Who is the Barbecue King of Texas?:* A Geo-Spatial Approach to Finding Local Experts on Twitter

Zhiyuan Cheng, James Caverlee, Himanshu Barthwal, and Vandana Bachani
Department of Computer Science and Engineering
Texas A&M University
College Station, TX, USA
{zcheng, caverlee, barthwal, vbachani}@cse.tamu.edu

ABSTRACT

This paper addresses the problem of identifying local experts in social media systems like Twitter. Local experts – in contrast to general topic experts – have specialized knowledge focused around a particular location, and are important for many applications including answering local information needs and interacting with community experts. And yet identifying these experts is difficult. Hence in this paper, we propose a geo-spatial-driven approach for identifying local experts that leverages the fine-grained GPS coordinates of millions of Twitter users. We propose a local expertise framework that integrates both users' topical expertise and their local authority. Concretely, we estimate a user's local authority via a novel spatial proximity expertise approach that leverages over 15 million geo-tagged Twitter lists. We estimate a user's topical expertise based on expertise propagation over 600 million geo-tagged social connections on Twitter. We evaluate the proposed approach across 56 queries coupled with over 11,000 individual judgments from Amazon Mechanical Turk. We find significant improvement over both general (non-local) expert approaches and comparable local expert finding approaches.

Categories and Subject Descriptors

H.2.8 [**Database Applications**]: Data Mining

Keywords

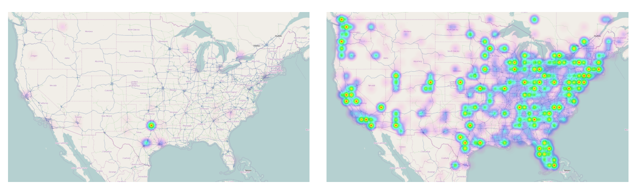
Twitter; expert finding; local expert; social tagging; crowdsourcing

1. INTRODUCTION

We tackle the problem of finding *local experts* in social media systems like Twitter. Local experts bring specialized knowledge about a particular location and can provide insights that are typically unavailable to more general topic experts. For example, a "foodie" local expert is someone who is knowledgeable about the local food scene, and may be able to answer local information needs like: what's the best barbecue in town? Which restaurants locally source their vegetables? Which pubs are good for hearing new bands? Similarly, a local "techie" expert could be a conduit to connecting with local entrepreneurs, identifying tech-oriented neighborhood hangouts, and recommending local talent (e.g., do you know any good, available web developers?). Indeed, a recent Yahoo! Research survey found that 43% of participants would like to directly contact local experts for advice and recommendations (in the context of online review systems like Yelp), while 39% would not mind being contacted by others [1].

And yet finding local experts is challenging. Traditional expert finding has focused on either small-scale, difficult-to-scale curation of experts (e.g., a magazine's list of the "Top 100 Lawyers in Houston") or on automated methods that can mine large-scale information sharing platforms. Indeed, many efforts have focused on finding experts in online forums [29], question-answering sites [18], enterprise corpora [3, 5], and online social networks [8, 11, 22, 25, 30]. These approaches, however, have typically focused on finding general topic experts, rather than *local experts*.

In this paper, we investigate new approaches for mining local expertise from social media systems. Our approach is motivated by the widespread adoption of GPS-enabled tagging of social media content via smartphones and social media services (e.g., Facebook, Twitter, Foursquare). These services provide a *geo-social* overlay of the physical environment of the planet with billions of check-ins, images, Tweets, and other location-sensitive markers. This massive scale geo-social resource provides unprecedented opportunities to study the connection between people's expertise and locations and for building localized expert finding systems.



(a) @BBQsnob (b) @JimmyFallon

Figure 1: Heatmap of the location of Twitter users who have listed @BBQsnob or @JimmyFallon

Concretely, we propose a local expertise framework – LocalRank – that integrates both a person's topical expertise and their local authority. The framework views a local expert as *someone who is well recognized by the local community*, where we estimate this local recognition via a novel spatial proximity expertise approach that leverages over 15 million geo-tagged Twitter lists. Figure 1(a) shows a heatmap of the locations of Twitter users who have labeled Daniel Vaughn (@BBQsnob) on Twitter. Vaughn – the newly-named Barbecue Editor of Texas Monthly – is one of the foremost barbecue experts in Texas. We can see that his expertise is recognized regionally in Texas, and more specifically by local barbecue centers in Austin and Dallas. In contrast, late-night host Jimmy Fallon's heatmap suggests he is recognized nationally, but without a strong local community. Intuitively, Daniel Vaughn is recognized as a *local expert* in Austin in the area of Barbecue; Jimmy Fallon is certainly an expert (of comedy and entertainment), but his expertise is diffused nationally.

