The Virtue of Patience: Offloading Topical Cellular Content through Opportunistic Links

Wei Peng*, Feng Li†, Xukai Zou*, and Jie Wu‡
*Department of Computer and Information Science
†Department of Computer, Information, and Leadership Technology
Indiana University-Purdue University Indianapolis, Indianapolis, IN, U.S.A.
‡Department of Computer and Information Sciences
Temple University, Philadelphia, PA, U.S.A.

Abstract*—Mobile data offloading is an approach to alleviating overloaded cellular traffic through alternative communication technologies on smartphones. Inspired by the prospect of spontaneous, peer-assisted, bulk data transfer through NFC or Wi-Fi Direct between proximate users' smartphones, we propose a model for mobile data offloading through the opportunistic proximity (e.g., Wi-Fi Direct) links with bounded content delivery delay and differential interests in content. Unlike the previous formulation of mobile data offloading as a target-set selection problem, which, essentially, asks the question "who (will download the content through the cellular link)," we ask "who" and "when." We present methods for individual users to locally estimate (their and their acquaintances') topological importance on the opportunistic proximity-link-based networks and aggregated interests in content. These factors are consolidated into a time-dependent function that embodies the concept of users' *patience* for the content. Each individual user, then, periodically make a probabilistic cellular download decision based on its patience at that time. Our motivation and insights are: 1) Involving topologically important, but otherwise disinterested, users in downloading and forwarding content helps improve offloading efficiency; 2) situation awareness embodied in the time-dependent patience function is desirable, since it allows users to react to hard-to-predict contact opportunities on the fly. Through trace-driven simulations, we corroborate our insights, and demonstrate the effectiveness of our proposed method in reducing cellular costs.

***Index Terms*—mobile data offloading, probabilistic algorithm, distributed algorithm, ego-centric betweenness centrality, interest aggregation, patience.**

I. Introduction

The cellular infrastructure is overloaded by an expanding user base and increasing bandwidth demand from smartphone applications. Indeed, driven primarily by smartphones, AT&T's wireless data traffic has grown 20000% over the five years between 2007 and 2011 [1].

Mobile data offloading, or mobile cellular traffic offloading, exploits alternative communication technologies on smartphones, and user mobility, to deliver information originally scheduled for transmission over the cellular networks. Previous works [2, 3, 4] demonstrate the feasibility of offloading cellular traffic by peer-to-peer assisted forwarding through Bluetooth. Recent developments in communication technology, embodied in the latest smart mobile devices (including Google Nexus 7 [5] and the rumored iPhone 5 [6]) that support NFC [7] and Wi-Fi Direct [8], makes spontaneous bulk data transfers between proximate users a reality. Furthermore, the current data usage cap and tiered pricing model [9] incentivizes smartphone users to offload their cellular data. These developments make further research in mobile data offloading relevant and worthwhile.

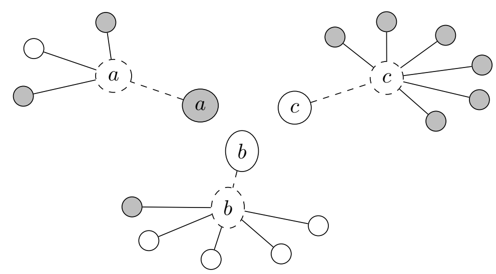


Fig. 1: Users' interests in content complicates the offloading strategy. Shaded nodes represent interested users; solid lines link acquaintances; dashed lines and nodes represent nodes' mobility.

Inspired by this vision, in our paper, we study the problem of offloading cellular traffic through proximity-based links (proximity link, for short) such as Wi-Fi Direct. In our model, we include a factor that was missing in existing mobile data offloading models: users' interests in content. Users' interests are particularly relevant for large-scale networks: Nobody desires (or is able) to consume all generated content. This lies behind the quest for better search engines and the rise of social taxonomy, or folksonomy, in tagging content. Additionally, we consider bounded delivery-delay tolerance to model the general case where the content, though having no hard real-time requirement, still needs to be delivered before too long, lest it becomes stale.

Figure 1 illustrates the complication brought by users' interests: When a, b, and c meet through a proximity link and if, due to limited budget, one and only one of them will download a piece of content through the cellular link: Who, among them, should download? Though b has more acquaintances than a does and therefore, in some sense, is more socially important, few of b's acquaintances are interested in the content, when compared with those of a: It is more cost effective for a to download and carry the content than

b. In another comparison, c is more socially important than a, and most of c's acquaintances are interested in the content: Though c is not interested in the content, if c downloads and carries the content, c can serve more users within a reasonable time than a can. In general, a cost-effective offloading strategy involves an interplay between users' interests and their social importance.

In addition to deciding *who* shall download the content through cellular links, as in the target-set selection formulation [4], we ask *when*. To appreciate the benefits of including time in the model, we consider a few scenarios.

- Every user downloads his[1] interested content through the cellular link immediately after the content is released. No offloading through the proximity link takes place in this case. This is the baseline *diligent* strategy that mobile data offloading measures against.
- Every user initially waits, in the hope that someone will download the content and forward the content (through the proximity link) to him through one of his acquaintances. However, nobody will receive the content, since nobody has downloaded it. Even if the content is eventually downloaded by some random user, and is forwarded to other interested users, it may have expired. This is the *lazy* strategy that introduces an unacceptably long delay.
- Some well-connected, or socially important, users, whose acquaintances are interested in the content, download the content through the cellular link, and forward the content to their acquaintances when they meet through the proximity link. As time passes by, and the risk of the content becoming stale increases, those users who have not received their interested content through either link become impatient in waiting, and eventually download the content through the cellular link if the content has still not been received after a long delay. This *adaptive* strategy is neither too diligent nor too lazy, and provides a trade-off between cellular traffic load and content delivery delay.

A challenge is to design such an adaptive strategy without resorting to central scheduling and coordination through the cellular link, which is costly and less scalable. Although human mobility exhibits patterns [10, 11], contact opportunities are hard to predict precisely. Therefore, effective central scheduling and coordination require prohibitively costly updating.

We address the challenge as follows. Users estimate their relative social importance in the dynamic, opportunistic, proximity-link-based network with a weighted ego-centric betweenness centrality metric (Equation (2)); users estimate their (and their acquaintances') aggregated interests (Equation (3)) based on their chances of meeting each other (Equation (1)); users use a function (Equation (4)), which embodies the concept of users' *patience* for the content, to consolidate users' social importance with aggregated interests. This function gives rise to a probabilistic cellular offloading strategy (Equation (5)) that assigns a cellular download probability to a user, according to his capability to help offload the topical cellular content. Users then periodically decide whether to download the content through the cellular link by their patience *at that time*.

