is shown in Table III, where we can see that the number of restaurant clusters and the number of cuisines are negatively correlated with each other, indicating that people with less cuisine variability are indeed likely to visit larger number of restaurant clusters, positively answering RQ#2. Moreover, we also present some metrics for the model performance in the last row of Table III, i.e., R-squared equals to 0.456, adjusted R-squared equals to 0.455, relative squared error (RSE) equals to 0.544, and root-mean-square error (RMSE) equals to 0.583, indicating a relatively strong linear relationship of the model for a social system.

As case studies, we chose two reviewers of restaurants in Pittsburgh and Las Vegas, denoted by A and B, respectively. A visited 18 restaurant clusters with 49 cuisine tags. B visited far more restaurant clusters, 71 in total, which have fewer cuisine tags, 26. It seems that A likes to try a variety of foods,

TABLE III
MLR MODEL FOR THE NUMBER OF RESTAURANT CLUSTERS AGAINST THE NUMBER OF CUISINES AND THE NUMBER OF RESTAURANTS

| | Estimate | Std. Error | $t$ value | Pr($>\|t\|$) |
|---|---|---|---|---|
| (Intercept) | 0.701 | 0.143 | 4.92 | 9.4e-07 |
| $E_R(T)$ | -0.459 | 0.061 | -7.58 | 4.8e-14 |
| $\log_2 N_S$ | 1.033 | 0.037 | 27.8 | <2.2e-16 |
| R-squared: 0.456; Adjusted: 0.455; RSE: 0.544; RMSE: 0.583 | | | | |

but is sensitive to the "cost" involved in finding them, and the reviews confirm that: *I have found yet another reason to hang around on Western Ave …… I often times just hurried passed storefront after storefront never taking a second glance at the many great establishments that line Western Ave.* ***Well now that Western Ave is within walking distance of my new home, I've been slowing down and exploring what the avenue has to offer*** *[posted on Sep. 29, 2012].* Another one: ***Nothing to write back to mom about. The only reason I eat here is because it located right downstair from my job. I eat at Damon's when I am desperate and really don't want to walk any where. I think their food is over priced a sald with bacon cost $9.00 that's a rip off*** *[posted on Feb. 2, 2012].*

On the other hand, B seems to care more about particular kinds of food, and pays less attention to the "cost" to find them. The reviews confirm this: *I hope more people find out this place is here because it's tucked away with other businesses. When I saw the sign after walking in that they had Boar's Head meats and cheeses I had a good feeling …… That meant of course the pastrami but it also meant* 3 *pepper cheese, pickles, peperoncinis, red onions, and seasoning. I can't even believe all that worked as well as it did together but damn it really did ……* ***After I ate I was full and satisfied. The only downside is that this location isn't on the side of town I live and that right now they are only open for*** 5 ***hours (10:30 AM til 3:30 PM) a day mon - fri. I do work in the field so when I am over here I will stop by for lunch but i do hope at some point they open up on weekends*** *[posted on Feb. 19, 2015].* Another one: *The cheesesteaks here are almost good enough to say Pop's who? …… I recommend trying the cheesesteak before anything else.* ***I still need to try a pizza that is more my style with sausage and peppers*** *but my wife ordered a white pie that I gave into and tried a slice of. It wasn't bad, just not my personal favorite toppings for a pizza. The crust on it was good. The mozzarella sticks are made fresh and by hand to order. They have a good amount of gooey cheese and cooked perfectly …… I do wish they had better signage out front.* ***I had a bit of a time finding it and at this time anyway there is nothing on the building saying who they are like the neighboring business in the center. I noticed the specials on their window is the only reason I found it.*** *Solid place with great lunch specials [posted on Jul. 3, 2014].*

### B. Predictability of Human Foraging Patterns

Similar to the work [12], for each patron, we calculate the random entropy $E_R$, the uncorrelated entropy $E_U$, and the Markov entropy $E_M$ of their geographical foraging pattern and then derive the probability density functions (pdfs) of the three kinds of entropy across all 2287 patrons, as shown