*Answer: Daniel Vaughn (@BBQsnob)

SIGIR'14, July 6–11, 2014, Gold Coast, Queensland, Australia.

http://dx.doi.org/10.1145/2600428.2609580.

Toward identifying local experts, this paper makes the following contributions.

• First, we propose the problem of *local expert finding* in social media systems like Twitter and propose a novel expertise framework – LocalRank. The framework decomposes local expertise into two key components: a candidate's topical authority (e.g., how well is the candidate recognized in the area of Barbecue or web development?) and his local authority (e.g., how well do people in Austin – the area of interest – recognize this candidate?).

• Second, to estimate *local authority*, we mine the fine-grained geo-tagged linkages among millions of Twitter users. Concretely, we extract Twitter list relationships where both the list creator and the user being labeled have revealed a precise location. The first local authority method considers the distance between an expert candidate's location and the location of interest, capturing the intuition that closer candidates are more locally authoritative. However, in many cases, an expert in one location may actually live far away – e.g., Daniel Vaughn is an expert in Austin Barbecue although he lives 200 miles away in Dallas. To capture these cases, we propose and evaluate a local authority method that considers the distance of the candidate expert's "core audience" from the location of interest (that is, to reward candidates who have many labelers near the location of interest, even if the candidate lives far away). So, if many people in Austin consider Daniel Vaughn an expert, then his Austin local authority should reflect that.

• Third, to estimate *topical authority*, we adapt a well-known language modeling approach to expertise identification, but augment it to incorporate the distance-weighted social ties of 24 million geo-tagged Twitter users. In this way, topical expertise can be propagated through the social network to identify local experts that are well connected to, and recognized by the local community in the topic.

• Finally, we evaluate the LocalRank framework across 56 local expertise queries coupled with 11,000 individual judgments from Amazon Mechanical Turk. We see a significant improvement in performance (35% improvement in $Precision$@10 and around 18% in $NDCG$@10) over the best performing alternative approach. We observe that the local authority approaches that consider the locations of a candidate's "core audience" perform much better than an alternative that only considers the distance between the candidate's location to the query location. In addition, we see that the expertise propagation through the social network can improve the baseline local expert finding approach.

These results demonstrate the viability of mining fine-grained geo-social signals for expertise finding, and highlight the potential of future geo-social systems that facilitate information flow between local experts and the local community.

2. RELATED WORK

The emergence of online geo-social systems provides unprecedented opportunities to bridge the gap between people's online and offline presence. However there are key challenges associated with these opportunities including location sparsity [2, 7] and location privacy [9, 28]. Given the geo-social footprints from these services, researchers have analyzed the spatio-temporal properties of these footprints [23], studied the semantics associated with these footprints [26], and investigated new location recommendation systems [27, 31].

Expert finding is an important task that has seen considerable research. Lappas et al. [16] provided a comprehensive survey about expert finding in social networks, and grouped the related work into two categories: (i) using text content posted by expert candidates; and (ii) using the expert candidates' online social connections. For example, Balog et al. [3] proposed two generative probabilistic models – a user model generated using documents associated to an expert, and a topic model generated using documents associated to the topic – to detect topic experts. Based on their evaluation over the TREC Enterprise corpora, the authors observed that the topic

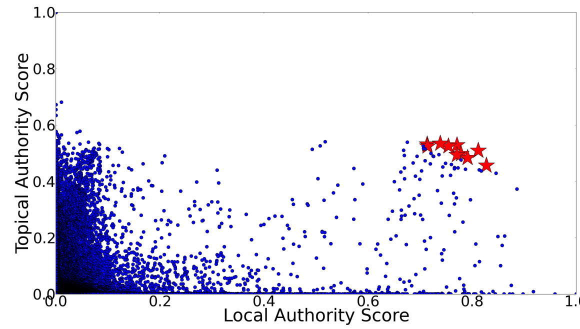


Figure 2: Our goal is to identify Local Experts (the red stars in the top-right section)

model outperforms the user model and other unsupervised techniques. On the other hand, Zhang et al. [29] applied link analysis approaches like PageRank and HITS to identify top experts in a Java forum, observing that both link analysis and network structure are helpful in finding users with extensive expertise.