[1]"He" ("his") is to be read "he/she" ("his/her") henceforth.

Thus, our solution is *social*, *content*, and *situation*-aware: Involving topologically important, but otherwise disinterested, users in downloading and forwarding content will help reduce the cellular traffic and improve the offloading efficiency, while satisfying users' content demand.

In the following sections, we formulate the problem (Section II), describe the design of our patience-based cellular offloading strategy (Section III), analyze its properties (Section IV), and complement the analysis with trace-driven simulations (Section V). Works that inspire ours are summarized in Section VI.

II. Model

Consider a group of smartphone users: Each user has a smartphone that can access the Internet through the cellular link, and connect with nearby smartphones through some proximity link. For example, Bluetooth is the current standard for linking proximate devices: Han et al. evaluated and confirmed the feasibility of using Bluetooth to offload cellular data [4]. Other possibilities include Wi-Fi co-location (two users connect to the same Wi-Fi access point) and the upcoming Wi-Fi Direct. The cellular link is persistent but expensive, while the proximity link is opportunistic but free: A smartphone can access the Internet immediately on demand through the cellular link, while two smartphones can connect with each other through the proximity link only when they are nearby. An opportunity for two smartphones to connect through the proximity link is an encounter between them.

The users are interested in some content that is generated and released by some publisher on the Internet. Each piece of content is tagged before its release. We model two aspects of the user-content relationship: users' interests and content's freshness. For a user u:

- The interests of u are represented by a set of tags I_u: u is interested in a piece of content if the content has a tag in I_u.
- For a tag g, u prefers to receive a piece of content with tag g, within a delay of up to f_g. Otherwise, if u has not received the content within that time, the content becomes stale for u. We call f_g the freshness of tag g.

The publisher publishes an aggregate feed, containing the summary, the tags, and the download link for each piece of released content. A user is notified through the feed when new content is released.

The problem is to find a localized strategy that minimizes the number of cellular downloads (which incur costs) while maximizing fresh content deliveries. A localized strategy is one in which each user makes decisions based on information obtained through encounters rather than requiring (global) coordination through the cellular link. Communication through cellular link incurs cost, and keeping track of local changes for global coordination aggravates the problem. In comparison, a localized strategy is cost-effective and adaptive. The decisions to be made include whether to download a piece of content through the cellular link and, if the answer is yes, when.

Before moving on to describe the concrete design, we make our assumptions explicit. Since the proximity link is virtually free, routing on the proximity-link-based network is not a focus of this paper: To maximize coverage, the content is epidemically forwarded across the opportunistic proximity-link-based network once it is downloaded through the cellular link. A more sophisticated forwarding strategy [12] can be adopted, but is beyond the scope of this paper. The users are honest and cooperative. In other words, each user will follow the protocols by honestly sharing information and cooperatively reducing the overall cost:

- A user will honestly report their interests to others upon request.
- If downloading, storing, and forwarding a piece of content will reduce the overall cellular cost of the whole network, a user will do it even if he is not interested in the content.

Enforcement and incentive [9, 13, 14], while important, are orthogonal to the current problem and are left for further studies.

III. Design

A. Overview

Intuitively, two groups of users are favored for directly downloading a piece of content with tag g through the cellular link:

- Those who are interested in g and meet with users who are interested in g;
- Those who are socially important, or equivalently, topologically important in the dynamic proximity-link-based network.

The rationale for favoring the first group is obvious: Those users have better chances of directly obtaining or forwarding the content to interested users. However, the scope of this case is restricted to direct acquaintances and is thus oblivious of the topology of the proximity-link-based network, for which the second case tries to remedy. The membership in the two groups may overlap; those who are members of both groups are favored over those who are members of only one group.

Concretely, a user decides his topological importance in the dynamic proximity-link-based network by locally computing his weighted ego-centric betweenness centrality (Section III-C). Along with the aggregated interest of both himself and his acquaintances (Section III-D), the user determines his patience for the content and periodically decides, with a temporal-dependent probability based on his patience, whether to download the content through the cellular link if he has not yet received the content (Section III-E).

In this section, we focus on the design details. Discussions on intuition and rationale are deferred to Section IV.

B. Temporal tie strength

Let the set of users that u has met through the proximity link be U_u: U_u is the set of u's acquaintances. For $v \in U_u$, let the chronologically ordered sequence of encounters between u and v be $[a_1, b_1], [a_2, b_2], \ldots, [a_k, b_k]$, and the current time be t; the average interval between consecutive encounters $\hat{s}_u(v)$ is defined as:

$$\hat{s}_u(v) = \begin{cases} \dfrac{(t - b_k) + \sum_{i=1}^{k-1}(a_{i+1} - b_i)}{k} & u \text{ and } v \text{ have met.} \\ +\infty & \text{otherwise.} \end{cases}$$

By definition, $\hat{s}_u(v)$ is symmetric: $\hat{s}_u(v) = \hat{s}_v(u)$; $\hat{s}_u(v) \geq 0$; u can locally compute $\hat{s}_u(v)$ for all $v \in U_u$.

Based on $\hat{s}_u(v)$, the temporal tie strength (tie for short) $s_u(v)$ is defined as:

$$s_u(v) = \begin{cases} \exp(-\alpha_s \hat{s}_u(v)) & s_u(v) \in [0, +\infty), \\ 0 & s_u(v) = +\infty, \end{cases} \tag{1}$$

in which $\alpha_s > 0$ is a scaling parameter, adapting to the given scenario, to prevent the tie $s_u(v)$ from dropping too fast with the increase of the average inter-encounter interval $\hat{s}_u(v)$.

Greater $s_u(v)$ corresponds to stronger tie between u and v; $s_u(v) \in [0, 1]$. Like $\hat{s}_u(v)$, $s_u(v)$ is symmetric: ($s_u(v) = s_v(u)$) and u can locally compute $s_u(v)$ for all $v \in U_u$.

