IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received April 9, 2018, accepted May 10, 2018, date of publication May 28, 2018, date of current version July 6, 2018.

Digital Object Identifier 10.1109/ACCESS.2018.2841432

Unraveling the Impact of Users' Interest on Information Dissemination in Wireless Networks

XIAOYING GAN[1,2], ZHIDA QIN[1], LUOYI FU[1], AND XINBING WANG[1]

[1]Department of Electronic Engineering, Shanghai Jiao Tong University, Shanghai 200240, China
[2]National Mobile Communications Research Laboratory, Southeast University, Nanjing 210096, China

Corresponding author: Xiaoying Gan (ganxiaoying@sjtu.edu.cn)

This work was supported in part by NSF China under Grant 61671478, Grant 61672342, Grant 61602303, and Grant 61532012, in part by the Science and Technology Innovation Program of Shanghai under Grant 17511105103, and in part by the open research fund of the National Mobile Communications Research Laboratory, Southeast University, under Grant 2018D06.

ABSTRACT In this paper, we analyze the throughput of data dissemination at the level of users' interests. We show that users' interests have the ability to drastically improve upon existing throughput scaling's established under the assumption that users show the same preference in any type of data they encounter. More precisely, we consider the scenario where each data source estimates the recipients that will be interested in its data based on user interest probability, which is described by a Zipf-distributed data popularity that decays of exponent α with data ranking. For such a user-centric model, we divide our analysis into different cases depending on data catalogue size K and study their respective throughput performance. With totally n users assumed, we present closed-form expressions of user-centric throughput versus n, α, and K. In particular, our results reveal that when $\alpha = 1$ where users' interests exhibit a moderate level of heterogeneity, the maximum throughput of $\Theta(\sqrt{n})$ (except for a poly-logarithmic factor) can be achieved in all the situations, with appropriate choice of K. The results augment the existing scaling laws derived in network-centric situation, in that given the same throughput data can be disseminated efficiently to more recipients in a user-centric network.

INDEX TERMS Wireless network, information exchange, user centered design.

I. INTRODUCTION

A new trend is nowadays entailed on the Internet, where a growing number of websites offer users the possibility to actively contribute content. It is reshaping the way of information dissemination in that more diverse information generated by users themselves can be available to others. Typical applications include You Tube [1]–[3], microblog [4], Facebook [8] and P2P networks [5]–[7], which have manifested themselves not only as websites on which people can read and share information, but also provided their visitors the possibility to create and maintain a social network. As opposed to the early days of network-centric scenario where users must unconditionally accept the information from some specific sources, such identical popularity (or unpopularity) of information among users is much more diluted in Internet. Constant waves of new information contributed by users are quickly personalizing their preference, leading to a great variability in user behavior and attention span in the Internet. Such scenario is dubbed user-centric network. User-centric network model has been widely investigated in existing works [38], [39]. Note that different from the content-centric network [40], user-centric network not only highlights the data popularity, but also the social connections.

The new trend of user-centric information dissemination is also penetrating from the Internet into wireless ad hoc networks, as the wireless networks themselves are developing toward a social medium, where people's preference during information exchange is more highlighted than before. The vast types of data generated by users themselves fit quite well to the property of wireless ad hoc networks, where there is no reinforcement from any centralized equipments on users' reception behavior. Also, as data can be stored in a distributed mode, ad hoc network can alleviate the heavy burden imposed on the central infrastructure from which all nodes must go through for certain data acquisition. And we therefore envision future wireless ad hoc networks shall be a good alternative of centralized network for user-centric data dissemination. Among numerous papers focusing on user-centric data transmission in wireless ad hoc networks, the most related ones consider social characteristics of nodes [9]–[14], where users' interests are identified from social perspective such as homophily or community-based assumption. However, these works mainly focus on proposals of routing schemes for satisfying users' interests, yet little is known about their impact on network performance metrics such as throughput, delay, etc. Established under Kumar's framework about scaling law analysis of throughput, some previous works have explored socially-connected wireless network throughput [15]–[17], while these works are mainly based on network structure analysis rather than the users' interests. These scenarios are essentially ''network-centric'', ignoring the satisfaction of users' interests. Data is usually forwarded to many users not interested in it, resulting in large amounts of network resources wasted.

This motivates us to first address the problem of network throughput in user-centric scenario from a social interest perspective. In present work, we consider satisfying users' interests and forward data only to users that are interested in it. Such recipients can also be called ''interesters'' in the rest of this paper. The major difficulty of user-centric data dissemination in wireless networks lies in that the interesters of a specific data class are generally unknown in advance at the data source, because it is difficult for the source to have knowledge about the interests of other users in the network. Such uncertainty of data recipients is different from the traffic pattern such as unicast [25]–[29] and multicast [30], where the destinations are fixed and pre-known. This makes routing strategy of user-centric data dissemination rather challenging.

Our main idea to overcome the aforementioned difficulty is to let a user estimate the possible recipients that will be interested in data as probability. Categorizing the data into K classes ranking in a descendent order, we formulate a user's interest in data item with a Zipf distribution decaying of exponent α with data ranking, as there exists ample evidence in the literature [3]–[19] that the data popularity in some practical cases follows such power laws. This probability is crucial to the determination of data recipients since a data source has no prior-information other than estimation of its possible recipients. We conduct our study in different scenarios based on data catalogue size, i.e., limited data catalogue where the total number of data classes K is no more than the total number of users n and large data catalogue where K is no less than n. In all the cases, the throughput bounds turn out to be quite delicate as we vary α and K. Notably, for $\alpha = 1$ we achieve the best performance, i.e., $\Theta(\sqrt{n})$ throughput scaling (except for a poly-logarithmic factor), and, over a wide range of values for α, the results may also manifest significant improvement over some existing bounds derived under the network-centric scenario. The performance benefit stems from the minimization of wireless hops, thanks to users' interest heterogeneity. The corresponding throughput-achieving communication schemes are also discussed.

It is worth noting that the purpose of our work is not to establish optimal information theoretic results, but a first attempt showing that there is an additional perspective to be exploited, i.e., users' interest, which has not been well considered in a lot of theoretical studies aimed at establishing fundamental scaling laws of ad-hoc networks. More sophisticated techniques can be added to our scheme, and can further improve the bounds presented here.

The roadmap of the paper is as follows. Section 2 lists literature review of some existing studies from the perspective of user interest as well as some scaling law analysis of throughput. We introduce the network model and list the definitions in Section 3. The main results of this paper are briefly introduced in Section 4. We give detailed analysis of user-centric throughput in Section 5. Section 6 is contributed to some discussions on our results and their implications. We give concluding remarks in Section 7.

II. RELATED WORKS

Among the previous literature, user-centric data transmission is mainly investigated under the background of content-based networking, where data requests are placed on content and routes are formed based on content provision and user interest. Two of the related works consider content delivery distribution [20], [21], the main idea of which is to store different copies of the content inside the networks so that terminal users can still retrieve the information even when one or more than one node disconnected. Some successful commercial distributed storage system include BigTable [22], OceanStore [23] and distributed hash table (DHT) [24].

This paper focuses on throughput bound of large-scale user-centric wireless networks. The fundamental asymptotic throughput study is initiated by Gupta and Kumar [25], who show that the maximal per-node unicast throughput achievable in wireless networks is $\Theta(1/\sqrt{n \log n})$ for a uniformly distributed destination. A series of works [26]–[32] have then followed, through either unicast or multicast. Nevertheless, all those studies are based on network-centric framework. Recently, there are only a few studies on throughput analysis in user-centric networks. In particular, the most related ones explore throughput scaling through backbone structure [33], content replication [34] and user mobility [35]. To address the concern of user demands, Rahul *et. al.* design a system that scales wireless throughput by enabling joint beamforming from distributed independent transmitters. However, it still remains unknown how user interests can affect the scaling performance in large-scale wireless networks.

III. SYSTEM MODEL

A. NETWORK TOPOLOGY

We consider a dense network $\mathcal{O}$ as a unit square. The size normalization and wrap-around conditions are also introduced here, which are common technical assumptions adopted in previous works to avoid tedious technicalities.

Note that these assumptions will not change the main results of this paper. n users with wireless communication capability are located in the network and exchange information in an ad hoc manner. Their locations, which can be denoted by a series of independent random variables, are uniformly distributed in $\mathcal{O}$.

B. COMMUNICATION MODEL

The well-known protocol model is introduced here to roughly represent the behavior of transmission constrained by interference. And the results are also applicable to the more refined physical model but it is beyond the scope of this paper. The model indicates under a fixed total bandwidth W, two node i is allowed to transmit to j if the positions of i and j, denoted by X_i and X_j, satisfy $\| X_i - X_j \| < r$, where r is a common transmission range employed by all the nodes and for every other node k transmitting, $\| X_j - X_k \| > (1+\Delta)r$, being Δ a guard factor.

C. USER INTEREST AND DATA MODELING

Our model estimates the interest of each user in data item as probability. We assume that data is categorized into K classes, each class contains one or multiple data items. The **interest profile** of a user i is a $K \times 1$ probability vector $\mathbb{P}_i = [q(1), q(2) \ldots, q(k), \ldots, q(K)]^T$, where $q(k)$ indicates the user probability to be interested the k-th class. Without loss of generality, we assume $\sum_{k=1}^{K} q(k) = 1$ and $\mathbb{P}_i$ can be considered as a discrete probabilistic distribution. In our paper, we assume a Zipf distribution, where the probability that a user is interested in a data class of rank k is

$$q(k) = \frac{1/k^\alpha}{\sum_{k=1}^{K} 1/k^\alpha}, \tag{1}$$

α being the power law parameter indicating the rate of popularity decline. Let $H \triangleq 1/\left(\sum_{k=1}^{K} 1/k^\alpha\right)$ and we can also express $q(k)$ as $q(k) = H/k^\alpha$. In the above equation, we sort different classes following their popularity, i.e., the most popular data class is the class 1, and the least popular one is class K.[1] This choice is supported by a number of measurements papers which have found Zipf distribution to be quite ubiquitous in experimental traces related to data dissemination patterns over the Internet [3], [19]. In [3], Cha *et al.* study the viewing patterns against videos with distinct ranks based on YouTube and other similar user content generated systems. The empirical video popularity distribution against the number of views is found to be Zipf-distributed. Qiu *et al.* [19] conduct in-depth analysis on channel popularity on a large collection of user channel access data from a nation-wide commercial IPTV network and find that channel popularity is highly skewed and can be well captured by a Zipf-like distribution. Note that the data popularity model comprises, as a special case, the network-centric scenario in which users cast identical interest to any incoming data. This extreme case occurs at $\alpha = 0$, making $q(k)$ give rise to uniform distribution where any one of the K data classes has equal probability $1/K$. A wider applications can be matched through adjusting α, of which low values corresponds to the data distribution in routers, intermediate values in proxies and higher values in mobile applications.