Along the direction of expert finding in online social networks, Weng et al. [25], proposed a link-analysis based approach to identify top experts in a topic. They considered both topical similarity between users and social connections. The authors observed their approach outperforms Twitter's system, PageRank, and topic-sensitive Pagerank. Similarly, Pal and Counts [22] introduced a probabilistic clustering framework to identify top authorities in a topic using both nodal and topical features. The Cognos system built by Ghosh et al. [11] leveraged Twitter lists to identify the candidate's expertise, and the authors reported that their system works as well as Twitter's official system (i.e., WTF: Who To Follow) to identify top users for a particular topic. Other works include expert finding in online forums [29], question-answering sites [18], enterprise corpora [5, 3], and online social network services [8, 11, 22, 25, 30].

In the context of local experts, Antin et al. [1] recently presented a survey designed to examine people's attitudes about local knowledge and personal investment in local neighborhoods. They observed that over 52% of the participants claimed having both local knowledge and personal investment in their local area. And in an encouraging direction, they found that many participants would like to contact local experts for advice (43%) and many would not mind being contacted by others (39%). To understand people's local expertise, some recent effort [17] proposed to apply points of interests as a possible categorization of expertise.

3. LOCALRANK: PROBLEM STATEMENT AND SOLUTION APPROACH

In this paper, we are interested in finding local experts with particular expertise in a specific location. We assume there is a pool of expert candidates $V = \{v_1, v_2, ..., v_n\}$, that each candidate v_i has an associated location $l(v_i)$ and a set of areas of expertise described by a feature vector $\vec{v_i}$. Each element in the vector is associated with a expertise topic word t_w (e.g., "technology"), and the element value indicates to what extent the candidate is an expert in the corresponding topic. As presented in our previous work [6], we define the **Local Expert Finding** problem as:

DEFINITION 1. ***(Local Expert Finding)*** *Given a query q that includes a query topic $t(q)$, and a query location $l(q)$, find the set of k candidates with the highest local expertise in query topic $t(q)$ and location $l(q)$.*

A location $l(q)$ can correspond to different spatial granularities, depending on the goal of expert finding – a region (e.g., Texas), a city (e.g., Austin), a neighborhood (e.g., downtown), or a latitude-longitude coordinate.

3.1 Topical vs. Local Authority

Identify a local expert requires that we can accurately estimate not only the candidate's expertise on a topic of interest (e.g., how much does this candidate know about barbecue), but also that we can identify the candidate's local authority (e.g., how well does the local community recognize this candidate's expertise). Hence, we propose to decompose the local expertise for a candidate v_i into two related dimensions:

- **Topical Authority**: which captures the candidate's expertise on the topic area $t(q)$.
- **Local Authority**: which captures the candidate's local authority in query location $l(q)$.

To illustrate, Figure 2 shows example candidates in this two-dimensional space for a particular topic (say, Barbecue) and a particular location (say, Austin):

- *Nobodies [bottom-left]*: For a particular area of interest, these candidates have both low topical and local authority.
- *Locals [bottom-right]*: These candidates have high local authority, but low topical authority. E.g., an artist living in Austin.
- *Experts [top-left]*: Candidates with high topical authority, but low local authority. These candidates are certainly experts on a topic, but are not well recognized locally for this expertise. E.g., an expert in pork barbecue originating in North Carolina, but not beef barbecue in Texas.
- *Local Experts [top-right: red stars]*: both great topical authority and local authority. E.g., Daniel Vaughn, the Barbecue Editor of Texas Monthly.

Note that a candidate is evaluated per-topic and per-location, so a local expert in one place may be considered as just an expert or even a nobody in a different location.