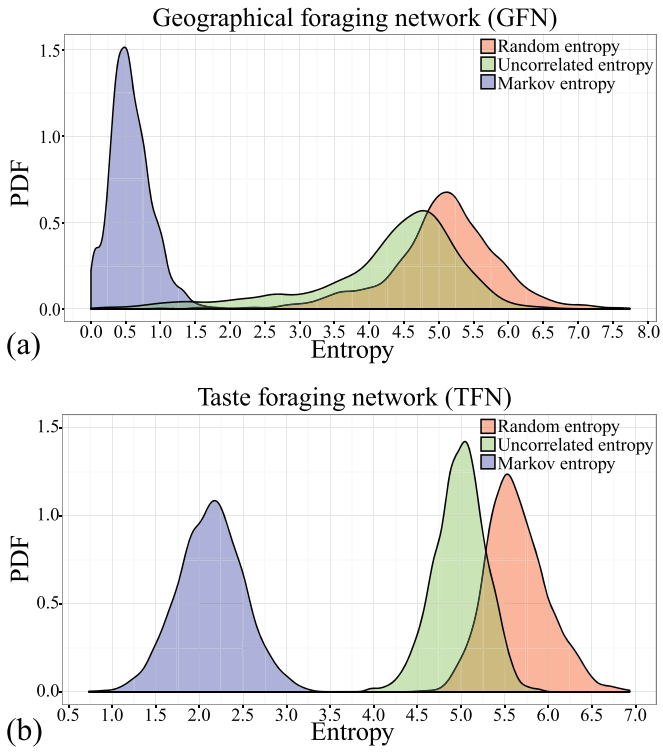



Fig. 6. PDFs of the random entropy $E_R$, the uncorrelated entropy $E_U$, and the Markov entropy $E_M$ of human foraging patterns, in (a) geography and (b) cuisine, for 2287 restaurant patrons.

in Fig. 6(a). We find that the pdf of $E_R$ peaks at 5.1, indicating that, on average, a patron may randomly visit $2^{E_R} \approx 34$ geographical restaurant clusters. More interestingly, the pdf of $E_M$ peaks at $E_M = 0.5$, suggesting that, in reality, the number of candidate clusters of restaurants that a patron will visit next time is only $2^{E_M} = 1.4$, fewer than two. This is strongly consistent with the study of general human mobility patterns [12].

Next, we study human cuisine foraging patterns. Using the same above-mentioned methods, we get the pdfs of $E_R$, $E_U$, and $E_M$ of the cuisine foraging patterns, as shown in Fig. 6(b). We find that the pdf of $E_R$ peaks at 5.5, meaning that, a patron may randomly choose among restaurants with $2^{E_R} \approx 45$ cuisines. However, the corresponding pdf of $E_M$ peaks at $E_M = 2.2$, i.e., the next cuisine an average person would choose will probably come from fewer than five candidate cuisines ($2^{E_M} = 4.6$). These results suggest that human foraging patterns, both in geography and cuisine, are of rather high regularity, answering RQ#3. Therefore, considering the finiteness and regularity of modern foraging behavior, utilizing people's restaurant visiting and cuisine choosing trajectories could help to design better recommendation system.

We also investigate whether people tend to switch between cuisines along taste-similarity links. We use congruence $C_1$ defined in (7). Note that to obtain a global perspective, we first integrate the TFNs of all people (by summing all the weights of directed links between the same pairs of nodes), and get the corresponding TSN. Then, we calculate the congruence between the overall TFN and TSN. We find that people indeed tend to switch cuisines along TSN links: the weights of TFN links between pairs of cuisines that also exist in TSN, $W_\alpha$, are on average about four times ($C_1 = 0.83$) larger than those of TFN-linked cuisines that are not linked in TSN, $W_\gamma$. To understand this tendency in more detail, we use the congruence $C_2$, defined as the Spearman correlation between the weights of TFN links between pairs of cuisines that are also connected in TSN, $W_\alpha$, and those of their corresponding TSN links, $W_\beta$. We find a significantly positive correlation $C_2 = 0.46$, indicating that people are more likely to switch between cuisines of higher similarity in successive times, positively answering RQ#3. Note that, here, we just use successive activities to establish TFN, without considering the time intervals between these activities. When we only keep those successive activities with time interval smaller than some time window, e.g., one day, one week, or one month, the above results almost keep the same. We thus think that our results are quite robust for various time windows.