[1] Note that in the present model a popular data class is popular among all individual users. It is also an interesting future work to further consider different users' preferences for data classes, e.g., some users are more interested in news while some are more interested in movies.

D. TRAFFIC PATTERN

Each of the n users can act as a data source, a recipient or just a relay which helps forward the data. If a user acts as a data source, it does not know a priori who the recipients are. Similarly, the recipients do not know who and when the data was generated from since they will optionally receive the data based on their own interests. Hence, in our traffic pattern, it is the data that will determine the recipients. And throughout the rest of our paper we will call these recipients "interesters" or destinations. This differs the traffic from unicast and multicast in that their source-destination pairs are pre-known and fixed. In our model the number of destinations per source may vary based on the class of the data stored at the source.

We also assume that each source has a data buffer storing the data he will generate. Besides, each user also has a relaying buffer for data to be forwarded to those intersters and a receiving buffer storing the data it is interested in. Note that data sent out from the same source always gains the same preference order among users, i.e., the same probability that a particular user will be interested in it.

E. THROUGHPUT DEFINITION

Definition 1: **User-centric Throughput**: For a specific data class k, the aggregate user-centric throughput, Λ_k, is said to be feasible if there is a spatial and temporal scheme for scheduling transmissions, such that by operating the network in a multi-hop fashion and buffering at intermediate nodes when awaiting transmission opportunities, each data source can transmit $g(k)$ bits/s of class k to its N_I^k intersters. That is, there is a $T < \infty$ such that in every time interval $[(i-1) \cdot T, i \cdot T]$, every data source can send $T \cdot g(k)$ bits of data from class k to each of its N_I^k interesters.

Definition 2: **The average aggregate user-centric throughput** Λ can be obtained through taking the average on Λ_k for all $k \in [1, 2, \ldots, K]$, i.e., $\Lambda = \sum_{k=1}^{K} \Lambda_k$.

Note that both Λ_k and Λ will be our major concern throughout the paper.

IV. MAIN RESULTS

A graphical representation of our results is reported in Figures 1-3, respectively. We adopt the order notation $\widetilde{\Theta}(\cdot)$ to hide poly logarithmic factors for better readability. Refined results are available in Section V.

Figures 1 and 2 plot the aggregate user-centric throughput Λ achieved versus different values of popularity parameter α

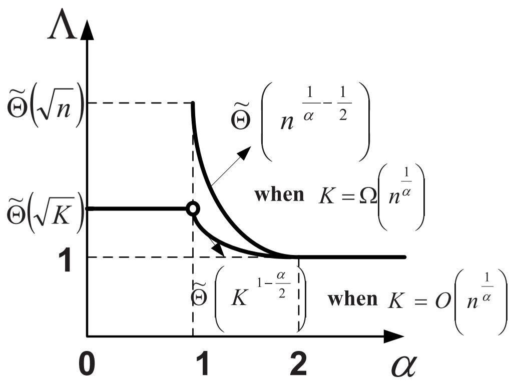


FIGURE 1. The aggregate user-centric throughput Λ versus α in the case of limited data catalogue, where $K = O(n)$ and each data class is stored by one data source.

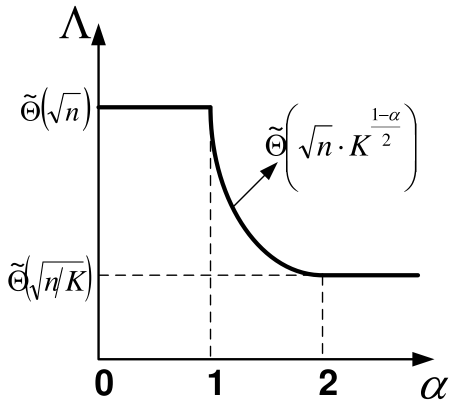


FIGURE 2. The aggregate user-centric throughput Λ versus α in the case of limited data catalogue, where $K = O(n)$ and each data class is stored by n/K data sources.

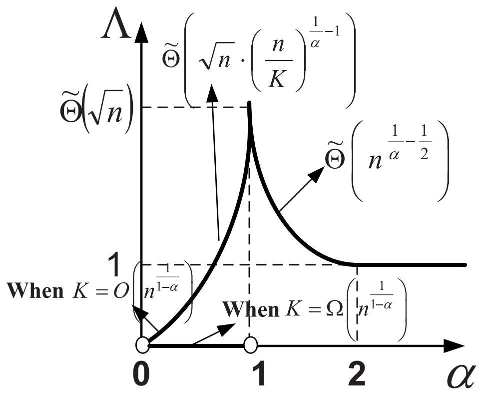


FIGURE 3. The aggregate user-centric throughput Λ versus α in the case of large data catalogue, where $K = \Omega(n)$ and each of the n users stores K/n non-overlapping classes.

in the case of limited data catalogue, where the total number of data classes K available is no more than the number of users n in the network. The difference between the scenarios in the two figures lies in that Figure 1 considers one data item per class with each data item stored initially in one data source (totally K sources) whereas Figure 2 considers n/K data items per classes distributed at n/K non-overlapping sources at the beginning (with totally n sources for all the K classes). As a counterpart, Figure 3 represents the throughput in large data catalogue size, where K is no less than the number of sources n, each with K/n data items from non-overlapping classes. Note that the two cases of limited data catalogue can model the real situations where limited contents are shared by a large number of users. Similarly, the case of large data catalogue reflects the circumstances where abundant contents are shared among users. All the three figures suggest that the aggregate throughput is maximized to $\widetilde{\Theta}(\sqrt{n})$ at the point where $\alpha = 1$, with appropriate choice of K in each figure ($K = \Theta(n)$ in Figure 1, any $K = O(n)$ in Figure 2 and any $K = \Omega(n)$ in Figure 3). The throughput decreases smoothly as α increases from 1 to 2 and is stabilized to a particular value after $\alpha > 2$ ($\Theta(1)$ in Figure 1 and Figure 3 and $\widetilde{\Theta}(\sqrt{n/K})$ in Figure 2). The results suggest that appropriate heterogeneity of users' interests (the best heterogeneity occurs at $\alpha = 1$) can lead to throughput improvement and are also likely to outperform the uniform case ($\alpha = 0$) where each user exhibits identical interests in all data types.

V. USER-CENTRIC THROUGHPUT ANALYSIS

A. PRELIMINARIES

Before we proceed, we first need to establish some important properties of our data popularity model. The following lemma shows how the parameter H in Zipf-distributed popularity assumed in our model varies with α, in order sense. Note that this result will be used throughout the paper.

Lemma 1: For the data popularity which follows a Zipf distribution of $q(k) = H/k^\alpha$, the expression of H is shown in the following cases, in different α:

$$H = 1/\left(\sum_{k=1}^{K} \frac{1}{k^\alpha}\right) = \begin{cases} \Theta(1) & \alpha > 1 \\ \Theta\left(\frac{1}{\log K}\right) & \alpha = 1 \\ \Theta\left(K^{\alpha-1}\right) & 0 \le \alpha < 1. \end{cases} \tag{2}$$

Proof: Let $f(x)$ denote a continuous decreasing function which satisfies $f(x) > 0$ for all $x \ge 1$. Then we have

$$\int_{x=1}^{K} f(x)dx \le \sum_{k=1}^{K-1} f(k) \le f(1) + \int_{x=1}^{K-1} f(x)dx.$$

We set $g(K) = \sum_{k=1}^{K} \frac{1}{k^\alpha}$. Then, equation (2) holds. With $H = 1/g(K)$, we can derive the results for H. This completes our proof. ∎

B. LIMITED DATA CATALOGUE

We first consider the limited data catalogue, which refers to the case where the total number of data classes K is no more than the total number of users n, i.e., $K = O(n)$. In limited data catalogue, data from K classes are initially stored at K out of n users acting as data source. We further assume one data item within each class to be disseminated over the network to span various users' attention. We will get into the finer resolution of different objects within each class in Section 5.3. The lemma given below indicates the properties exhibited in the case of limited data catalogue.

Lemma 2: For a data item belonging to class k, where $k \in [1, \ldots, K]$, the number of estimated intersters N_{I_k} in it is

$$N_{I_k} = \Theta\left(n \cdot \frac{H}{k^\alpha}\right). \tag{3}$$

And these N_{I_k} interesters are randomly and uniformly distributed in the whole network.

1) UPPER BOUND

We first analyze the upper bound of both user-centric throughput for data class k and the aggregate user-centric throughput in limited data catalogue, as are shown in the following lemma:

Lemma 3: In the case of limited catalogue where K out of n users serve as data sources for each of the K data classes, the user-centric throughput Λ_k for data class k and the aggregate user-centric throughput Λ can be upper-bounded by

$$\begin{cases} \Lambda_k \leq \dfrac{N_{I_k} \cdot W}{\mathscr{I}_k} \\ \Lambda \leq \displaystyle\sum_{k=1}^{K} \dfrac{N_{I_k} \cdot W}{\mathscr{I}_k}, \end{cases} \tag{4}$$

where $\mathscr{I}_k$ is the total number of non-interesters of class k overhearing a copy of data from class k and W is the total bandwidth for transmission.