C. Weighted ego-centric betweenness centrality

For $v, w \in U_u$, u can obtain $\hat{s}_u(v)$, $\hat{s}_u(w)$, and $\hat{s}_v(w)$ (or, equally, $\hat{s}_w(v)$) during their encounters. u can construct his neighborhood graph G_u, of which nodes are $\{u\} \cup U_u$ and the edge between $v, w \in U_u \cup \{u\}$ has a weight $\hat{s}_v(w) = \hat{s}_w(v)$ if $\hat{s}_w(v) \neq +\infty$. For $v, w \in U_u$ and $v \neq w$, let $p(v, w)$ be the proposition "(v, u, w) is a shortest path between v and w"; this can be determined by, for example, the Dijkstra's algorithm [15]. From G_u, u can compute a weighted ego-centric betweenness centrality β_u:

$$\beta_u = \begin{cases} \dfrac{\sum\limits_{v, w \in U_u, v \neq w} [p(v, w)]}{2\binom{|U_u|}{2}} & |U_u| \geq 2, \\ 0 & \text{otherwise.} \end{cases} \tag{2}$$

In Equation (2) and the following, when p is a proposition, the notation $[p]$ is the propositional indicator function:

$$[p] = \begin{cases} 1 & p \text{ is true,} \\ 0 & p \text{ is false.} \end{cases}$$

From Equation (2), $\beta_u \in [0, 1]$.

D. Interest aggregation

User u records the interests I_v of user v when they meet. u's aggregated interest $i_u(g)$ on tag g is:

$$i_u(g) = [g \in I_u] + \sum_{v \in U_u} s_u(v)[g \in I_v]. \tag{3}$$

$i_u(g) \geq 0$; $i_u(g) < 1$ only if $g \notin I_u$.

E. Patience and the probabilistic cellular downloading strategy

From the centrality β_u (Equation (2)) and aggregated interest $i_u(g)$ on tag g (Equation (3)), u determines his patience $p_{u,g}$ for tag g as a function:

$$p_{u,g}: \quad [0,1] \to [0,1], \tag{4}$$

defined as (for two scaling parameters $\alpha_i > 0$ and $\alpha_\beta > 1$, which correspond to the interest aggregation $i_u(g)$ and the centrality β_u, respectively):

$$p_{u,g}(x) = \begin{cases} \left(1 - e^{-\alpha_i i_u(g)}\right) x^{\alpha_\beta^{(1-2\beta_u)}} & g \in I_u, \\ \left(1 - e^{-\alpha_i i_u(g)}\right) (1-x)^{\alpha_\beta^{(1-2\beta_u)}} & g \notin I_u. \end{cases} \tag{5}$$

The patience function defined by Equations (4) and (5) gives u a strategy to make cellular download decision. u, according to the strategy and based on the situation at that time (Have u received the content? How close to the content expiration?), periodically (at a pre-defined interval for all users) makes a probabilistic cellular download decision as follows. At the moment $t + x \cdot f_g$ ($x \in [0,1]$) between the time t that u first learns about a piece of content with tag g and the time $t + f_g$ that the content becomes stale for u, u flips a biased coin and, with a probability $p_{u,g}(x)$, downloads the content through the cellular link. As a stipulation, if u is interested in the content himself, but has neither downloaded nor received the content by the time $t + f_g$, u will download the content directly through the cellular link to satisfy his content demand.

IV. Analysis

In this section, we take a closer look at the various parts of our design and how they fit together to make an efficient mobile data offloading strategy. Our agenda is to discuss the intuition behind the design and show the following: The probabilistic cellular downloading strategy based on the patience function defined by Equation (5) behaves in intuitively desirable ways.

A. Probabilistic cellular downloading strategy based on patience

We take the patience function $p_{u,g}(x)$ defined in Equation (5) apart:

- The maximal probability that u will download the content through the cellular link *in one round* is $1 - e^{-\alpha_i i_u(g)}$, which is monotonically increasing on $i_u(g)$: Greater aggregated interest $i_u(g)$ corresponds to higher maximal cellular downloading probability.
- The shape (i.e., bends upward or downwards, or mathematically, concave or convex) of the patience function $p_{u,g}$ depends on u's centrality β_u: $\beta_u = \frac{1}{2}$ corresponds to the diagonal; $\beta_u > \frac{1}{2}$ (u is more socially important) corresponds to a concave (bends upward) curve; $\beta_u < \frac{1}{2}$ (u is less socially important) corresponds to a convex (bends downward) curve.
- In all cases, the patience function $p_{u,g}$ is monotonic. The direction of change (i.e., increasing or decreasing) depends on whether u is interested in g himself, i.e., $g \in I_u$ or $g \notin I_u$. If $g \in I_u$, $p_{u,g}$ is monotonically increasing; if $g \notin I_u$, $p_{u,g}$ is monotonically decreasing.

The effect of the parts on the patience function $p_{u,g}$ is illustrated in Figure 2a. The effects of the scaling parameter α_i and α_β are shown in Figures 2b and 2c, respectively.

The probabilistic downloading strategy based on the patience function in Equation (5) has a few desirable properties.

Property 1. *If u has higher chances of serving users (possibly including himself) before content expiration, the maximal probability that u will download the content in one round is higher.*

We can validate Property 1 by noticing that the probability $1 - e^{-\alpha_i i_u(g)}$ is a monotonically increasing function depending only on $i_u(g)$ (given the system scaling parameter α_i): We will see in Section IV-D that the intuition behind $i_u(g)$ is exactly to reflect the chances of u being able to serve content in time.

Property 2. *Other things being equal, more socially important users have higher cellular downloading probability.*

Property 2 is evident by comparing each pair of $\beta_u = 0$ and $\beta_u = 1$ curves with the same set of other parameters. Analytically, by Equation (5), it is straightforward to verify that a larger β_u leads to a larger $p_{u,g}(x)$ for the same $x \in [0,1]$.

The intuition behind Property 2 is that a more socially important user has better chances of meeting others, and passing on the downloaded content. Therefore, letting them download with higher probabilities may help offloading the cellular traffic to the proximity link.

Property 3. *If u is* not *interested in a tag g, u's downloading probability will decrease over time; otherwise, u's downloading probability will increase over time.*

Property 3 is evident by noticing that, in Equation (5), $p_{u,}$ is monotonically increasing if $g \in I_u$ and monotonically decreasing if $g \notin I_u$.

The intuition behind Property 3 is as follows.

If u is not interested in a tag g, u is being purely altruistic in downloading content with g. u can start downloading with a high probability in the hope that he can forward the content to others when they meet later. With the chances of meeting others (and hence forwarding the content to others through the proximity link) dwindling over time, the value of celluar downloading decreases. This is reflected by the monotonically decreasing downloading probability in the second case in Equation (5).