### C. Foraging Patterns and Social Status

In many online communities, including Yelp, people rely on experts in particular areas, and mimic their behaviors so as to get relevant information. The popularity, or social status, of a person in this community thus can be measured by the number of their followers, which we expect is related to the foraging behavior.

We have already proposed three entropy metrics to measure the geographical foraging behavior. Since these metrics are significantly correlated with each other, we first use orthogonal decomposition to decompose their correlation and get three uncorrelated parts: the random entropy $E_R(G) = \log_2 N$, the distributional complexity $P(G)$, and the structural complexity $Q(G)$, given by (8) and (9), respectively. For the foraging behavior in cuisine, we can get similar uncorrelated metrics $E_R(T)$, $P(T)$, and $Q(T)$. We use them as well as the average price level across each patron's reviews, denoted by $PL$, of food to build an MLR model for the community popularity of a person. Such a model can help us understand to what extent the complexities of foraging behavior are associated with individual popularity, when controlled for other properties including the number of reviews and the number of votes, denoted by $N_R$ and $N_V$, respectively.

We logged the number of followers (popularity), the number of reviews, and the number of votes to stabilize the variance, and get seven independent variables with significance $p < 0.05$, as shown in Table IV. Since the number of variables is relatively large in this model, we check the magnitude of multicollinearity of the model by calculating the *variance inflation factors* (VIFs), and find that VIFs for all the variables are smaller than 3, indicating that the multicollinearity is low. Similarly, the metrics for the goodness of fit are presented in the last row of Table III.

We find that both the number of reviews and the number of votes are positively correlated with popularity, suggesting that, typically, a reviewer will get larger number of followers if they wrote more reviews of higher quality. Patrons reviewing higher priced restaurants also attract more followers. This may be because in many cases, a higher price level means better quality and service, which could be paid more attention.

TABLE IV

MLR MODEL FOR POPULARITY AGAINST THE RANDOM ENTROPY, DISTRIBUTIONAL AND STRUCTURAL COMPLEXITIES OF GFN AND TFN, AND OTHER PROPERTIES INCLUDING THE NUMBER OF REVIEWS, THE NUMBER OF VOTES, AND THE AVERAGE PRICE LEVEL OF FOOD

| | Estimate | Std. Error | $t$ value | Pr($>\|t\|$) |
|---|---|---|---|---|
| (Intercept) | -6.624 | 0.387 | -17.10 | <2.2e-16 |
| $\log_2 N_R$ | 0.109 | 0.028 | 3.90 | 9.7e-05 |
| $\log_2 N_V$ | 0.740 | 0.015 | 49.23 | <2.2e-16 |
| $PL$ | 0.405 | 0.080 | 5.09 | 3.9e-07 |
| $E_R(G)$ | -0.154 | 0.027 | -5.63 | 2.0e-08 |
| $P(G)$ | 0.132 | 0.051 | 2.59 | 0.0098 |
| $E_R(T)$ | 0.354 | 0.075 | 4.76 | 2.1e-06 |
| $Q(T)$ | 0.176 | 0.053 | 3.32 | 9.3e-04 |
| R-squared: 0.763; Adjusted: 0.762; RSE: 0.787; RMSE: 0.786 | | | | |

Or other way around: because the patrons have more followers, they thus are more likely to review the restaurants of higher prices. After controlling for other variables, the patrons who visited fewer geographical clusters of restaurants, with their reviews distributed in these clusters more uniformly, tend to have a larger number of followers, possibly due to their relatively narrower but stronger influence in local regions. On the other hand, patrons with a higher variety of cuisines are more likely to get higher popularity, i.e., both $E_R(T)$ and $Q(T)$ are positively correlated with the number of followers. This is reasonable since the patrons who taste a wider variety of food can provide more comprehensive comparisons between different cuisines and, thus, may attract more followers of diverse tastes. These results answer RQ#4.