Proof: Given the set $\mathcal{K}$ of K data classes, let $\Lambda_{\mathcal{K}} = (\Lambda_1, \Lambda_2, \ldots, \Lambda_K)$ be the rate vector of all K data classes. Then, for a data class k, assume that $\mathscr{I}_k$ users who are not interested in data from data class k but overhear a copy from it during data transmission. Obviously, $\Lambda_k \cdot \mathscr{I}_k \leq N_{I_k} \cdot W$, where the right side represents the total data that can be disseminated for data class k over the whole network. Taking summation of Λ_k on all $k \in [1, \ldots, K]$, we can obtain the upper bound of the average aggregate user-centric throughput, i.e., $\Sigma_{\Lambda_k \in \Lambda_{\mathcal{K}}} \Lambda_k \leq \Sigma_{k=1}^{K} \frac{N_{I_k} W}{\mathscr{I}_k}$. ■

To obtain the closed-form expressions for the upper bound of both Λ_k and Λ, we need to calculate the lower bound of $\mathscr{I}_k$ described in Lemma 3. This is usually figured out through the average length of spanning tree in traditional multicast. However, the problem is complicated in our model by the fact that the number of destinations is not fixed and is dependent on data popularity probability, which is different from the spanning tree established in multicast [30]. In order to proceed, we can simplify the problem by focusing on establishing a tree T_k for a specific data class k first. The tree is estimated to span from source to N_{I_k} interesters, where N_{I_k} varies according to different k. Lemmas 4-6 provide some properties of T_k.

Lemma 4: For the data belonging to class $k \in [1, \ldots, K]$, the total transmission length, denoted by $\| T_k \|$, of any tree T_k spanning $\frac{H \cdot n}{k^\alpha}$ nodes that are randomly placed over the whole network almost surely is at least $\tau \cdot \sqrt{\frac{H \cdot n}{k^\alpha}}$, where τ is a constant.

Lemma 5: The area region A_k containing the number of non-interesters of data class k is at least $c \cdot r\sqrt{N_{I_k}}$ when the number of interesters of class k satisfies $N_{I_k} \leq n/\log n$.

Proof: Let L_k denote the area covered by the number of leaf nodes in T_k. Obviously, since there are N_{I_k} interesters of data class k, the number of leaf nodes is thus N_{I_k}. These N_{I_k} occupy a total region with area $N_{I_k} \cdot \pi r^2$. Let In_k denote union region covered by the internal nodes of T_k. Due to the random and uniform distribution of these N_{I_k} users (Lemma 2), the problem can be converted to figure out the minimum tree spanning from the source of data class k to its N_{I_k} interesters and get it to be $\sigma\sqrt{N_{I_k}}$ according to [37]. Then, we can get $\mathrm{In}_k = \frac{\sigma\sqrt{N_{I_k}}}{r} \cdot \pi r^2 = \sigma\pi\sqrt{N_{I_k}} \cdot r$. Hence, the area of the region covering those non-interesters that get a copy of data from class k is $A_k \geq \mathrm{In}_k - L_k = \sigma\pi\sqrt{N_{I_k}} \cdot r - N_{I_k} \cdot \pi r^2$. Note that $\sigma\pi\sqrt{N_{I_k}} \cdot r > N_{I_k} \cdot \pi r^2$ in order sense when $N_{I_k} = O\left(\frac{1}{r^2}\right)$. Therefore, we can conclude that $A_k = c \cdot r\sqrt{N_{I_k}}$, where $c = \sigma\pi$. This completes our proof. ■

Lemma 6: If the estimated number of interesters N_{I_k} on data class k exceeds a threshold $\eta\frac{n}{\log n}$, where η is a constant, then the number of empty sub-squares occupied by those interesters is at most a constant fraction of the total number of sub-squares. It has already been inferred from Lemma 2 that the estimated $N_{I_k} \geq \eta\frac{n}{\log n}$ is randomly and uniformly chosen among the n users during data dissemination process.

Proof: We partition the unit square into sub-squares with side length r. Hence, the square will be partitioned into $L = \lfloor\frac{1}{r^2}\rfloor$, denoted as $\{M_i\}_{i=1}^{L}$.

Let X be the number of squarelets that do not have any of these N_{I_k} interesters inside. We define the variable X_i as

$$\begin{cases} 1, & \text{if there is no interesters of class } k \text{ in } M_i. \\ 0, & \text{if there are interesters of class } k \text{ in } M_i. \end{cases}$$

Let $X = \sum_{i=1}^{n} X_i$. According to Chebyshevąŕs Inequality, we have

$$\Pr(|X - \mathbb{E}[X]| \geq \tau \cdot L) \leq \frac{\mathrm{Var}(X)}{\tau^2 L^2}, \tag{5}$$

where τ is a constant in the range [0, 1]. Obviously, we have

$$\begin{aligned} \mathrm{Var}(X_i) &= \mathbb{E}[X_i^2] - \mathbb{E}^2[X_i] \\ &= \left(1 - \frac{1}{L}\right)^{N_{I_k}} - \left(1 - \frac{1}{L}\right)^{2N_{I_k}}, \end{aligned}$$

and

$$\begin{aligned} \mathrm{Var}(X) &= \sum_{i=1}^{n} \mathrm{Var}(X_i) = n \cdot \mathrm{Var}(X_i) \\ &= L\left[\left(1 - \frac{1}{L}\right)^{N_{I_k}} - \left(1 - \frac{1}{L}\right)^{2N_{I_k}}\right]. \end{aligned}$$

Substituting it into Inequality (5), we obtain

$$\Pr((X-\mathbb{E}[X])\geq\tau\cdot L)\leq\frac{L\left[\left(1-\frac{1}{L}\right)^{N_{I_k}}-\left(1-\frac{1}{L}\right)^{2N_{I_k}}\right]}{\tau^2L^2}.$$

Since $N_{I_k}\geq\eta\frac{n}{\log n}\geq\eta\cdot L, L\cdot\left(1-\frac{1}{L}\right)^{N_{I_k}}\leq L\cdot e^{-\eta}$. Hence,

$$\Pr(X\geq\mathbb{E}[X]+\tau\cdot L)\leq\frac{e^{-\eta}-e^{-2\eta}}{\tau^2L^2}. \tag{6}$$

The right side of Inequality (6) goes to zero as L goes to infinity. Moreover,

$$\mathbb{E}[X]+\tau\cdot L\leq L\cdot e^{-\eta}+\tau\cdot L=L(e^{-\eta}+\tau)\leq\varphi\cdot L,$$

where ψ is a constant belonging to [0, 1]. We draw the conclusion that the number of empty sub-squares occupied by those interesters is at most a constant fraction of the total number of sub-squares. ■

A consideration we should take into is the transmission range r mentioned in Lemmas 5 and 6. Previous investigations have shown that transmission ranges should not be too large to increase power consumption while maintaining network connectivity. A range $r=\Theta\left(\sqrt{\frac{\log n}{n}}\right)$ is commonly adopted to make the whole network fully connected, which has already been verified in [25]. Based on this, the A_k in Lemma 5 can further be expressed as $A_k=\sqrt{N_{I_k}\log n/n}$. Since $N_{I_k}=n\cdot H/k^{\alpha}$, substituting it into A_k we can get $A_k=\sqrt{H\cdot\log n/k^{\alpha}}$. With A_k derived under N_{I_k} satisfying $O\left(\frac{n}{\log n}\right)$ and $\Omega\left(\frac{n}{\log n}\right)$, we can figure out the number of non-interesters $\mathscr{I}_k$ receiving the data from class k, as are shown in Lemmas 7 and 8:

Lemma 7: When $N_{I_k}=O(n/\log n)$, we can express the number of non-interesters $\mathscr{I}_k$ obtaining data from class k as $\mathscr{I}_k=n\sqrt{H\cdot\log n/k^{\alpha}}$.

Proof: Since all the users are uniformly distributed over the whole network, the number of users located in A_k is on average $\frac{A_k\cdot K}{1}=n\sqrt{\frac{H\cdot\log n}{k^{\alpha}}}$. This completes our proof. ■

Lemma 8: When $N_{I_k}=\Omega\left(\frac{n}{\log n}\right)$, the area A_k occupies a constant fraction of the whole network area, i.e., $A_k=\varrho\cdot 1=\varrho$. Thus, the number of non-interester of class k that obtain a copy of it can be expressed as $\mathscr{I}_k=\varrho n$.

The last step before we obtain the upper bound of Λ_k is to specify the ranges of k, in which the condition $N_{I_k}=\Omega\left(\frac{n}{\log n}\right)$ and $N_{I_k}=O\left(\frac{n}{\log n}\right)$ can be satisfied respectively. The next lemma gives the ranges of k where $N_{I_k}=\Omega\left(\frac{n}{\log n}\right)$, varying with α:

Lemma 9: In a network with totally n users, for a specific data class k, the number of intersters $N_{I_k}=O(\frac{n}{\log n})$ if and only if k satisfies

$$k=\begin{cases}\Omega\left(\log^{\frac{1}{\alpha}} n\right) & \alpha>1\\ \Omega(1) & \alpha=1\\ \Omega\left(n\left(\frac{\log n}{n}\right)^{\frac{1}{\alpha}}\right) & 0<\alpha<1\\ \forall k\in[1,\ldots,K] & \alpha=0.\end{cases} \tag{7}$$

Proof: According to Lemma 2, we have already known that for data class k, N_{I_k} is estimated to be $\Theta\left(n\cdot\frac{H}{K^{\alpha}}\right)$. Hence, substituting it into the equation $N_{I_k}=O(\frac{n}{\log n})$, we obtain that the equation holds when $k\geq H^{\frac{1}{\alpha}}\log^{\frac{1}{\alpha}} n$. According to the result of H shown in Lemma 1, the results of this lemma naturally hold. ■

With the closed-form expression of $\mathscr{I}_k$, we can express the upper bound shown in Lemma 3 for both $N_{I_k}=O(n/\log n)$ and $N_{I_k}=\Omega(n/\log n)$ as follows:

Lemma 10: In each subnetwork with n users, the upper bound for the user-centric throughput Λ_k of data class k and the aggregate user-centric throughput Λ is

$$\begin{cases}\Lambda_k=O\left(\sqrt{\frac{H}{k^{\alpha}\log n}}\right) & k=\Omega\left((H\log n)^{\frac{1}{\alpha}}\right)\\ \Lambda_k=O\left(\frac{H}{k^{\alpha}}\right) & k=O\left((H\log n)^{\frac{1}{\alpha}}\right)\end{cases} \tag{8}$$

2) THROUGHPUT-ACHIEVING SCHEME

Now we will propose an optimal routing scheme, under which the throughput achieved can reach the upper bound proposed above. For user-centric data routing, the cell partition TDMA scheme is highly efficient for scheduling active transmissions in the network. However, routing becomes a major issue in user-centric traffic pattern since an optimal routing tree needs to be established. Moreover, the destinations (also called interesters) of data from class k cannot be determined and fixed in prior. Instead, they can be estimated only during data transmission, based on the ranking of data class for the incoming data. Our main idea is to first construct a Euclidean spanning tree using Prim's algorithm, and then convert it to a user-centric routing tree. We present the user-centric routing scheme shown as follows:

Optimal User-Centric Routig Tree:

Step 1: Construct a spanning tree using Prim's algorithm:

(a) Interesters of data class k form N_{I_k} components.
(b) The network is partioned into $N_{I_k}-g$ squares with side length of each square being $1/\lceil\sqrt{N_{I_k}-g}\rceil$. ($g=1,2,\ldots,N_{I_k}-1$.)
(c) Find a square that contains two nodes from two different connected components. Merge the two components by adding a edge between the two nodes.
(d) For each $g\in[1,\ldots,N_{I_k}-1]$, repeat step (b) and (c) until $g=N_{I_k}-1$. Return the User-centric Spanning Tree, denoted by UCST(k) for data class k.