Conversely, if u is interested in a tag g, u is helping both himself and others in downloading content with g. u can afford to start downloading with a low probability in the hope that he can receive the content from another user who has the content, and thus, save cellular bandwidth. With the chances of meeting others (and hence receiving the content from others through the proximity link) dwindling over time, u becomes increasingly *impatient* in waiting. This is reflected by the monotonically increasing downloading probability in the first case in Equation (5).

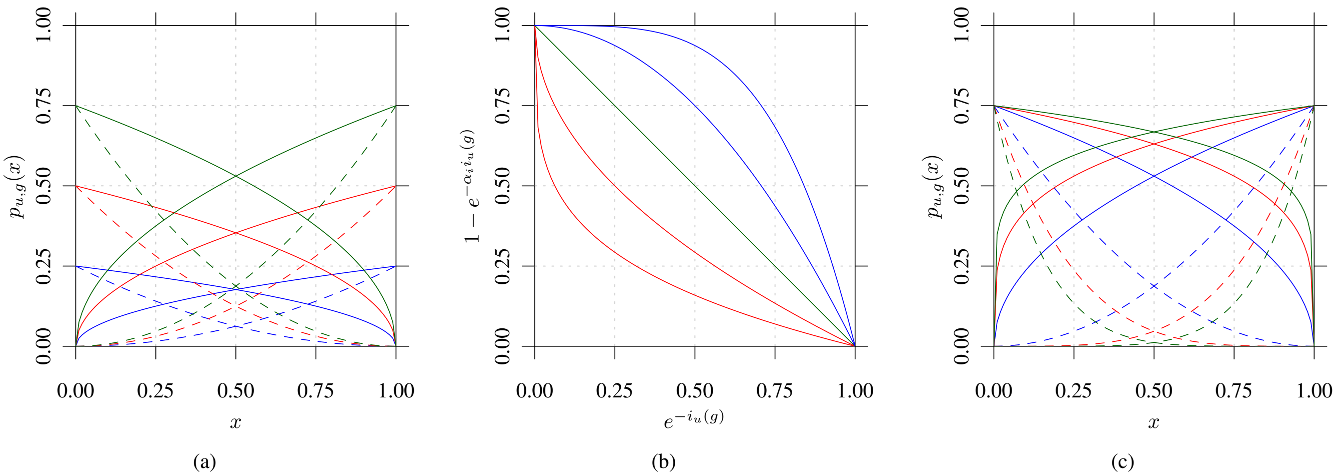


Fig. 2: The patience function $p_{u,g}$ and the scaling parameters α_i (interest $i_u(g)$) and α_β (centrality β_u). (a) Given the scaling parameters α_i and α_β, the patience $p_{u,g}$ function is jointly determined by the aggregated interest $i_u(g)$ and the centrality β_u. For $\alpha_i = 1$ and $\alpha_\beta = 2$, the patience functions corresponding to the 12 ($3 \times 2 \times 2$) combinations $i_u(g) = 0.29, 0.69, 1.39$ (corresponding to $1 - e^{-\alpha_i i_u(g)} = 0.25, 0.50, 0.75$; blue, red, green), $\beta_u = 0, 1$ (dashed, solid), and the cases $g \in I_u, g \notin I_u$ (increasing, decreasing) are plotted for comparison. (b) The effect of the (interest) scaling parameter α_i. The maximum of the patience function ($1 - e^{-\alpha_i i_u(g)}$), which corresponds to the maximal probability that u will download the content through the cellular link in one decision, are plotted against the inverse exponential of the aggregated interest ($e^{-i_u(g)}$) with $\alpha_i = 0.25, 0.5, 1, 2, 4$ (greater than 1: blue; less than 1: red; equal to 1: green) for comparison. (c) The effect of the (centrality) scaling parameter α_β. For (interest) scaling parameter $\alpha_i = 1$ and aggregated interest $i_u(g) = 1.39$ (corresponding to the maximal cellular downloading probability $1 - e^{-\alpha_i i_u(g)} = 0.75$), the patience functions $p_{u,g}$ corresponding to the 12 ($2 \times 2 \times 3$) combinations $\beta_u = 0, 1$ (dashed, solid), $\alpha_\beta = 2, 4, 6$ (blue, red, green), and $g \in I_u, g \notin I_u$ (increasing, decreasing) are plotted for comparison.

B. *Temporal tie strength*

The average interval between consecutive encounter $\hat{s}_u(v)$ quantifies the frequency of encounters between u and v (thus, the opportunities to offload the cellular traffic to the proximity link), based on their past encounters: If they met frequently in the past, they are more likely to do so in the future. The assumption behind this is that human social contacts are regular and thus predictable, which is confirmed by studies on human mobility [10, 11, 16] and is taken by previous social-assisted routing schemes [12, 17].

$\hat{s}_u(v)$ can be computed efficiently by keeping a running sum of past intervals, a count of encounters, and the timestamp of the last encounter. This is amenable for implementation in a large network where the nodes, which are resource-constrained, have to keep track of a large number of neighbors.

The temporal tie strength $s_u(v)$ between u and v is a monotonically decreasing function on $\hat{s}_u(v)$ that maps into $[0, 1]$: The more frequently u and v meet, the stronger their (social) tie is. The reason for making $s_u(v)$ a number between 0 and 1 is to avoid marginalizing u's own interests in the aggregated interest $i_u(g)$ in Equation (3), which will be further discussed later in Section IV-D.

C. *Weighted ego-centric betweenness centrality*

The weighted ego-centric betweenness centrality β_u defined in Equation (2) is inspired and loosely based on the ego-centric betweenness centrality [18]. The difference between the two are the weights on the edges and that, for a pair of u's opportunistic neighbors v and w, we do not divide $[p(v, w)]$ by the number of shortest (weighted) paths between them. The reason is that, given the heuristic nature of the centrality metric, minor relaxation is justified by computation efficiency. The rationale for considering a weighted graph is that, on an intermittently connected graph like the proximity-link-based network, the delay (characterized by the weights on the edges of the graph) matters.

Intuitively, β_u is the ratio (thus, $\beta_u \in [0, 1]$) that, among all pairs of u's opportunistic neighbors, u can pass on content with the shortest delay (the geodesic, or the shortest path). The greater β_u is, the more topologically important, or socially important, that u is on the proximity-link-based network.

D. *Interest aggregation*

u's aggregated interest $i_u(g)$ on a tag g (Equation (3)) gives an estimation on the content demand by u and u's acquaintances.