3.2 Local Expertise in Twitter Lists

To identify these local experts (the red stars), we propose to exploit the geo-social information embedded in Twitter lists to find candidates who are *well recognized by the local community*. Twitter lists are a form of crowd-sourced knowledge, whereby aggregating the individual lists constructed by distinct users can reveal the crowd perspective on how a Twitter user is perceived [11]. Concretely, for each expert candidate v_i, we assume that there is a set of people $V_l(v_i)$ that recognize v_i's expertise, and label v_i in their own lists. We refer to the set of people as candidate v_i's *list labelers* or more concisely *labelers*. Candidate v_i is the *labelee*. Unique to this study in comparison with previous efforts that use Twitter lists, *for each labeler v_j (such that $v_j \in V_l(v_i)$), we assume that v_j's location $l(v_j)$ is known*. For example, Figure 3 shows a geo-tagged Twitter list relationship in which the list labeler (@jerry from San Antonio) has placed the labelee @BBQsnob (from Dallas) on his "BBQ" list.

But how do we sample such geo-tagged list relationships? Are there sufficient users to support local expertise finding? And if so, do these lists actually reveal topics of potential expertise interest, or are they focused mainly on other dimensions (e.g., for organizing a user's friends)? In the following, we present our Twitter geo-tagged data collection (summarized in Table 1) and address the potential of geo-tagged lists to support local expertise finding, before turning to the development of our local expert finding approach.

Geo-Locating Users. We sample 54 million Twitter user profiles based on the ID range of 12 (starting from Twitter co-founder Jack Dorsey @Jack) to 100 million, as well as 3 billion geo-tagged tweets we collected earlier [14]. For each user, we seek to assign a *home location*; however, it is widely observed that many Twitter users reveal overly coarse or no location at all in the self-reported location field (see, e.g., [7], [13]). While no approach guarantees a perfect geo-location assignment for each user (due to noise and sparsity in self-reported locations), we adopt a home finding method that relies on a user's geo-tagged tweets akin to a similar approach previously used for check-ins and geo-tagged images [20]. First, we group the user's locations where he posted his tweets into squares of one degree latitude by one degree longitude (covering about 4,000 square miles). Next, we select the square containing the most geo-tagged tweets as the center, and select the eight neighboring squares to form a lattice. We divide the lattice into squares measuring 0.1 by 0.1 square degrees, and repeat the center and neighbor selection procedures. This process repeats until we arrive at squares of size 0.001 by 0.001 square degrees (covering about 0.004 square miles). Finally, we select the center of the square with the most geo-tagged tweets as the "home" of the user. In total, we geo-locate about 24 million out of the 54 million users (about 45.1%) with fine-grained latitude-longitude coordinates (using a minumum of 5 geo-tagged tweets per user).

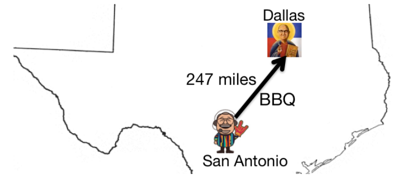


Figure 3: Example: @jerry lists @BBQsnob with label "BBQ".

Table 1: Geo-tagged Twitter data

Data Type	Total # of Records
User Profiles	53,743,459
Geo-Tagged User Profiles (45.1%)	24,252,450
Lists	12,882,292
User List Occurrences	85,988,377
Geo-Tagged List Relationships (17.2%)	14,763,767
Friendship Links	166,870,858
List-peer Relationships	430,186,408

Geo-Labeled List Relationships. Of the 24 million geo-tagged Twitter users, we collect 13 million lists that these users occur on or that these users have created. In total, the 24 million users occur 86 million times in the 13 million lists. Among these 86 million occurrences of a user in a list, almost 15 million of them are geo-tagged, indicating a direct link from a list creator's location to a list member's location. In addition to this network of list relationships, we additionally collect two additional networks around these users: (i) 167 million friendship links connecting these geo-tagged users; and (ii) 430 million links connecting a pair of geo-tagged users that co-occur in the same list.