Step 2: Divide the network into cells with side length r. For each edge uv in UCST(k), randomly select a point w that is in the same row as u and the same column as v. Then select a node in each of the cells which uw and wv are crossed by. Connect those users to form a path from u to v.

Step 3: Combine the paths and remove cycles. Return the obtained user-centric routing tree UCRT(k) for data class k.

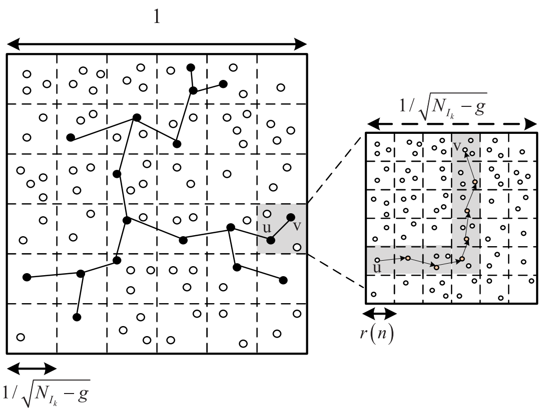


FIGURE 4. User-centric routing tree and multi-hop in step 2.

An illustration of user-centric routing is shown in Figure 4. One of the most significant factors related to routing tree is the tree length spanning from a specific source to all destinations. Denoting $\parallel$ UCRT(k) $\parallel$ as the tree length for data class k, we have

Lemma 11: For $\frac{Hn}{k^\alpha}$ interesters of data in class k randomly located in a unit square, the length of User-Centric Routing Tree $\parallel$ UCRT(k) $\parallel$ is at most $\rho\sqrt{\frac{Hn}{k^\alpha}}$, where ρ is a constant.

Note that employing transmission for UCRT(k) of a specific $k \in [1, \ldots, K]$ is conducted through TDMA scheme. Each cell is treated as a scheduling unit and transmits in a round-robin fashion. For each active cell, randomly select a user located in that cell. Which data class to disseminate depends on which data class the selected user will send if it acts as a data source, or on which data class it will help to relay if it acts as a relay for that class. In order to analyze the throughput for each data class k, it is important to figure out the load for k in each cell, as is shown in Lemma 12.

Lemma 12: Given a squarelet c, the probability that the flow for data class k is routed through c is upper bounded by $\kappa \parallel \text{UCRT}(k) \parallel r$.

Proof: Recall that we will construct the UCRT as the method described in Algorithm 1, which is composed of N_{I_k} steps. For a given data class k, the squarelet c may be used in any one of the N_{I_k} steps to build the tree. For step g (with $1 \leq g \leq N_{I_k}$), the network is divided into $\lfloor N_{I_k} - g \rfloor^2$ cells with side length $1/\sqrt{\lfloor N_{I_k} - g \rfloor}$. Let I_g represent the indicator that c is used in step g. Then the probability that $\Pr(I_g = 1|N_{I_k})$ can be expressed as

$$\Pr(I_g = 1|N_{I_k}) = \frac{1}{\lfloor N_{I_k} - g \rfloor} \cdot p_c(g), \tag{9}$$

where $\frac{1}{\lfloor N_{I_k} - g \rfloor}$ is the probability that the cell containing squarelet c is used and p_c represents the probability that squarelet c is used when that cell containing c is used. With this cell further tessellated into squarelets with area r, assume that c locates in the i_{th} row and j_{th} column, then we have

$$\begin{aligned} p_c(g) &= (i-1) \cdot \left[r^2 \lfloor N_{I_k} - g \rfloor - (i-1) r^3 \lfloor N_{I_k} - g \rfloor^{\frac{3}{2}} \right] \\ &\quad + (j-1) \cdot \left[r^2 \lfloor N_{I_k} - g \rfloor - (j-1) r^3 \lfloor N_{I_k} - g \rfloor^{\frac{3}{2}} \right] \\ &\leq 2r \lfloor \sqrt{N_{I_k} - g} \rfloor, \end{aligned} \tag{10}$$

where the first (second) term in Equation 10 is the probability that squarelet c is used when u (resp. v) is on the same row (resp. column) as c. Taking summation from 1 to N_{I_k}, we get

$$\begin{aligned} \Pr(I_g = 1) &= \sum_{g=1}^{N_{I_k}} \Pr(I_g = 1|N_{I_k}) \\ &= \sum_{g=1}^{N_{I_k}} \frac{2r}{\lfloor N_{I_k} - g \rfloor} \lfloor \sqrt{N_{I_k} - g} \rfloor \\ &\leq \sum_{g=1}^{N_{I_k}} \frac{2r}{\lfloor N_{I_k} - g \rfloor} = \Theta\left(r \cdot \sqrt{N_{I_k}} \right) \\ &= \kappa \parallel \text{UCRT}(k) \parallel r. \end{aligned} \tag{11}$$

■

Lemma 13: For a specific data class k, the number of sessions for k that invoke c for routing uniformly over all cells is

$$\lim_{K \to \infty} \Pr\left(\bigcap_{c=1}^{K} \{ N(c) \leq \kappa' K \parallel \text{UCRT}(k) \parallel r \} \right) = 1.$$

Proof: For a specific squarelet c, we have $N(c) = \sum_{k=1}^{K} I_c$, where I_c represents the indicator function that squarelet c is invoked by transmission of data class k. According to Lemma 12, I_c is *i.i.d.* Bernoulli random variables with probability $p \leq \kappa \parallel \text{UCRT}(k) \parallel r$. By Chernoff bounds, we have

$$\begin{aligned} \Pr\left(N(c) > 2\mathbb{E}\left[\sum_{k=1}^{K} I_c \right] \right) &< \Pr\left(\sum_{k=1}^{K} I_c > 2\mathbb{E}\left[\sum_{k=1}^{K} I_c \right] \right) \\ &< \left(\frac{e}{4} \right)^{n \cdot \kappa \|\text{UCRT}(k)\| r} \\ &< (e)^{-n \cdot \kappa \|\text{UCRT}(k)\| r/8}. \end{aligned} \tag{12}$$

Since $r = \Theta(n/\log n)$, we can further get

$$\begin{aligned} &\Pr\left(\bigcap_{c=1}^{K} \{ N(c) \leq \kappa' K \parallel \text{UCRT}(k) \parallel r \} \right) \\ &\geq 1 - \sum_{c} \Pr\left(N(c) > 2\mathbb{E}\left[\sum_{k=1}^{K} I_c \right] \right) \\ &\geq 1 - ne^{-\sqrt{n \log n}/8} \to 1. \end{aligned} \tag{13}$$

Note that the last row of Equaqtion (13) holds as long as n goes to infinity. This completes our proof. ■

The last step before we obtain the closed-form expression of user-centric throughput is to find out the data classes that will get $\Omega(1)$ interesters, as is shown in Lemma 14.

Lemma 14: In a network with total number of users being n, for a specific data class k, the number of interesters N_{I_k} satisfies $N_{I_k} = \Omega(1)$ if

$$k = \begin{cases} O\left(n^{\frac{1}{\alpha}}\right) & \alpha > 1 \\ O\left(\left(\frac{n}{\log K}\right)^{\frac{1}{\alpha}}\right) & \alpha = 1 \\ O\left(n^{\frac{1}{\alpha}} K^{\frac{1}{\alpha}-1}\right) & 0 \leq \alpha < 1. \end{cases} \tag{14}$$

Proof: The estimated number of interesters of data class k is $N_{I_k} = n \cdot \frac{H}{k^\alpha} = \Omega(1)$ is equivalent to $k = O(n \cdot H)^{\frac{1}{\alpha}}$. According to Lemma 1, we can complete the proof. ■

Finally, by employing a TDMA strategy such that every cell has a constant fraction of time for transmission, and further dividing one time slot into mini-slots such that a cell can deliver the traffic for each data class session invoking it, we can get the results shown in the following lemma:

Lemma 15: The user-centric throughput for data class k obtained in our optimal user-centric routing scheme is

$$\Lambda_k = \begin{cases} \Omega\left(\sqrt{\frac{H}{k^\alpha \log K}}\right) & k = \Omega\left((H\log K)^{\frac{1}{\alpha}}\right) \\ \Omega\left(\frac{H}{k^\alpha}\right) & k = O\left((H\log K)^{\frac{1}{\alpha}}\right). \end{cases} \tag{15}$$

Note that the lower bound derived in Equation (15) matches the upper bound derived in Equation (8). Hence, we can obtain the tight bound of Λ_k by simply replacing the Ω with Θ in Equation (15). The detailed expression for both Λ_k and Λ can now be derived for the whole network, as shown in Theorem 1.