The rationale for u to weigh an acquaintance v's interests by their tie $s_u(v)$ is as follows. u needs to decide whether downloading a piece of content will help meet v's content demand. This is restricted by their chances of meeting each other, as characterized by their tie $s_u(v)$. Even v is interested in a piece of content, if u has little chance of meeting v, there is little point for u to download the content for v.

Again, the rationale for making the tie $s_u(v)$ a number between 0 and 1 in Equation (1) is to avoid marginalizing u's own interests in the aggregated interest $i_u(g)$ in Equation (3). Downloading a piece of content that u is interested in, himself, will *immediately* satisfy his content demand, while others' content demand will be met by u's celluar downloading only if they meet some time *later*, before the content expires. Therefore, in u's cellular downloading decision, u's own interests are more important than others: Making $s_u(v)$ a number between 0 and 1 does exactly that in Equation (3).

By Equation (3), if u is interested in a tag g ($g \in I_u$),

$i_u(g) \geq 1$: The only possibility that $i_u(g) < 1$ is that $g \notin I_u$. The greater $i_u(g)$ is, the more likely that u downloading a piece of content with tag g will help satisfy users' content demand through the proximity link.

V. SIMULATIONS

We compare the performance of the proposed patience-based offloading strategy with a recent work by Han et al. on cellular offloading, which is based on the target-set formulation [4]. The comparison is based on simulations driven by two publicly available contact traces: a real-world collected trace, Haggle INFOCOM 2006 [19], and a synthesized trace, NUS contact [20].

A. Methodology

1) Dataset: The Haggle INFOCOM 2006 contact trace (Haggle, henceforth) contained Bluetooth sightings of 78 attendees and 20 stationary nodes in the conference venue during the 3 days of the 2006 INFOCOM conference. It is widely cited due to its closed-world nature: The attendees met each other often in the conference venue, which produced a trace with repetitive contact patterns in a short time and a confined space. The time-resolution of this dataset is one second.

The NUS contact trace was synthesized from the class schedules and rosters for the Spring 2006 semester in National University of Singapore (NUS). Students attending the same session of a class were considered to have contacts with each other. In our simulation, we chose a group of 1,000 students who shared a class schedule with at least one other student in the group. The time-resolution of this dataset is one hour.

2) Procedure: Han et al. [4] proposed a deterministic, centralized, and heuristic algorithm to choose a set of nodes to serve as the offloading *target set*, i.e., nodes that download the content at the beginning and serve as initial seeds for subsequent proximate propagation). Although the target-set formulation of the cellular offloading problem is elegant, to select the target set, the algorithm (the emphtarget-set strategy henceforth) is centralized and requires the SP to collect individual nodes' contact information (which intrudes users' privacy) through either cellular links (which is costly under the current mobile computing business model) or other non-cellular links (e.g., WiFi, which is either inconvenient or untimely). Moreover, it is unclear what is the best size for the target set. Follow the method used by Han et al. [4], we resolved this by statistically summarizing the simulation results on multiple target sets with different sizes. Since Han et al. did not consider users' interest in their model [4], to fairly compare their target-set strategy with our patience-based one, we set the upper limit on the target set's size to the number of interested users, to eliminate the cases that (unfairly) favors patience-based strategy due to the absence of a parameter in the target-set model; this allows us to assess the performance of the patience-based offloading strategy more objectively.

In contrast, the patience-based strategy is localized and adaptive. The parameters in Equation (5) can be used to tune the balance between maximizing offloading efficiency and minimizing content delivery delay in the patience-based strategy. To study this flexibility, we used three sets of parameters to instantiate the patience-based strategy. The resulting instances differ in their maximal downloading probability $(1-exp(-\alpha_i i_u(g))$ in Equation (5); explained in Section IV-A) or, more intuitively, the eagerness in downloading the content through the cellular link early. The three instances are identified as *eager*, *moderate*, and *lazy* and their parameters (from Equations (5) and (2)) are shown Table I.

		eager	moderate	lazy
Haggle	α_i	0.5	0.1	0.05
	α_β	2		
	α_s	0.01		
NUS	α_i	0.05	0.03	0.01
	α_β	2		
	α_s	0.01		

TABLE I: Parameters (from Equations (5) and (2)) for the three instances (eager, moderate, lazy) of the patience strategy used in the simulation.

Since the focus of our study was on reducing cellular traffic, we adopted a simple strategy in the opportunistic forwarding between proximate users: Once a node u obtained a piece of content (by either downloading through the cellular link or receiving from other nodes through the proximity link), the content would be forwarded through the proximity link to all of u's neighbors when they met u. This is known in literature as epidemic forwarding or flooding.

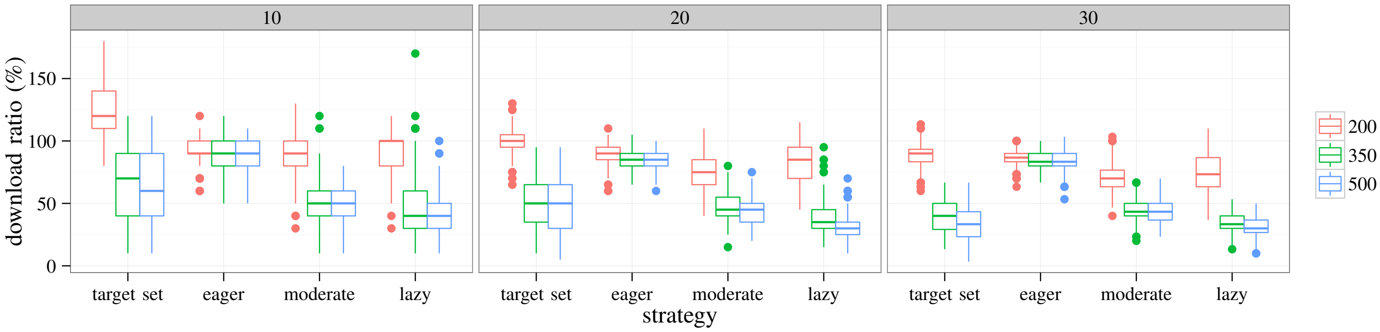
We simulated the cellular downloading decision processes under these strategies with various numbers of interested users. For each given number of interested users, we generated over 100 interest distributions among the users, and for each interest distribution, the stochastic decision process was repeated 50 times to reduce statistical bias.

3) Metrics: Performance of the strategies are measured by two metrics, *downloading ratio* and *content delivery delay*.