Expertise Potential of List Names. We parse the list names that are associated with all 14 million geo-tagged list labeling relationships (i.e., links connecting list creator to list member). Table 2 shows the most frequent unigrams. We are encouraged to see that 15 of the 21 most frequent unigrams are related to either people's expertise or interests (the others focus on friendship and celebrity); as has been observed by Kwak et al. [15], Twitter serves as a form of news media as well as a social network, so there is good potential for expertise mining.

Spatial Patterns of Expertise. What do these geo-tagged lists reveal? For four example topics – "tech", "entertain", "travel", and "food" – we plot in Figure 4 the cumulative distribution of frequency of list labeling relationships over distance. That is, how far apart are list labelers from the list labelees? We observe almost 40% of Twitter users who are labelees in a "food"-relevant list are within a hundred miles to the labelers. However, only about 10% to 15% of the labelees in a list of other three topics are within a hundred miles to their labelers. In addition, the average distance between a pair of list labeler and list labelee for "food" is also much

Table 2: Most frequent words in list names of geo-tagged list labeling relationships

news	2.66%	**media**	1.87%	**music**	1.71%
twibe	1.27%	**tech**	1.11%	people	1.06%
celeb	1.04%	**social**	1.04%	**sport**	1.01%
design	0.84%	**market**	0.81%	**politic**	0.80%
follow	0.70%	celebrity	0.69%	**food**	0.61%
art	0.58%	**business**	0.55%	friend	0.52%
entertain	0.50%	**web**	0.48%	**travel**	0.47%

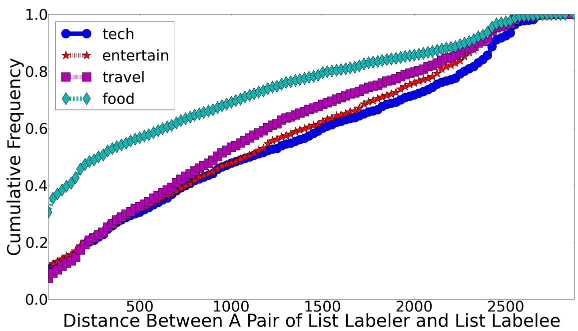


Figure 4: Cumulative frequency of list relationship distances

smaller than the average distance for other topics. These observations suggest that certain topics are inherently more "local" and that identifying local experts in topics that are inherently more local could be easier than identifying local experts in other topics.

3.3 Local Expert Finding with LocalRank

Based on these encouraging observations – (i) that there is a wealth of geo-tagged list data in Twitter; (ii) that these lists tend to focus on areas of potential expertise; and (iii) that distance impacts list labeling (and possibly revealing the localness of particular topics) – we turn in the next two sections to developing methods for identifying local experts.

Recall that we propose to measure a candidate v_i's local expertise by a combination of both the candidate's topical authority and local authority. While there are many ways to integrate these two scores, we propose a simple combination in this first study. We formally define candidate v_i's **LocalRank** (LR) $s(v_i, q)$ in query q as:

$$s(v_i, q) = s_l(l(v_i), l(q)) * s_t(\vec{v_i}, G, t(q))$$

where $s_l(l(v_i), l(q))$ denotes the *Local Authority* of v_i in query location $l(q)$, and $s_t(\vec{v_i}, G, t(q))$ denotes the *Topical Authority* of v_i in query topic $t(q)$ that is estimated using the candidate's expertise vector $\vec{v_i}$, and the social graph G that the candidate is involved in. In the following two sections we investigate how to estimate these values.

4. ESTIMATING LOCAL AUTHORITY

In this section, we present three approaches for estimating a candidate expert's *local authority*. The first local authority method considers the distance between an expert candidate's location and the location of interest, capturing the intuition that closer candidates are more locally authoritative. The latter two leverage the fine-grained geo-tagged linkages among the sampled Twitter users as revealed through list relationships, where both the list creator and the user being labeled have revealed a precise location.