## V. CONCLUSION

In this paper, we studied the modern food foraging patterns based on Yelp data set and made a dual contribution as follows.

First, in methodology, we studied human foraging patterns in two dimensions, i.e., geography and cuisine, based on patron activities on Yelp. We defined the similarity between different cuisines and established a TSN as our cuisine space. We used the sequential review data of patrons to establish the GFN and the TFN, and proposed three types of entropies to capture the complexity of patrons' foraging behavior. We transformed the entropy metrics by using orthogonal decomposition to get uncorrelated terms which, together with other properties, were used to build an MLR model for reviewer popularity.

Second, based on the methods proposed, we found that people tend to search for food following basic optimal rules: from the near to the distant in geography and from the familiar to the novel in cuisine; their foraging behaviors thus are far away from random, and are quite predictable in both geography and cuisine. Moreover, Yelp reviewers writing about higher priced food, with a wider range of cuisines but smaller number of foraging locations tend to gain higher popularity in the community, indicating that such behaviors are more likely to be followed and mimicked by others.

In the future, we will focus on integrating these foraging patterns in both geography and cuisine, combined with machine learning algorithms, to design a comprehensive, bespoke temporal recommendation system. Moreover, we will relate the foraging patterns of users with their health and wellness attributes, leading to healthier behaviors.

## REFERENCES


[1] E. Danchin, L.-A. Giraldeau, and F. Cézilly, *Behavioural Ecology*. New York, NY, USA: Oxford Univ. Press, 2008.

[2] G. H. Pyke, "Optimal foraging theory: A critical review," *Annu. Rev. Ecol. Syst.*, vol. 15, no. 1, pp. 523–575, 1984.

[3] G.-H. Tzeng, M.-H. Teng, J.-J. Chen, and S. Opricovic, "Multicriteria selection for a restaurant location in Taipei," *Int. J. Hospitality Manage.*, vol. 21, no. 2, pp. 171–187, 2002.

[4] H. R. Moskowitz, "Taste and food technology: Acceptability, aesthetics, and preference," in *Handbook of Perception*. Orlando, FL, USA: Academic, 1978, pp. 157–194.

[5] G. B. Colombo, M. J. Chorley, V. Tanasescu, S. M. Allen, C. B. Jones, and R. M. Whitaker, "Will you like this place? A tag-based place representation approach," in *Proc. IEEE Int. Conf. Pervasive Comput. Commun. Workshops (PERCOM Workshops)*, Mar. 2013, pp. 224–229.

[6] N. E. Raine and L. Chittka, "The correlation of learning speed and natural foraging success in bumble-bees," *Proc. Roy. Soc. London B, Biol. Sci.*, vol. 275, no. 1636, pp. 803–808, 2008.

[7] Q. Xuan, H. Fang, C. Fu, and V. Filkov, "Temporal motifs reveal collaboration patterns in online task-oriented networks," *Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top.*, vol. 91, no. 5, p. 052813, 2015.

[8] H. Kwak, C. Lee, H. Park, and S. Moon, "What is Twitter, a social network or a news media?" in *Proc. 19th Int. Conf. World Wide Web*, 2010, pp. 591–600.

[9] S. A. Myers, A. Sharma, P. Gupta, and J. Lin, "Information network or social network?: The structure of the Twitter follow graph," in *Proc. 23rd Int. Conf. World Wide Web*, 2014, pp. 493–498.

[10] S. Wasserman and K. Faust, *Social Network Analysis: Methods and Applications*, vol. 8. Cambridge, U.K.: Cambridge Univ. Press, 1994.

[11] D. Knoke and S. Yang, *Social Network Analysis*, vol. 154. Newbury Park, CA, USA: Sage, 2008.

[12] C. Song, Z. Qu, N. Blumm, and A.-L. Barabási, "Limits of predictability in human mobility," *Science*, vol. 327, no. 5968, pp. 1018–1021, 2010.