Theorem 1: In limited data catalogue where each data class contains only one data item stored initially at one source, the aggregate throughput for a specific data class k, denoted by Λ_k can be expressed as the results shown in Table 1.

And the aggregate average user-centric throughput Λ can be expressed as

$$\Lambda = \begin{cases} \Theta(1) & \alpha > 2 \\ \Theta\left(\sqrt{\log n}\right) & \alpha = 2 \\ \Theta\left(\frac{n^{\frac{1}{\alpha}-\frac{1}{2}}}{\sqrt{\log n}}\right) & 1 < \alpha < 2 \& K = \Omega\left(n^{\frac{1}{\alpha}}\right) \\ \Theta\left(\frac{K^{1-\frac{\alpha}{2}}}{\sqrt{\log n}}\right) & 1 < \alpha < 2 \& K = O\left(n^{\frac{1}{\alpha}}\right) \\ \Theta\left(\sqrt{\frac{K}{\log n \log K}}\right) & \alpha = 1 \\ \Theta\left(\sqrt{\frac{K}{\log n}}\right) & 0 \leq \alpha < 1. \end{cases} \tag{16}$$

Proof: 1. Proof of the aggregate user-centric throughput Λ_k for data class k.

TABLE 1. User-centric throughput vs α and k.

α	k	Λ_k
$\alpha > 2$	$\Omega\left(\min\left((K^2\log n)^{\frac{1}{\alpha}}, n^{\frac{1}{\alpha}}, K\right)\right)$	$\Theta\left(\frac{1}{k^\alpha}\right)$
	$\left[\Omega\left(\log^{\frac{1}{\alpha}} n\right), O\left(\min\left((K^2\log n)^{\frac{1}{\alpha}}, n^{\frac{1}{\alpha}}, K\right)\right)\right]$	$\Theta\left(\frac{\log^{-\frac{1}{2}} n}{\sqrt{k^\alpha}}\right)$
	$O\left(\log^{\frac{1}{\alpha}} n\right)$	$\Theta\left(\frac{1}{k^\alpha}\right)$
$1<\alpha\leq 2$	$\left[\Omega\left(\log^{\frac{1}{\alpha}} n\right), O\left(\min\left(n^{\frac{1}{\alpha}}, K\right)\right)\right]$	$\Theta\left(\frac{\log^{-\frac{1}{2}} n}{\sqrt{k^\alpha}}\right)$
	$O\left(\log^{\frac{1}{\alpha}} n\right)$	$\Theta\left(\frac{1}{k^\alpha}\right)$
$\alpha = 1$	$\left[\Omega\left(\frac{\log n}{\log K}\right), O\left(\frac{K}{\log K}\right)\right]$	$\Theta\left(\frac{\log^{-\frac{1}{2}} n}{\sqrt{k}}\right)$
	$O\left(\frac{\log n}{\log K}\right)$	$\Theta\left(\frac{1}{k\log K}\right)$
$0\leq\alpha<1$	$[1,\ldots,K]$	$\Theta\left(\frac{1}{\sqrt{K\log n}}\right)$

According to Lemma 15, we know that $\Lambda_k = \Theta(\sqrt{H}/\sqrt{k^\alpha \log n})$ when $K > \sqrt{k^\alpha/H\log n}$ and $N_k = O(n/\log n)$. This implies that $k = O\left(K^{\frac{2}{\alpha}}(H\log n)^{\frac{1}{\alpha}}\right)$ and $k = \Omega\left((H\log n)^{\frac{1}{\alpha}}\right)$. We also should bound the minimum number of interesters N_k for class k to be larger than 1, i.e., $n \cdot H/k^\alpha > 1$. This is equivalent to $k = O\left((n\cdot H)^{\frac{1}{\alpha}}\right)$. When $\alpha < 2$, it is obvious that $K^{\frac{2}{\alpha}}(H\log n)^{\frac{1}{\alpha}} > K$. Thus, the condition that $K > \sqrt{k^\alpha/H\log n}$ holds for all $k \in [1,\ldots,K]$ when $\alpha < 2$. Besides, the condition that $k = \Omega\left((H\log n)^{\frac{1}{\alpha}}\right)$ holds for all $k \in [1,\ldots,K]$ when $0 < \alpha < 1$ since $(H\log n)^{\frac{1}{\alpha}} = K^{\frac{\alpha-1}{\alpha}}\log^{\frac{1}{\alpha}} n < 1$. Considering that fact that when $K > \sqrt{k^\alpha/H\log n}$, $\Lambda_k = \Theta(\sqrt{H/k^\alpha \log n})$ when $k = \Omega(H\log n)^{\frac{1}{\alpha}})$ and $\Lambda_k = \Theta(H/k^\alpha)$ when $k = O(H\log n)^{\frac{1}{\alpha}})$, we can obtain the result of Λ_k for $\alpha \leq 1$ shown in Table 1.

For $1 < \alpha \leq 2$, the only concern is if the data class guaranteeing the minimum number of interesters N_{I_k} exceeds the maximum data class K. That is, if $K > n^{1/\alpha}$. Choosing the minimum value between K and $n^{1/\alpha}$, we can get Λ_k for $1 < \alpha \leq 2$ listed in Table 1.

Finally, for the case where $\alpha > 2$, we are concerned about if $K > n^{1/\alpha}$ as well as if $K^{\frac{2}{\alpha}}(\log n)^{\frac{1}{\alpha}} > \min(n^{1/\alpha}, K)$. Listing all the possible range of k satisfying one or both conditions, we can get the throughput results Λ_k presented in Table 1 for $\alpha > 2$.

2. Proof of the aggregate user-centric throughput Λ.

For the aggregate user-centric throughput Λ for all data classes, we only prove the part where $\alpha > 2$. The result of Λ for $\alpha < 2$ can be obtained using the similar analysis. The user-centric throughput Λ_k is reduced to $\Theta(1/k^\alpha)$ when $k > K^{\frac{2}{\alpha}}(\log n)^{\frac{1}{\alpha}}$, $K^{\frac{2}{\alpha}}(\log n)^{\frac{1}{\alpha}} < n^{1/\alpha}$ and $n^{1/\alpha} < K$; Λ_k is bounded by $\Theta(1/\sqrt{k^\alpha \log n})$ when $\log^{\frac{1}{\alpha}} n < k < K^{\frac{2}{\alpha}}(\log n)^{\frac{1}{\alpha}}$ and $n^{1/\alpha} < K^{\frac{2}{\alpha}}(\log n)^{\frac{1}{\alpha}}$; Λ_k reduces to $\Theta(1/k^\alpha)$

when $k < \log^{\frac{1}{\alpha}} n$. Hence, we can obtain Λ for $\alpha > 2$ in the following three cases:

- When $K < n^{\frac{1}{\alpha}}$, we have

$$\begin{aligned}\Lambda &= \sum_{k=1}^{\log^{\frac{1}{\alpha}} n} \left(\frac{1}{k^\alpha} + \Lambda_k\right)\\ &= \sum_{k=\log^{\frac{1}{\alpha}} n}^{K^{\frac{2}{\alpha}} \log^{\frac{1}{\alpha}} n} \frac{1}{\sqrt{k^\alpha \log n}} + \sum_{k=K^{\frac{2}{\alpha}} \log^{\frac{1}{\alpha}} n}^{K} \frac{K}{\sqrt{k^\alpha}} = \Theta(1).\end{aligned} \tag{17}$$

- When $n^{\frac{1}{\alpha}} < K < \sqrt{n}$, we have

$$\begin{aligned}\Lambda &= \sum_{k=1}^{\log^{\frac{1}{\alpha}} n} \left(\frac{1}{k^\alpha} + \Lambda_k\right)\\ &= \sum_{k=\log^{\frac{1}{\alpha}} n}^{K^{\frac{2}{\alpha}} \log^{\frac{1}{\alpha}} n} \frac{1}{\sqrt{k^\alpha \log n}} + \sum_{k=K^{\frac{2}{\alpha}} \log^{\frac{1}{\alpha}} n}^{n^{\frac{1}{\alpha}}} \frac{n^{\frac{1}{\alpha}}}{\sqrt{k^\alpha}} = \Theta(1).\end{aligned} \tag{18}$$

- When $K > \sqrt{n}$, we have

$$\Lambda = \sum_{k=1}^{\log^{\frac{1}{\alpha}} n} \frac{1}{k^\alpha} + \sum_{k=\log^{\frac{1}{\alpha}} n}^{n^{\frac{1}{\alpha}}} \frac{1}{\sqrt{k^\alpha \log n}} = \Theta(1). \tag{19}$$

The proof can be completed by using the similar calculation of Λ to the cases where $\alpha < 2$. ■

C. MULTIPLE SOURCES PER CLASS IN LIMITED DATA CATALOGUE

Another common application in limited data catalogue, and meanwhile our major concern in this subsection is the case of multiple objects within the same data class. More precisely, we assume each data classes has n/K sources randomly distributed over the network, storing different data items from each other (See Figure 5 for illustration.). Since the n/K data items gain the same preference ranking from users, the probability that a user is interested in one of the n/K data items within a specific class k can be expressed as $q(k)_i = \frac{H}{k^\alpha} \cdot \frac{K}{n}$. Hence, with total number of interesters N_{I_k} for each data class, the number of interesters for each of the n/K data items within class k is $N_{I_k} \cdot \frac{K}{n} = n \cdot \frac{H}{k^\alpha} \cdot \frac{K}{n} = K \cdot \frac{H}{k^\alpha}$.