Candidate Proximity. The first (and perhaps most intuitive) approach to estimate candidate v_i's local authority for query q is to use the distance between candidate v_i's location $l(v_i)$ and the query location $l(q)$. For example, if we are looking for experts on Austin Barbecue, then all candidates located in Austin will be considered more authoritative than candidates outside of Austin. We define this *Candidate Proximity* ($s_{l_{CP}}$) as:

$$s_{l_{CP}}(l(v_i), l(q)) = \left(\frac{d_{min}}{d(l(v_i), l(q)) + d_{min}}\right)^{\alpha}$$

where $d(l(v_i), l(q))$ denotes the distance between $l(v_i)$, and $l(q)$ (using the Haversine formula which accounts for the curvature of the earth), and we set $d_{min} = 100$ miles. In this case α indicates how fast the local authority of candidate v_i for query location $l(q)$ diminishes as the candidate moves farther away from the query location. This first local authority approach captures the intuition that closer candidates are more locally authoritative. Figure 5(a) shows a candidate expert in Baltimore (the green pentagon); if we are looking for an expert in New York (the gold star), such a Baltimore candidate's local expertise will be a function of the distance from Baltimore to New York. While simple, this approach cannot capture local expertise of candidates who do indeed live far from a location of interest. As we have mentioned before, Daniel Vaughn is an expert in Austin Barbecue although he lives 200 miles away in Dallas.[1]

To capture these cases where expertise is not dictated solely by distance from a candidate to an area of interest, we next propose two local authority methods that consider the distance of the candidate expert's "core audience" from the location of interest (that is, to reward candidates who have many labelers near the location of interest, even if the candidate lives far away).

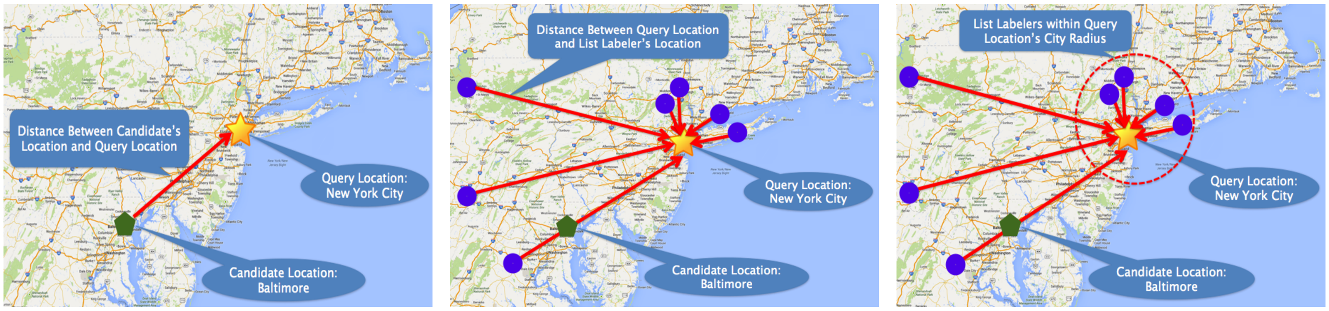
Spread-based Proximity. The first of these geo-tagged list methods is the *Spread-based Proximity* that measures the "spread" of a candidate's core audience's locations compared to the query location:

$$s_{l_{SP}}(L(V_l(v_i)), l(q)) = \frac{\sum_{v_{l_j} \in V_l(v_i)} s_{l_{CP}}(l(v_{l_j}), l(q))}{|V_l(v_i)|}$$

where v_{l_j} denotes one of the core audience $V_l(v_i)$ of candidate v_i. Basically, the "spread" it measures considers how far candidate v_i's core audience are from the query location $l(q)$ on average. If the core audience of a candidate is close to a query location on average, the candidate gets a high score of $s_{l_{SP}}$. For example, in Figure 5(b), the green pentagon and the gold star represent the expert candidate's location and the query location, respectively. However, the spread-based proximity for the candidate in the query location emphasizes the distance of the links (plotted as red arrows) between the candidate's list labelers' locations (plotted as blue dots) and the query location.

Focus-based Proximity. In some cases, the spread-based proximity approach may underestimate a candidate's local authority. For example, for a couple of "foodies" v_a and v_b both in New York City, suppose v_a has a large audience in New York City recognizing his food expertise, and is well appreciated by a lot of people on the west coast, and even abroad; while v_b is much less well recognized by the local community in New York City, but has more people recognizing his expertise in mid-east United States, North Carolina, and Florida. Despite a much better local community recognition in New York, user v_a has a lower value of spread-based core audience query spatial proximity, due to the higher spatial spread of his

[1]In addition, the home location of an expert candidate may not even be accurate: recall that our home locator estimates a location based on a single user's geo-tagged tweets. In contrast, the following two local authority methods consider the aggregated perspectives of many list labelers, so there is a clearer signal of a candidate's location of expertise.