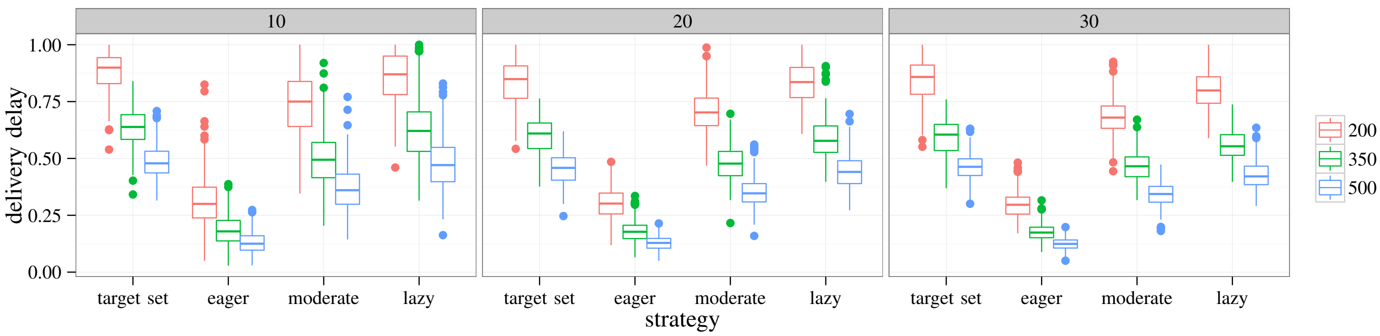
Download ratio. An offloading strategy's efficiency can be measured the number of cellular downloads by the end of the decision process (which is determined by the content's freshness). Quantitatively, if there are n_i users who are interested in the content and, by the end of the offloading process, the content is downloaded through the cellular link d times, the downloading ratio of the offloading strategy is $\frac{d}{n_i} \times 100\%$. An offloading strategy that can satisfy users' content demand with fewer cellular downloads is more efficient.

Content delivery delay. While delay is inevitable for an offloading strategy that does not have every interested user download a piece of content as soon as it becomes available, it is desirable that the delay is minimized. Thus, another aspect of an offloading strategy's efficiency is the content delivery delay that it introduces. Quantitatively, for a piece of content u that is released at the moment 0 and must be delivered by the moment 1^2, let the time of delivery to an interested user $i \in I_u$ be $t_d(i)$, the (average) content delivery delay is $\sum_{i \in I_u} t_d(i)/|I_u|$, which is a value between 0 and 1.

[2] We can normalize the delay by content's delivery deadline to make the delivery delay to 1. Since all interested users who have not received the content by the delivery deadline download the content directly through cellular link, normalized content delivery delay is never greater than 1.

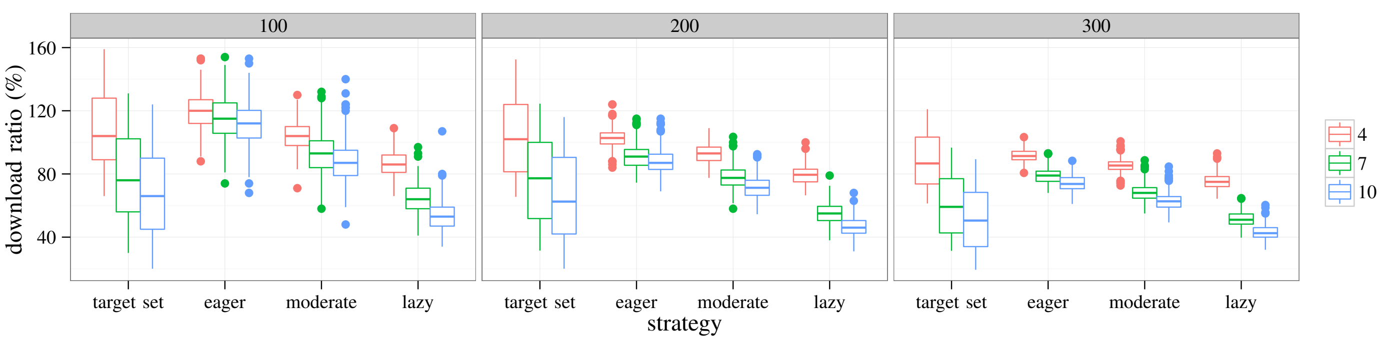


(a)

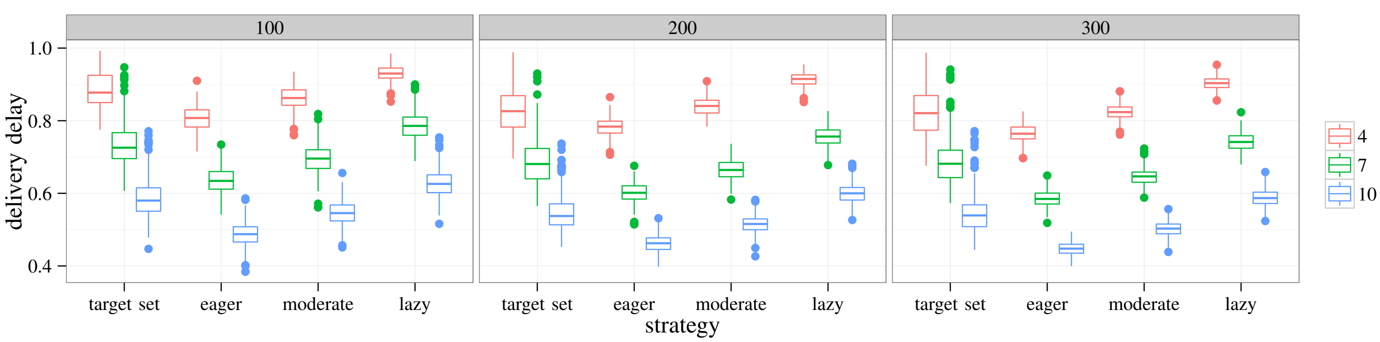


(b)

Fig. 3: Haggle: download ratio and (normalized) delivery delay. Results with different numbers of interested users (10, 20, and 30 interested users out of the 98 nodes) and content delivery deadline (200, 350, and 500 seconds) are compared. For the patience strategies, a downloading decision is made every 50 seconds.



(a)



(b)

Fig. 4: NUS: download ratio and (normalized) delivery delay. Results with different numbers of interested users (100, 200, and 300 interested users out of the 1,000 nodes) and content delivery deadline (4, 7, and 10 hours) are compared. For the patience strategies, a downloading decision is made every 1 hour.

While it would be nice to have an offloading strategy that has low downloading ratio and content delivery delay, these objectives are not always compatible with each other: A trade-off between cellular bandwidth usage and content delivery delay often needs to be made. This is discussed in more detail, in the context of simulation results, in Section V-B.