[13] A. Noulas, S. Scellato, C. Mascolo, and M. Pontil, "An empirical study of geographic user activity patterns in foursquare," in *Proc. ICwSM*, vol. 11. Jul. 2011, pp. 570–573.

[14] K. Zhang, Q. Jin, K. Pelechrinis, and T. Lappas, "On the importance of temporal dynamics in modeling urban activity," in *Proc. 2nd ACM SIGKDD Int. Workshop Urban Computing*, 2013, p. 7.

[15] A. Komninos, V. Stefanis, A. Plessas, and J. Besharat, "Capturing urban dynamics with scarce check-in data," *IEEE Pervasive Comput.*, vol. 12, no. 4, pp. 20–28, Oct. 2013.

[16] Y. Mejova, S. Abbar, and H. Haddadi, "Fetishizing food in digital age:#FoodPorn around the world," in *Proc. ICWSM*, 2016, pp. 250–258.

[17] J. Chang and E. Sun, "Location 3: How users share and respond to location-based data on social networking sites," in *Proc. 5th Int. AAAI Conf. Weblogs Social Media*, 2011, pp. 74–80.

[18] L. Hu, A. Sun, and Y. Liu, "Your neighbors affect your ratings: On geographical neighborhood influence to rating prediction," in *Proc. 37th Int. ACM SIGIR Conf. Res. Develop. Inf. Retr.*, 2014, pp. 345–354.

[19] M. Luca, "Reviews, reputation, and revenue: The case of yelp.com," Harvard Bus. School, Boston, MA, USA, Working Paper 12-016, 2016.

[20] M. Luca and G. Zervas, "Fake it till you make it: Reputation, competition, and yelp review fraud," *Manage. Sci.*, vol. 62, no. 12, pp. 3412–3427, 2016.

[21] D. A. Raichlen, B. M. Wood, A. D. Gordon, A. Z. Mabulla, F. W. Marlowe, and H. Pontzer, "Evidence of Lévy walk foraging patterns in human hunter–gatherers," *Proc. Nat. Acad. Sci. USA*, vol. 111, no. 2, pp. 728–733, 2014.

[22] G. M. Viswanathan, M. G. Da Luz, E. P. Raposo, and H. E. Stanley, *The Physics of Foraging: An Introduction to Random Searches and Biological Encounters*. Cambridge, U.K.: Cambridge Univ. Press, 2011.

[23] M. C. González, C. A. Hidalgo, and A.-L. Barabási, "Understanding individual human mobility patterns," *Nature*, vol. 453, no. 7196, pp. 779–782, 2008.

[24] S. Hasan, X. Zhan, and S. V. Ukkusuri, "Understanding urban human activity and mobility patterns using large-scale location-based data from online social media," in *Proc. 2nd ACM SIGKDD Int. Workshop Urban Comput.*, 2013, p. 6.

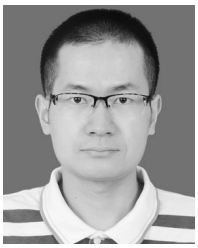
[25] M. Akbari, X. Hu, L. Nie, and T.-S. Chua, "From tweets to wellness: Wellness event detection from Twitter streams," in *Proc. AAAI*, 2016, pp. 87–93.

[26] Y. Mejova, H. Haddadi, A. Noulas, and I. Weber, "#FoodPorn: Obesity patterns in culinary interactions," in *Proc. 5th Int. Conf. Digit. Health*, 2015, pp. 51–58.