The upper bound of Λ_k can be obtained through the similar techniques in Section 5.2. Also, an optimal user-centric routing scheme to Section 5.2 can be adopted here to reach the upper bound. For the sake of brevity, we skip the corresponding analysis and instead present the key point in throughput calculation, i.e., to find out the maximum data class k_{M_i} that can attract at least one interester in each of the n/K objects per class. This can be figured out in the similar manner of Lemma 14, but deep down into the scale of each data item within the same class. As another significant step, Lemma 16 gives the range of data classes within which

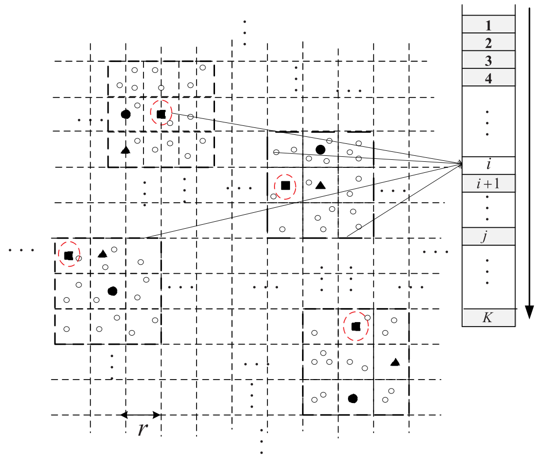


FIGURE 5. **Illustration of multiple sources per data class in limited data catalogue. Each of the K classes has n/K sources randomly distributed over the whole network, storing different data items from each other. The squares, triangles and the circles shown in the figure represent some typical sources. Note that sources of the same shape store a data item that belongs to the same class.**

the number of interesters can be bounded by $n/\log n$, in the similar idea of Lemma 9, but again at the level of each data item within the same class.

Lemma 16: Denote $N_{I_{k_i}}$ $(i \in [1, \ldots, n/K])$ as the number of interesters in each of the n/K contents for a specific data class k. Then, for all $k \in [1, \ldots, K]$, $N_{I_{k_i}}$ satisfies the condition that $N_{I_{k_i}} = O\left(\frac{n}{\log n}\right)$.

Proof: The condition $N_{I_{k_i}} = O\left(\frac{n}{\log n}\right)$ is equivalent to $\frac{KH}{k^\alpha} = O\left(\frac{n}{\log n}\right)$. This implies that $k > \left(\frac{K}{n} H \log n\right)^{\frac{1}{\alpha}}$. Since $K = O(n)$, we know that $K/n < 1$, which further indicates that $\left(\frac{K}{n} H \log n\right)^{\frac{1}{\alpha}} < 1$. The minimum value of k is 1, which is already larger than $\left(\frac{K}{n} H \log n\right)^{\frac{1}{\alpha}}$. Thus, the conclusion in this lemma naturally holds for all $k \in [1, \ldots, K]$. ■

Now we can derive the specific expression of Λ_k and Λ given in Theorem 2.

Theorem 2: In limited data catalogue with multiple data items contained in each data class, the aggregate throughput for a specific data class k, denoted by Λ_k can be expressed as the results shown in Table 2.

The aggregate average user-centric throughput Λ can be expressed as

$$\Lambda = \begin{cases} \Theta\left(\sqrt{\frac{n}{K \log n}}\right) & \alpha > 2 \\ \Theta\left(\sqrt{\frac{n \log K}{K \log n}}\right) & \alpha = 2 \\ \Theta\left(\sqrt{\frac{n}{\log n}} K^{\frac{1-\alpha}{2}}\right) & 1 < \alpha < 2 \\ \Theta\left(\sqrt{\frac{n}{\log K \log n}}\right) & \alpha = 1 \\ \Theta\left(\sqrt{\frac{n}{\log n}}\right) & 0 \leq \alpha < 1. \end{cases} \tag{20}$$

TABLE 2. User-centric throughput vs α and k.

α	k	Λ_k
$\alpha > 1$	$k = \left[1, O\left(K^{\frac{1}{\alpha}}\right)\right]$	$\Theta\left(\sqrt{\frac{n}{k^{\alpha} K \log n}}\right)$
$\alpha = 1$	$k = \left[1, O\left(\frac{K}{\log K}\right)\right]$	$\Theta\left(\sqrt{\frac{n}{kK \log K \log n}}\right)$
$0 \leq \alpha < 1$	$\forall k \in [1, \ldots, K]$	$\Theta\left(\sqrt{\frac{nK^{\alpha-2}}{k^{\alpha} \log n}}\right)$

Proof: 1. Proof of the aggregate user-centric throughput Λ_k for data class k.

Since there are n/K sources in each of the K classes, the user-centric throughput of a specific data class k is the sum of the throughput of each of the n/K sources. Denote λ_i as the throughput of the i_{th} ($i \in [1, \ldots, n/K]$) source at class k, then the user-centric throughput of class k is $\Lambda_k = \sum_{i=1}^{n/K} \lambda_i$. Moreover, since the n/K data items are at the same level of popularity, we can further express Λ_k as $\Lambda_k = \frac{n}{K}\lambda_i$. For each λ_i, according to Lemma 16, we know $\Lambda_k = \Theta(\sqrt{H}/\sqrt{k^{\alpha}\log n})$ for all $k \in [1, \ldots, K]$. Also, we should bound the minimum number of interesters $N_{I_{k_i}} > 1$. This leads to $k = O(H \cdot K^{1/\alpha})$. Based on these conditions, we can obtain the results of $\Lambda_k = \frac{n}{K}\lambda_i$ shown in Table 2.

2. Proof of the aggregate user-centric throughput Λ.

The aggregate user-centric throughput for all data classes can be obtained by taking the summation of Λ_k over all k, i.e., $\Lambda = \sum_{k=1}^{K} \Lambda_k = \sum_{k=1}^{K} \sum_{i=1}^{n/K} \lambda_i$. We only prove the results of $\alpha > 1$ and the results in other scenarios are easy to obtain by taking the summation of Λ_k on all k.

- When $\alpha > 1$, we have $\Lambda = \sum_{k=1}^{K^{\frac{1}{\alpha}}} \Theta\left(\sqrt{\frac{n}{Kk^{\alpha}\log n}}\right)$.
- When $\alpha \geq 2$, we have $\Lambda = \left(\sqrt{\frac{n}{K \log n}}\right)$ for $\alpha > 2$ and $\Theta\left(\sqrt{\frac{n \log K}{K \log n}}\right)$ for $\alpha = 2$.
- When $1 < \alpha < 2$, Λ can be expressed as $\Lambda = \Theta\left(\sqrt{\frac{n}{K \log n}}\right) \cdot K^{1-\frac{\alpha}{2}} = \Theta\left(\sqrt{\frac{n}{\log n}} K^{\frac{1-\alpha}{2}}\right)$.

Similarly, we can obtain the results in the case where $\alpha = 1$, $0 < \alpha < 1$ and $\alpha = 0$ using the same method above. The results of Λ corresponding to different α are shown in Equation (20). This completes our proof. ■

D. LARGE DATA CATALOGUE SIZE

A counterpart of limited data catalogue is the case of large data catalogue size, where the total number of data classes satisfies $K = \Omega(n)$. And the K classes are uniformly stored at n users. More precisely, on average, each user has $M = K/n$ data classes. For a specific user i, denote $\{k_{i_j}\}|_{j=1}^{M}$ as the data classes stored at it. Then, for another arbitrarily chosen user j storing data classes $\{k_{j_l}\}|_{l=1}^{M}$, we assume that $\{k_{i_l}\}|_{j=1}^{M} \bigcap \{k_{j_l}\}|_{l=1}^{M} = \phi$, which implies that each of the K classes is possessed by one and only one user. And we restrict our consideration here to one data item within each class. A finer resolution of different data items within each class will be taken into account as our future work.

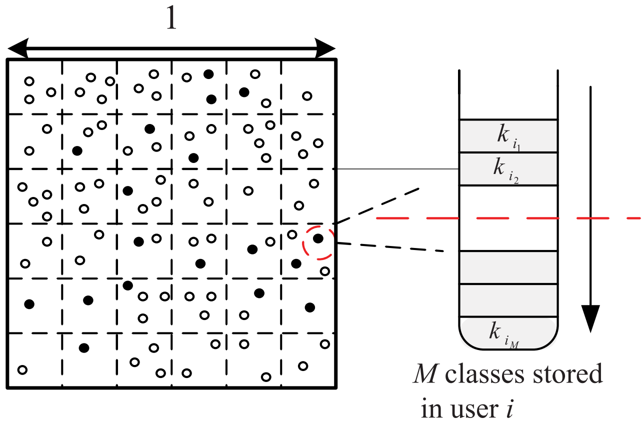


FIGURE 6. Illustration of the case of large data catalogue, where $K = \Omega(n)$ and each of the n users stores $M = K/n$ non-overlapping classes. The red dotted line represents a threshold k_M on the ranking of data classes stored at a generic node. Data classes with their ranking below k_M will have zero interester.

As can be derived using Lemma 2, the estimated number of interesters $N_{k_{ij}}$ on data class j stored at user i is $\Theta(n \cdot H/k_{ij}^{\alpha})$. Taking the summation of $N_{k_{ij}}$ on all $i \in [1, \ldots, n]$ and $j \in [1, \ldots, M]$, we have

$$\sum_{i=1}^{n}\sum_{j=1}^{M} N_{k_{ij}} = \sum_{i=1}^{n}\sum_{j=1}^{M} n\left(\frac{H}{k_{ij}^{\alpha}}\right) = n. \tag{21}$$

More precisely speaking, a data source in large data catalogue needs to act as multiple sources to deliver data from different classes to their corresponding interesters. However, due to the existence of k_M, some data classes ranking behind k_M have zero interesters (see Figure 7 for illustration) and therefore should be excluded in throughput calculation. From Lemma 14, we can obtain the k_M in this case, as is shown in Lemma 17.

Lemma 17: For $K = \Omega(n)$, the maximum data class k_M that can guarantee $N_k > 1$ is

$$k_M = \begin{cases} O\left(n^{\frac{1}{\alpha}}\right) & \alpha > 1 \\ O\left(\frac{n}{\log K}\right) & \alpha = 1 \\ O\left(n^{\frac{1}{\alpha}} K^{\frac{\alpha-1}{\alpha}}\right) & 0 \leq \alpha < 1 \,\&\, n = \Omega(K^{1-\alpha}) \\ 0 & 0 \leq \alpha < 1 \,\&\, n = O(K^{1-\alpha}). \end{cases} \tag{22}$$

Lemma 17 implies that k_M never exceeds n regardless of the fact that the total number of data classes K is much larger than n. An observation from this lemma is that when $0 \leq \alpha < 1$, to guarantee $k_M > 1$ we must have $K > n^{\frac{1}{1-\alpha}}$. Hence, there is no interester for any of the K data classes when $K > n^{\frac{1}{1-\alpha}}$ and leads to zero throughput in such case.