B. Results

The simulation results of the Haggle dataset are shown in the form of boxplots[3] [21] in Figures 3a and 3b, respectively. Results with different numbers of interested users (10, 20, and 30 interested users out of the 98 nodes) and content delivery deadline (200, 350, and 500 seconds) are compared. As noted in Section V-A2, the target-set strategy was enhanced to eliminate the cases in which the size of the target set is greater than the number of interested users; this allows us to access the performance of the strategies more objectively.

An ideal offloading strategy has a small download ratio and short delivery delay (Section V-A3). In reality, these two goals are usually not compatible. This is evident in the three variants of the patience-based strategy: While the eager variant has the shortest delivery delay (Figure 3b at the expense of largest download ratio (Figure 3a), the lazy variant has the opposite performance trade-off and the moderate variant comes in between. Patience-based strategy, through its parameters (e.g., Table I), provides a control over the trade-off between a small download ratio and short delivery delay.

One type of content that benefits the most from the situation awareness in our patience-based offloading strategy is the content that needs to be delivered quickly. One example is the content that expires after 200 seconds in Figures 3a and 3b: All variants of the patience-based strategy deliver the content with a significantly lower celluar download ratio and delivery delay than that of the target-set strategy. In this case, an interested user who chooses to wait for content is very likely to either 1) receive the content quickly from other users through the proximate channel or 2) do not receive the content till the content delivery deadline. For the latter case, the patience-based offloading strategy allow those users to realize that they are unlikely to receive the content from others (i.e., become *impatient*) and, hence, download the content directly. In contrast, the same group of users will wait the end of content delivery deadline to download the content under the target-set strategy. This corresponds to the shorter delivery delay of the patience-based strategy in Figure 3b.

For a similar reason, the benefit of situation awareness, which is the essence of the patience function (Equation (5)), is more pronounced for the type of content with fewer interested users: The sparsity of the interested users will make those interested users to have a higher probability than their (probably uninterested) neighbors to download the content (through the interest aggregation in Equation (3)); the interested users who have not received the content will become impatient in waiting and hence download the content sooner than they otherwise would under the target-set strategy.

For the popular content that stays fresh longer, such as the one that is interested by 30 users and has content freshness of 500 seconds, the target-set strategy may have a smaller (i.e., better) download ratio than some variants of the patience-based strategy, as shown for the case of eager variant with 30 interested users in Figure 3a. An examination of corresponding cases in Figure 3b suggests that this advantage of the target-set strategy is attained at the expense of a significantly longer delivery delay (3 times as long as that of the patience-based strategy). It also suggests that, in these situations (i.e., many interested users and long content freshness), having only a few users to download the content initially and allow the content to propagate through the proximate channel could greatly reduce cellular downloading cost. The patience-based strategy could be optimized for these situations by tuning the parameters to have a small maximal downloading probability (Equation (5)), as demonstrated by the lazy variant of patience-based strategy in Figure 3a.

Although it is not evident from Figures 3a and 3b, we note that, unlike the target-set strategy that requires collection of users' contact traces for offline training (to find the target set), the patience-based strategy only requires exchange of information between opportunistically encountered users while achieving comparable, or even better, performance: The patience-based strategy is localized and online. The benefit of the patience-based offloading is that it is more less intrusive to users, has lower maintenance overhead for the service provider, is more scalable, and adapts easily to the preference and connection changes among the users.

Comparable results on the NUS dataset are shown in Figures 4a and 4b. Despite the increase of scale (from around 100 nodes in Haggle to $1,000$ nodes in NUS) and trace regularity (NUS is synthesized from class schedules and roasters, as described in Section V-A2), the observation and discussion on performance trade-off and the benefits and limitation of the patience-based strategy drawn from Haggle still hold for NUS.

VI. Related works

Mobile data offloading, or mobile cellular traffic loading, is about the trade-off between the persistent but expensive cellular links and the intermittent but cheap (often free) local links. Balasubramanian et al. [22] and Lee et al. [23] conducted empirical studies, and confirmed the feasibility of offloading cellular traffic through intermittent Wi-Fi links in urban vehicular and pedestrian settings, respectively. Han et al. [3] proposed using Bluetooth to offload cellular traffic. The follow-ups [2, 4] formulated mobile data offloading as a target-set selection problem [24], and proved the approximation bound of a greedy approximation algorithm [25]. Ioannidis et al. [26] proved the convexity of the "timely content distribution over mobile social network" problem and studied how the average age of content changes when the number of users increases. Our work complements their contributions by studying the

[3]Boxplots show the second quartile (i.e., the median) along with the first and third quartiles (i.e., 25% and 75%) in the middle box. The whiskers extend to the extrema within 1.5 times of the inter-quartile range (i.e., the distance between the first and third quartiles). Data beyond the end of the whiskers are outliers and plotted as points [21].

distribution of topical content and modeling users' content preference and time-varying patience for content.

The concept of centrality, which originated in sociology to measure relative importance of social actors, has been applied in studying computer networks. Borgatti [27] surveyed common definitions of the centrality concept (degree, closeness, betweenness, and eigenvector [28]). Hui et al. [17], among others, used centrality as a hint for routing in delay-tolerant networks. Kim and Anderson [29] adapted centralities to temporal-evolving graphs. The significant overhead of gathering information to compute traditional socio-centric centralities prompts researchers to investigate alternative ego-centric centralities, especially ego-centric betweenness centrality [18]. A finding from these investigations is that although the socio-centric and ego-centric versions of the betweenness centrality do not usually match in raw values, they often agree in relative ranking [30]. Brandes, in seeking a faster algorithm to compute the (socio-centric) betweenness centrality, implicitly extends betweenness centrality to weighted graph [31]. Based on the regularity of human mobility pattern [10, 11, 16], we, adapting the ideas of Nanda and Kotz [30] and Brandes [31], define a weighted ego-centric betweenness centrality to help users locally decide their relative temporal topological importance.

Users' content preference was previously considered in the context of content-centric routing [32] and publisher/subscriber architecture [12]. Given the preference variance for the large number of cellular subscribers, it is also relevant for mobile data offloading. Routing through proximity links is not a focus of this work, and we assume flooding. We include content preference in our model, discuss its interplay with social importance and bounded delay tolerance, and provide a method to consolidate them in an adaptive probabilistic cellular offloading strategy.