[27] Y. Mejova, H. Haddadi, S. Abbar, A. Ghahghaei, and I. Weber, "Dietary habits of an expat nation: Case of qatar," in *Proc. Int. Conf. Healthcare Informat. (ICHI)*, Oct. 2015, pp. 57–62.
[28] S. Abbar, Y. Mejova, and I. Weber, "You tweet what you eat: Studying food consumption through Twitter," in *Proc. 33rd Annu. ACM Conf. Human Factors Comput. Syst.*, 2015, pp. 3197–3206.
[29] I. Weber and P. Achananuparp, "Insights from machine-learned diet success prediction," in *Proc. Pacific Symp. Biocomput.*, 2016, pp. 540–551.
[30] A. Farseev and T.-S. Chua, "TweetFit: Fusing multiple social media and sensor data for wellness profile learning," in *Proc. AAAI*, 2017, pp. 95–101.
[31] A. Noulas, S. Scellato, R. Lambiotte, M. Pontil, and C. Mascolo, "A tale of many cities: Universal patterns in human urban mobility," *PLoS ONE*, vol. 7, no. 5, p. e37027, 2012.
[32] M. Ester, H.-P. Kriegel, J. Sander, and X. Xu, "A density-based algorithm for discovering clusters in large spatial databases with noise," in *Proc. 2rd Int. Conf. Knowl. Discovery Data Mining*, 1996, pp. 226–231.
[33] F. Zhang, N. J. Yuan, K. Zheng, D. Lian, X. Xie, and Y. Rui, "Exploiting dining preference for restaurant recommendation," in *Proc. 25th Int. Conf. World Wide Web*, 2016, pp. 725–735.
[34] R. Cheng, J. Pang, and Y. Zhang, "Inferring friendship from check-in data of location-based social networks," in *Proc. IEEE/ACM Int. Conf. Adv. Social Netw. Anal. Mining*, 2015, pp. 1284–1291.
[35] A. K. Jain, M. N. Murty, and P. J. Flynn, "Data clustering: A review," *ACM Comput. Surv.*, vol. 31, no. 3, pp. 264–323, Sep. 1999.
[36] M. N. Gaonkar and K. Sawant, "AutoEpsDBSCAN: DBSCAN with Eps automatic for large dataset," *Int. J. Adv. Comput. Theory Eng.*, vol. 2, no. 2, pp. 11–16, 2013.
[37] Q. Xuan, A. Okano, P. Devanbu, and V. Filkov, "Focus-shifting patterns of OSS developers and their congruence with call graphs," in *Proc. 22nd ACM SIGSOFT Int. Symp. Found. Softw. Eng.*, 2014, pp. 401–412.
[38] C. W. Gardiner, *Handbook of Stochastic Methods*, vol. 3. Berlin, Germany: Springer, 1985.
[39] P. Moulin and J. Liu, "Analysis of multiresolution image denoising schemes using generalized Gaussian and complexity priors," *IEEE Trans. Inf. Theory*, vol. 45, no. 3, pp. 909–919, Apr. 1999.
[40] D. Wang, D. Pedreschi, C. Song, F. Giannotti, and A.-L. Barabási, "Human mobility, social ties, and link prediction," in *Proc. 17th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2011, pp. 1100–1108.
[41] S. Chang and A. Pal, "Routing questions for collaborative answering in community question answering," in *Proc. IEEE/ACM Int. Conf. Adv. Social Netw. Anal. Mining*, 2013, pp. 494–501.
[42] M. Cataldo, P. A. Wagstrom, J. D. Herbsleb, and K. M. Carley, "Identification of coordination requirements: Implications for the design of collaboration and awareness tools," in *Proc. 20th Anniversary Conf. Comput. Supported Cooperative Work*, 2006, pp. 353–362.
[43] M. Cataldo and J. D. Herbsleb, "Coordination breakdowns and their impact on development productivity and software failures," *IEEE Trans. Softw. Eng.*, vol. 39, no. 3, pp. 343–360, Mar. 2013.
[44] N. R. Draper and H. Smith, *Applied Regression Analysis*, 2nd ed. New York, NY, USA: Wiley, 1981.
[45] D. Cherney, T. Denton, and A. Waldron, *Linear Algebra*. Davis, CA, USA: Univ. California Davis, 2013.

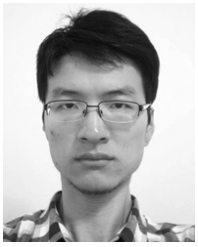
**Qi Xuan** (M'18) received the B.S. and Ph.D. degrees in control theory and engineering from Zhejiang University, Hangzhou, China, in 2003 and 2008, respectively.