(a) Candidate Proximity (b) Spread-based Proximity (c) Focus-based Proximity

Figure 5: Three methods for estimating local authority

labelers. To overcome this type of expertise underestimation, we propose the *Focus-based Proximity* as:

$$s_{l_{FP}}(L(V_l(v_i)), l(q)) = \frac{|\{v_{l_j} | d(l(v_{l_j}), l(q)) \leq r(l(q))\}|}{|V_l(v_i)|}$$

where $r(l(q))$ represents a radius around a location $l(q)$. This focus-based proximity measures how focused a candidate's audience is in the query location by measuring the percentage of the core audience that resides within the radius of the query location. For example, in Figure 5(c), 4 out of 7 labelers (blue dots) for the candidate (green pentagons) are within the radius (plotted as the red dashed circle) of the query location (gold star), and the focus-based proximity in this case is $\frac{4}{7} \approx 0.57$.

These two local authority methods – the spread-based and focus-based approaches – are designed to capture the expert candidate's spatial influence measured via collective intelligence contributed by the people who labeled the candidate.

5. ESTIMATING TOPICAL AUTHORITY

In this section, we discuss how we estimate the topical authority score of candidate v_i being as a local expert in query q. Specifically, we propose to use both the crowd-sourced geo-tagged labels and the social connections between people to quantify a candidate's topical expertise score given a query.

5.1 Directly Labeled Expertise

We begin with a topical authority approach that leverages the directly labeled expertise of candidate v_i, as revealed through the sampled Twitter lists. Specifically, we adapt the user-centric model that Balog et al. proposed in [3] to estimate the *Topical Authority Score* $s_t(\vec{v_i}, G, t(q))$ of v_i with respect to the query topic $t(q)$ (ignoring for now the social graph G). Balog et al. applied the user-centric model to identify an expert's knowledge based on the documents (emails and web pages) that they are associated with. In our scenario, we apply the user-centric model to identify expert candidates' expertise based on the list labels that the crowd has applied to them.

The model is built on standard language modeling techniques: a user v_i can be represented by a multinomial probability distribution over the vocabulary of topic words (i.e., $p(t_w|\theta_{v_i})$, where θ_{v_i} denotes a user model). In this case, for each user v_i, we infer a user model θ_{v_i} such that the probability of a topic word t to occur in user v_i's list labels can be estimated via $p(t_w|\theta_{v_i})$.

Given user v_i's user model θ_{v_i}, for a query q, user v_i's *Topical Authority Score* $s_t(\vec{v_i}, G, t(q))$ in query q is measured as the probability of query text $t(q)$ to be generated from the users' user model:

$$s_t(\vec{v_i}, G, t(q)) = p(t(q)|\theta_{v_i}) = \prod_{t_w \in t(q)} p(t_w|\theta_{v_i})$$

where t_w denotes a topic word in query text $t(q)$.

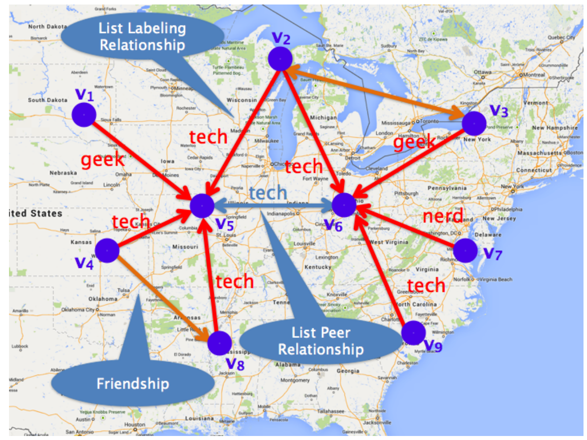


Figure 6: Examples of social and list-based connections

Since we are expecting that most of the users will be labeled by a small number of unique labels, most of the topic words will have zero probabilities for a particular user v_i. Thus we smooth $p(t_w|\theta_{v_i})$ using the probability of the topic word to occur in the whole corpus of labels $p(t_w|\theta_{C_v})$ when estimating $p(t_w|\theta_{v_i})$:

$$p'(t_w|\theta_{v_i}) = (1 - \lambda) * p(t_w|\theta_{v_i}) + \lambda * p(t_w|\theta_{C_v})$$

Here λ represents the extent of smoothing. A large value of λ indicates that the probability $p(t_w|\theta_{v_i})$ is more weighted towards the probability of the topic word t_w to occur in the corpus $p(t_w|\theta_{C_v})$. In the experiments, we fix the value of λ to 0.1.