The upper bound of Λ_k can still be derived in the similar manner as previous two cases. Also, the throughput-achieving routing scheme can follow the main idea of that in case 2, with some minor modifications: Considering the fact that K/n data classes stored at each user, with at most one of the K/n classes having non-zero interesters, when a user is scheduled to transmit, he only picks out the one with the highest ranking stored in his buffer and then transmits it. The rest data classes stored at the buffer can be simply ignored for that they will hardly attract interesters. Hence, the same result can be informed by analyzing the scenario where there are totally k_M data classes and each of these k_M classes is possessed by one and only one user. As a consequence, we can obtain Λ_k and Λ shown in Theorem 3.

Theorem 3: In large data catalogue, the user-centric throughput Λ_k for data class k is shown in Table 3. And the aggregate average user-centric throughput Λ can be expressed as

$$\Lambda = \begin{cases} \Theta(1) & \alpha > 2 \\ \Theta\left(\frac{n^{\frac{1}{\alpha}-\frac{1}{2}}}{\sqrt{\log n}}\right) & 1 < \alpha \leq 2 \\ \Theta\left(\frac{\sqrt{n}}{\sqrt{\log n \log K}}\right) & \alpha = 1 \\ \Theta\left(\frac{n^{\frac{1}{\alpha}-\frac{1}{2}} K^{1-\frac{1}{\alpha}}}{\sqrt{\log n}}\right) & 0<\alpha<1 \& n=\Omega(K^{1-\alpha}) \\ 0 & 0<\alpha<1 \& n=O(K^{1-\alpha}). \end{cases} \tag{23}$$

TABLE 3. User-centric throughput vs α and k.

α	k	Λ_k
$\alpha > 2$	$[1, \ldots, n^{\frac{1}{\alpha}}]$	$\Theta\left(\frac{1}{k^\alpha}\right)$
$1 < \alpha \leq 2$	$\left[\Omega\left(\log^{\frac{1}{\alpha}} n\right), O\left(n^{\frac{1}{\alpha}}\right)\right]$	$\Theta\left(\frac{1}{\sqrt{k^\alpha \log n}}\right)$
	$O\left(\log^{\frac{1}{\alpha}} n\right)$	$\Theta\left(\frac{1}{k^\alpha}\right)$
$\alpha = 1$	$\left[1, O\left(\frac{n}{\log K}\right)\right]$	$\Theta\left(\frac{n}{K\sqrt{k \log n \log K}}\right)$
$0 \leq \alpha < 1$	$O\left(n^{\frac{1}{\alpha}} K^{\frac{\alpha-1}{\alpha}}\right)$	$\Theta\left(\sqrt{\frac{K^{\alpha-1}}{k^\alpha \log n}}\right)$
	$[1, \ldots, k_M]$	0

Proof: According to Lemma 17, the total valid number of data classes is $k_M < n$ for all $\alpha \geq 0$. Since the K data classes are randomly and uniformly stored among n users, the k_M valid data classes are also randomly and uniformly stored among n users. That is to say, each user stores on average $\Theta(K/n)$ data classes with at most one valid class among them. Since the invalid data classes contribute zero throughput to the whole network, we can focus mainly on the contribution from those k_M classes of the throughput. This is equivalent to the scenario where there are n users with totally k_M data classes and each of these k_M classes is stored by one and only one source. Then, we can obtain the Λ_k and Λ using the similar analysis in case 2. ■

VI. DISCUSSION

A. IMPACT OF α ON THROUGHPUT

Although mainly of theoretical interest, we believe that our work can provide fundamental principles to smart design of user-centric network architecture. We first discuss about the impact of user-centric data distribution on network throughput. In all three scenarios, a common phenomenon from our results is that the aggregate user-centric throughput Λ is reduced to a constant in order sense when $\alpha > 2$ and decreases gradually as α increases from 1 to 2. This is because users' interests are highly skewed to popular data when $\alpha > 2$. The preference on popular data makes it similar to the broadcast case where the data from each source needs to be disseminated to all the other nodes in the network.

Another significant phenomenon is that the maximum throughput can be achieved at $\alpha = 1$ in all three cases. Note that this outperforms the throughput at $0 \leq \alpha < 1$, where each user's interest is flattened towards the uniform over all K data classes. The reason behind is that the average source-destinations' transmission length, a key factor that determines throughput in static networks is minimized when $\alpha = 1$. A useful heuristic can be brought about is that throughput improvement can be achieved by the existence of diverse users' interests in real network applications.

B. IMPACT OF K ON THROUGHPUT

Data catalogue size (say K) also has a strong impact on user-centric throughput performance. And the graphical results shown in Figures 1-3 all indicate that changing K can affect the aggregate throughput results Λ, in the range $\alpha \in [0, \ldots, 2]$. In the case of limited data catalogue, when each data class is possessed by only one source (Figure 1), the throughput Λ increases linearly with $\sqrt{K}$ in the range $\alpha \in [0, 1]$ and reaches the maximum throughput when $K = \Theta(n)$; In the range $\alpha \in [1, 2]$, there is a critical threshold of $K = n^{\frac{1}{\alpha}}$, above which the throughput jumps to $n^{\frac{1}{\alpha}-\frac{1}{2}}$, independent of K; When each data class contains multiple data items (Figure 2), the throughput remains to be $\sqrt{n}$ when $\alpha \in [0, 1)$, regardless of how K varies. Because users exhibit identical interest in each of the K data classes when $\alpha \in [0, 1)$. For each data class, users cast the same interest at each of the n/K data belonging to that class. Overally, users can be treated as having identical interest in each of them given n data items totally, which is equivalent to unicast traffic pattern. On the contrary, the throughput decreases as we increases K when $\alpha > 1$, since a larger K will impose more users' concentration on data from different classes but dilute users' identical interests among data items within the same class. In the case of large data catalogue (Figure 3), the aggregate throughput Λ is independent of how K varies but only constrained by the number of users n when $\alpha \in [1, 2]$. In contrast, Λ decreases sharply as K increases when $\alpha \in [0, 1)$, given the constraint that $K = O(n^{\frac{1}{1-\alpha}})$. The throughput reduces to zero if $K = \Omega(n^{\frac{1}{1-\alpha}})$. The reason is that the users show identically very low interests in each of the K classes in too large data catalogue, resulting in almost zero bandwidth utility.

C. RELATIONSHIP WITH THROUGHPUT RESULTS IN PREVIOUS WORKS

Now we will associate user-centric throughput obtained in our work with those in previous works. Note that most of previous works are based on the network-centric scenario, where whether a user is interested in the data he encounters is not taken into account. The typical aggregate unicast throughput is already shown to be $\Theta(\sqrt{n})$ in [25], where each source is assigned to send data to $\Theta(1)$ destinations. Note that the same throughput of $\Theta(\sqrt{n})$ is achievable in our work, where $\omega(1)$ users can be treated as destinations of a data item. To this regard, both data dissemination and bandwidth utility are exploited in a more efficient way in our paper. Because the data can be disseminated to as many interesters as possible, resulting in little extra bandwidth wasted in transferring data to lots of users not interested in it.

The aggregate throughput for classic multicast is demonstrated to be $\Theta\left(\sqrt{\frac{n}{k\log n}}\right)$ in static network [30], where k denotes the number of destinations for each source. Note that the throughput of $\Theta(\sqrt{n})$ can be achieved only when $k = \Theta(1)$ and the multicast throughput reduces to $\Theta(1)$ when $k = \Theta(n)$. Our results outperform such multicast throughput in that a maximum throughput $\Theta(\sqrt{n})$ can be achieved provided that $\omega(1)$ users can receive the data they are interested in. Furthermore, the user-centric throughput can remain to be $\Theta(\sqrt{n})$ regardless of data catalogue size. In the case of large data catalogue, throughput of $\Theta(\sqrt{n})$ remains unchanged even if the total number of data classes available is much more larger than n. This brings about a heuristic that we can introduce as many data classes as possible to the network without sacrifice on throughput.

Note that here the throughput gain refers to the number of extra destinations that data can be delivered to given the same throughput. The gain is attributed to the efficient user-centric data dissemination where data is delivered to the users that are interested in it.

VII. CONCLUSION

This paper analyzes the throughput of user-centric data dissemination in large scale ad hoc wireless networks, which considers data transmission from perspective of users' interests. We show that users' interests can drastically improve upon existing scaling laws established under the assumption that users show the same preference in any data they encounter. More precisely, we consider the scenario in which each data source estimates the interesters of its data based on Zipf-distributed user interest probability. For such user-centric model, we study throughput performance and present closed-form expression of user-centric throughput versus different value of α and the total number of data classes K available. Notably, for $\alpha = 1$ the maximum capacity of $\Theta(\sqrt{n})$ (except for a poly-logarithmic factor) can be achieved for all three situations, under appropriate choice for K in each situation.