He was a Post-Doctoral Researcher with the Department of Information Science and Electronic Engineering, Zhejiang University, from 2008 to 2010, and a Research Assistant with the Department of Electronic Engineering, City University of Hong Kong, Hong Kong, in 2010 and 2017. From 2012 to 2014, he was a Post-Doctoral Researcher with the Department of Computer Science, University of California at Davis, Davis, CA, USA. He is currently a Professor with the College of Information Engineering, Zhejiang University of Technology, Hangzhou. His current research interests include network-based algorithm design, social network data mining, social synchronization and consensus, reaction-diffusion network dynamics, machine learning, and computer vision.



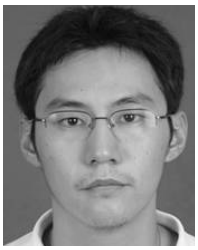
**Mingming Zhou** received the B.S. degree in automation from Jiangnan University, Wuxi, China, in 2015. He is currently pursuing the M.S. degree in control theory and engineering with the College of Information Engineering, Zhejiang University of Technology, Hangzhou, China.

His current research interests include social network analysis, empirical software engineering, and machine learning.



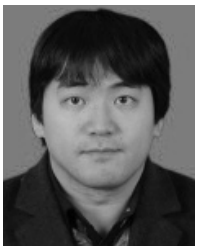
**Zhi-Yuan Zhang** received the B.S. degree in electrical and electronic engineering from the University of Nottingham, Ningbo, China, in 2016. He is currently pursuing the M.S. degree with the Department of Computer Science, University of California at Davis, Davis, CA, USA.

His current research interests include software engineering, network science, and social network data mining, especially focus on mining open source software projects.



**Chenbo Fu** received the B.S. degree in physics from the Zhejiang University of Technology, Hangzhou, China, in 2007, and the M.S. and Ph.D. degrees in physics from Zhejiang University, Hangzhou, in 2009 and 2013, respectively.

In 2014, he was a Research Assistant with the Department of Computer Science, University of California at Davis, Davis, CA, USA. He was a Post-Doctoral Researcher with the College of Information Engineering, Zhejiang University of Technology, where he is currently a Lecturer. His current research interests include network-based algorithm design, social network data mining, chaos synchronization, network dynamics, and machine learning.



**Yun Xiang** received the B.S. degree in information and electrical engineering from Zhejiang University, Hangzhou, China, in 2006, the M.A.Sc. degree in electrical engineering from the University of Massachusetts, Amherst, MA, USA, in 2008, and the Ph.D. degree in electrical engineering from the University of Michigan, Ann Arbor, MI, USA, in 2014, respectively.

He is currently as an Associate Professor with the College of Information Engineering, Zhejiang University of Technology, Hangzhou, China. His current research interests include wireless sensor network, data fusion, machine learning, and network algorithms.

**Zhefu Wu** received the Ph.D. degree from the College of Information Science & Electronic Engineering, Zhejiang University, Hangzhou, China, in 2000.

He is currently an Associate Professor with the College of Information Engineering, Zhejiang University of Technology, Hangzhou. His current research interests include social network data mining, complex network dynamics, machine learning, and wireless sensor network algorithms and applications.

**Vladimir Filkov** received the Ph.D. degree in computer science from Stony Brook University, Stony Brook, NY, USA, in 2002.

He is the Co-Director of the DECAL Lab. with the Department of Computer Science, University of California at Davis, Davis, CA, USA. In 2002, he joined as an Assistant Professor with the University of California at Davis, Davis, CA, USA, where he is currently a Professor of computer science. His research has been funded by the U.S. National Science Foundation (NSF), U.S. Department of Agriculture (USDA), U.S. Forest Service (USFS), U.S. Air Force Office of Scientific Research (AFOSR), and others. His current research interests include empirical software engineering, systems biology and gene networks, applied network theory, and data mining and algorithms.

Dr. Filkov is member of the ACM. His papers have received a number of best paper awards and other recognitions.