VII. Conclusion

In offloading topical cellular content, the virtue of patience is to allow the more capable to have better chances of serving the common good. The patience function (Equation (5)) shows one approach to locally synthesizing topological importance and content demand for better offloading efficiency. The simulation results suggest that properly involving topologically important, but disinterested, users in downloading and forwarding content helps in reducing cellular traffic.

These are just the beginnings; plenty of work is left to be done. Enforcement and incentive are two important issues to be further studied once the offloading framework is established. Other practical issues, like packetization, buffer management, and node churning, are omitted in the current work for simplicity, but are unavoidable in real-world implementations. We also plan to implement and deploy the proposed patience-based mobile data offloading strategy in a test-bed environment, which, we expect, will yield further insights.

Acknowledgment

This research was supported in part by NGCRC and NSF grants ECCS 1231461, ECCS 1128209, CNS 1138963, CNS 1065444, CNS-1262984, CCF 1028167, and DUE-1303325.

References

[1] J. Donovan. (2012) Wireless Data Volume on AT&T's Network Continues to Double Annually. [Online]. Available: http://goo.gl/JtNKT

[2] B. Han, P. Hui, V. Kumar, M. Marathe, G. Pei, and A. Srinivasan, "Cellular traffic offloading through opportunistic communications: a case study," in *Proc. ACM CHANTS*, 2010.

[3] B. Han, P. Hui, and A. Srinivasan, "Mobile data offloading in metropolitan area networks," *ACM SIGMOBILE Mob. Comput. Commun. Rev. (MC2R)*, vol. 14, no. 4, pp. 28–30, 2011.

[4] B. Han, P. Hui, V. Kumar, M. Marathe, J. Shao, and A. Srinivasan, "Mobile data offloading through opportunistic communications and social participation," *IEEE Trans. Mob. Comput. (TMC)*, vol. 99, no. 5, pp. 821–834, 2012.

[5] J. Bookwater. (2012) Google Nexus 7 review. [Online]. Available: http://goo.gl/yigHn

[6] R. Ritchie. (2012) iPhone 5 rumored to be getting low power, Wi-Fi Direct enabled chipset... and AirDrop? [Online]. Available: http://goo.gl/pJhTI

[7] NFC Forum. About NFC. [Online]. Available: http://goo.gl/zSJqb

[8] Wi-Fi Alliance. Wi-Fi Direct. [Online]. Available: http://goo.gl/fZuyE

[9] J. Wright, "Access pricing under competition: An application to cellular networks," *Jrnl. Industrial Economics*, vol. 50, no. 3, pp. 289–315, 2002.

[10] M. Gonzalez, C. Hidalgo, and A. Barabási, "Understanding individual human mobility patterns," *Nat.*, vol. 453, no. 7196, pp. 779–782, 2008.

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

[12] F. Li and J. Wu, "MOPS: Providing content-based service in disruption-tolerant networks," in *Proc. IEEE ICDCS*, 2009.

[13] X. Zhuo, W. Gao, G. Cao, and Y. Dai, "Win-coupon: An incentive framework for 3G traffic offloading," in *Proc. IEEE ICNP*, 2011.

[14] W. Peng, F. Li, X. Zou, and J. Wu, "A privacy-preserving social-aware incentive system for word-of-mouth advertisement dissemination on smart mobile devices," in *Proc. IEEE SECON*, 2012.

[15] D. Johnson, "A note on Dijkstra's shortest path algorithm," *Jrnl. ACM (JACM)*, vol. 20, no. 3, pp. 385–388, 1973.

[16] I. Rhee, M. Shin, S. Hong, K. Lee, S. Kim, and S. Chong, "On the Levy-walk nature of human mobility," *IEEE/ACM Trans. Netw. (ToN)*, vol. 19, no. 3, pp. 630–643, 2011.

[17] P. Hui, J. Crowcroft, and E. Yoneki, "Bubble rap: social-based forwarding in delay tolerant networks," in *Proc. ACM MobiHoc*, 2008.

[18] M. Everett and S. Borgatti, "Ego network betweenness," *Soc. Netwrk.*, vol. 27, no. 1, pp. 31–38, 2005.

[19] J. Scott, R. Gass, J. Crowcroft, P. Hui, C. Diot, and A. Chaintreau, "CRAWDAD trace cambridge/haggle/imote/infocom2006 (v. 2009-05-29)," Downloaded from http://goo.gl/u58AG, May 2009.

[20] V. Srinivasan, M. Motani, and W. T. Ooi, "CRAWDAD data set nus/contact (v. 2006-08-01)," Downloaded from http://goo.gl/KRtDE, Aug. 2006.

[21] R. McGill, J. Tukey, and W. Larsen, "Variations of box plots," *The American Statistician*, vol. 32, no. 1, pp. 12–16, 1978.

[22] A. Balasubramanian, R. Mahajan, and A. Venkataramani, "Augmenting mobile 3G using WiFi," in *Proc. ACM MobiSys*, 2010.

[23] K. Lee, I. Rhee, J. Lee, S. Chong, and Y. Yi, "Mobile data offloading: how much can WiFi deliver?" in *Proc. ACM CoNEXT*, 2010.

[24] M. Richardson and P. Domingos, "Mining knowledge-sharing sites for viral marketing," in *Proc. ACM SIGKDD*, 2002.

[25] D. Kempe, J. Kleinberg, and É. Tardos, "Maximizing the spread of influence through a social network," in *Proc. ACM SIGKDD*, 2003.

[26] S. Ioannidis, A. Chaintreau, and L. Massoulié, "Optimal and scalable distribution of content updates over a mobile social network," in *Proc. IEEE INFOCOM*, 2009.

[27] S. Borgatti, "Centrality and network flow," *Soc. Netwrk.*, vol. 27, no. 1, pp. 55–71, 2005.

[28] P. Bonacich, "Power and centrality: A family of measures," *Amrn. Jrnl. Soc.*, pp. 1170–1182, 1987.

[29] H. Kim and R. Anderson, "Temporal node centrality in complex networks," *Phys. Rev. E*, vol. 85, no. 2, p. 026107, 2012.

[30] S. Nanda and D. Kotz, "Localized bridging centrality for distributed network analysis," in *Proc. IEEE ICCCN*, 2008.

[31] U. Brandes, "A faster algorithm for betweenness centrality," *Jrnl. Math. Soc.*, vol. 25, no. 2, pp. 163–177, 2001.

[32] A. Carzaniga, M. Rutherford, and A. Wolf, "A routing scheme for content-based networking," in *Proc. IEEE INFOCOM*, 2004.