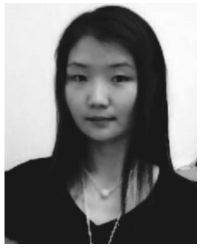
REFERENCES

[1] *YouTube*. Accessed: 2017. [Online]. Available: http://www.youtube.com
[2] S. Bausch and L. Han. (2006). *YouTube U.S. Web Traffic Grows 75 Percent Week over Week*. [Online]. Available: http://www.nielsen-netratings.com/pr/pr0607212.pdf
[3] M. Cha, H. Kwak, P. Rodriguez, Y. Y. Ahn, and S. Moon, "I tube, you tube, everybody tubes: Analyzing the world's largest user generated content video system," in *Proc. ACM SIGCOMM IMC*, 2007, pp. 1–14.
[4] S. Dann, "Benchmarking micro-blog performance: Twitter content classification framework," in *Maximizing Commerce and Marketing Strategies Through Micro-Blogging*. Hershey, PA, USA: IGI Global, 2015, pp. 313–332.
[5] G. Liu, H. Shen, and L. Ward, "An efficient and trustworthy P2P and social network integrated file sharing system," *IEEE Trans. Comput.*, vol. 64, no. 1, pp. 54–70, Jan. 2015.
[6] H. Zhou, H. Wang, X. Li, and V. Leung, "A survey on mobile data offloading technologies," *IEEE Access*, vol. 6, no. 1, pp. 5101–5111, 2018.
[7] K. Chen, H. Shen, K. Sapra, and G. Liu, "A social network based reputation system for cooperative P2P file sharing," *IEEE Trans. Parallel Distrib. Syst.*, vol. 26, no. 8, pp. 2140–2153, Aug. 2015.
[8] E. Bakshy, S. Messing, and L. A. Adamic, "Exposure to ideologically diverse news and opinion on Facebook," *Science*, vol. 348, no. 6239, pp. 1130–1132, 2015.
[9] M. Xiao, J. Wu, and L. Huang, "Community-aware opportunistic routing in mobile social networks," *IEEE Trans. Comput.*, vol. 63, no. 7, pp. 1682–1695, Jul. 2014.
[10] H. Hu, Y. Wen, T. S. Chua, J. Huang, W. Zhu, and X. Li, "Joint content replication and request routing for social video distribution over cloud CDN: A community clustering method," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 26, no. 7, pp. 1320–1333, Jul. 2016.
[11] L. Yao, Y. Man, Z. Huang, J. Deng, and X. Wang, "Secure routing based on social similarity in opportunistic networks," *IEEE Trans. Wireless Commun.*, vol. 15, no. 1, pp. 594–605, Jan. 2016.
[12] H. Zhou, C. M. V. Leung, C. Zhu, S. Xu, and J. Fan, "Predicting temporal social contact patterns for data forwarding in opportunistic mobile networks," *IEEE Trans. Veh. Technol.*, vol. 66, no. 11, pp. 10372–10383, Nov. 2017.
[13] C. Zhu, H. Zhou, V. C. M. Leung, K. Wang, Y. Zhang, and L. T. Yang, "Toward big data in green city," *IEEE Commun. Mag.*, vol. 55, no. 11, pp. 14–18, Nov. 2017.
[14] W. Gao and G. Cao, "User-centric data dissemination in disruption tolerant networks," in *Proc. IEEE INFOCOM*, Apr. 2011, pp. 3119–3127.
[15] L. Fu, W. Huang, X. Gan, F. Yang, and X. Wang, "Capacity of wireless networks with social characteristics," *IEEE Trans. Wireless Commun.*, vol. 15, no. 2, pp. 1505–1516, Feb. 2016.
[16] B. Azimdoost, H. R. Sadjadpour, and J. J. Garcia-Luna-Aceves, "Capacity of wireless networks with social behavior," *IEEE Trans. Wireless Commun.*, vol. 12, no. 1, pp. 60–69, Jan. 2013.
[17] Z. Qin, X. Gan, J. Wang, L. Fu. and X. Wang, "Capacity of social-aware wireless networks with directional antennas," *IEEE Trans. Commun.*, vol. 65, no. 11, pp. 4831–4844, Nov. 2017.
[18] J. Nielsen. (1997). *Zipf Curves and Website Popularity*. [Online]. Available: www.useit.com/alertbox/zipf.html
[19] T. Qiu, Z. Ge, S. Lee, J. Wang, Q. Zhao, and J. Xu, "Modeling channel popularity dynamics in a large IPTV system," in *Proc. ACM SIGMETRICS*, Seattle, WA, USA, Jun. 2009, pp. 275–286.
[20] J. Stribling *et al.*, "Flexible, wide-area storage for distributed systems with WheelFS," in *Proc. NSDI*, Boston, MA, USA, 2009, pp. 43–58.
[21] V. Shah and G. D. Veciana, "Performance evaluation and asymptotics for content delivery networks," in *Proc. IEEE INFOCOM*, Apr./May 2014, pp. 2607–2615.
[22] F. Chang *et al.*, "Bigtable: A distributed storage system for structured data," in *Proc. OSDI*, 2006, p. 15.
[23] S. Rhea *et al.*, "Maintenance-free global data storage," *IEEE Internet Comput.*, vol. 5, no. 5, pp. 40–49, Sep./Oct. 2001.
[24] F. Dabek, J. Li, E. Sit, J. Robertson, M. F. Kaashoek, and R. Morris, "Designing a DHT for low latency and high throughput," in *Proc. NSDI*, Mar. 2004, p. 7.
[25] P. Gupta and P. R. Kumar, "The capacity of wireless networks," *IEEE Trans. Inf. Theory*, vol. 46, no. 2, pp. 388–404, Mar. 2000.
[26] M. J. Neely and E. Modiano, "Capacity and delay tradeoffs for ad-hoc mobile networks," *IEEE Trans. Inf. Theory*, vol. 51, no. 6, pp. 1917–1937, Jun. 2005.

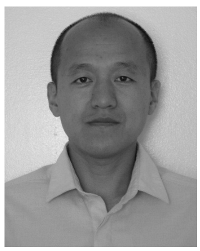
[27] X. Lin and N. Shroff, "The fundamental capacity-delay tradeoff in large mobile ad hoc networks," in *Proc. 3rd Annu. Med-Hoc-Netw. Workshop*, 2004.
[28] P. Li, Y. Fang, and J. Li, "Throughput, delay, and mobility in wireless ad hoc networks," in *Proc. IEEE INFOCOM*, San Diego, CA, USA, Mar. 2010, pp. 1–9.
[29] M. Garetto and E. Leonardi, "Restricted mobility improves delay-throughput tradeoffs in mobile ad hoc networks," *IEEE Trans. Inf. Theory*, vol. 56, no. 10, pp. 5016–5029, Oct. 2010.
[30] X. Li, "Multicast capacity of wireless ad hoc networks," *IEEE/ACM Trans. Netw.*, vol. 17, no. 3, pp. 950–961, Jun. 2009.
[31] X. Wang, W. Huang, S. Wang, J. Zhang, and C. Hu, "Delay and capacity tradeoff analysis for motioncast," *IEEE/ACM Trans. Netw.*, vol. 19, no. 5, pp. 1354–1367, Oct. 2011.
[32] K. Lee, Y. Kim, S. Chong, I. Rhee, Y. Yi, and N. B. Shroff, "On the critical delays of mobile networks under Lévy walks and Lévy flights," *IEEE/ACM Trans. Netw.*, vol. 21, no. 5, pp. 1621–1635, Oct. 2013.
[33] W. Liu, K. Lu, J. Wang, Y. Qian, T. Zhang, and L. Huang, "Capacity of distributed content delivery in large-scale wireless ad hoc networks," in *Proc. IEEE INFOCOM*, Orlando, FL, USA, Mar. 2012, pp. 19–27.
[34] S. Gitzenis, G. S. Paschos, and L. Tassiulas, "Asymptotic laws for joint content replication and delivery in wireless networks," in *Proc. IEEE INFOCOM*, Orlando, FL, USA, Mar. 2012, pp. 2760–2776.
[35] G. Alfano, M. Garetto, and E. Leonardi, "Content-centric wireless networks with limited buffers: When mobility hurts," in *Proc. IEEE INFOCOM*, Apr. 2013, pp. 1815–1823.
[36] H. Rahul, S. Kumar, and D. Katabi, "MegaMIMO: Scaling wireless capacity with user demands," in *Proc. ACM SIGCOMM*, 2012, pp. 235–246.
[37] J. M. Steele, "Growth rates of Euclidean minimal spanning trees with power weighted edges," *Ann. Probability*, vol. 16, no. 4, pp. 1767–1787, Oct. 1988.
[38] K. Lin, W. Shen, M. S. Chen, and K. Tan, "User-centric network MIMO with dynamic clustering," *IEEE/ACM Trans. Netw.*, vol. 25, no. 3, pp. 1910–1923, Jun. 2017.
[39] S. Chen, F. Qin, B. Hu, X. Li, and Z. Chen, "User-centric ultra-dense networks for 5G: Challenges, methodologies, and directions," *IEEE Wireless Commun.*, vol. 23, no. 2, pp. 78–85, Apr. 2016.
[40] A. V. Vasilakos, Z. Li, G. Simon, and W. You, "Information centric network: Research challenges and opportunities," *J. Netw. Comput. Appl.*, vol. 52, pp. 1–10, Jun. 2015.

XIAOYING GAN received the Ph.D. degree in electronic engineering from Shanghai Jiao Tong University, Shanghai, China, in 2006. From 2009 to 2010, she was a Visiting Researcher with the California Institute for Telecommunications and Information, University of California at San Diego, La Jolla, CA, USA. She is currently an Associate Professor with the Department of Electronic Engineering, Shanghai Jiao Tong University. Her current research interests include network economics, social aware networks, heterogeneous cellular networks, multiuser multi-channel access, and dynamic resource management.

ZHIDA QIN received the B.S. degree from the Huazhong University of Science and Technology, Wuhan, China, in 2014. He is currently pursuing the Ph.D. degree in electronic engineering with Shanghai Jiao Tong University, Shanghai, China. His current research interests include wireless network capacity, social aware networks, and information diffusion.



LUOYI FU received the B.E. degree in electronic engineering and the Ph.D. degree in computer science and engineering from Shanghai Jiao Tong University, China, in 2009 and 2015, respectively. She is currently an Assistant Professor with the Department of Computer Science and Engineering, Shanghai Jiao Tong University. Her research interests are in the areas of social networking and big data, scaling laws analysis in wireless networks, connectivity analysis, and random graphs.



XINBING WANG received the B.S. degree (Hons.) from the Department of Automation, Shanghai Jiao Tong University, Shanghai, China, in 1998, the M.S. degree from the Department of Computer Science and Technology, Tsinghua University, Beijing, China, in 2001, and the Ph.D. degree (major from the Department of Electrical and Computer Engineering and minor from the Department of Mathematics) from North Carolina State University, Raleigh, NC, USA, in 2006. He is currently a Professor with the Department of Electronic Engineering, Shanghai Jiao Tong University. He is a member of the technical program committees of several conferences, including the IEEE INFOCOM from 2009 to 2014, the ACM MobiCom 2012, and the ACM MobiHoc 2012 and 2013. He has been an Associate Editor of the IEEE/ACM Transactions on Networking and the IEEE Transactions on Mobile Computing.