5.2 Expertise Propagation

In addition to the directly labeled expertise derived from our collection of geo-tagged Twitter lists, we are interested to explore whether the social and list-based connections of Twitter users also provide strong signals of expertise. Specifically, we consider three graphs that include three types of connections: (i) User Friendship; (ii) List-labeling Relationship; and (iii) List-peer Relationship (see the data collection described in Table 1). Recall that each user v_i is characterized as a vector $\vec{v_i}$ of his topical expertise generated from the directly labeled expertise method. Can we enrich the expertise signals from the Twitter lists by propagating expertise along these three graphs? The intuition is that people with particular expertise have a higher likelihood to be connected to other people with the same expertise, and that having multiple connections to people with a particular expertise raises the possibility of an individual also having that expertise.

User Friendship. The first expertise propagation approach is based on user friendship, as represented by a direct link $e(v_i, v_j)$ from user v_i to user v_j. In Figure 6, we show nine expert candidates (plotted as blue dots that are labeled from v_1 to v_9). Here, a *friendship link* (plotted as an orange arrow) connects a candidate to another candidate that he follows, and an example would be the orange arrow on the bottom left from v_4 to v_8. The motivation for propagation along friendship links is that a candidate has a higher likelihood to be an expert in query topic $t(q)$ if he has friend(s) that are also expert(s) in query topic $t(q)$.

Given users' friendship linkages, we can generate the friendship graph $G_f(V, E)$ for a set of users V, and a set of friendship links E that connect users in V. For every edge $e_f(v_i, v_j)$, the weight $w_f(v_i, v_j)$ is simply $\frac{1}{|E_{out}(v_i)|}$, where $E_{out}(v_i)$ represents the set of out links from user v_i.

In addition, from the perspective of the "First Law of Geography" [24] that "everything is related to everything else, but near things are more related than distant things", we hypothesize that a user knows a friend nearby better than a friend farther away. Thus, we generate an alternative $G_f{}'(V, E)$ to reflect the effect of distance between a pair of connected users v_i, and v_j on how well user v_i knows v_j (i.e., how much credit v_j gets from v_i), by introducing the local authority score to the calculation of the weight $w_f{}'(v_i, v_j)$ for edge $e(v_i, v_j)$:

$$w_f{}'(v_i, v_j) = \frac{s_l(l(v_i), l(v_j))}{|E(v_i)|}$$

List-labeling Relationship. The second expertise propagation approach considers the *list-labeling relationship* derived from the sampled Twitter lists. The motivation for the propagation here is: if an expert v_i in a topic $t(q)$ labels another user v_j as an expert in the same topic, user v_j also has a high likelihood to be an expert in the topic.

For example, user v_i lists user v_j as a tech expert in one of his lists on Twitter, generating a direct link $e_l(v_i, v_j)$ from v_i to v_j indicating a relationship connected by expertise recognition. In this way, a graph G_l capturing the expertise recognition can be constructed. Returning to Figure 6, we show this list-labeling relationship (plotted as a red arrow) that links a list labeler to the candidate that he listed, and an example would be the red arrow on the top left from v_1 to v_2 with a list label "geek".

As in the friendship case, we can similarly construct two graphs – one with the weight $w_l(v_i, v_j)$ and the other one with the distance-based weight $w_l{}'(v_i, v_j)$ for the link $e_l(v_i, v_j)$ according to the number of out links from v_i in G_l, and $G_l{}'$ respectively.

List-peer Relationship. Finally, we can propagate expertise along peers that appear on the same list. Returning to Figure 6, this list-peer relationship (plotted as a blue arrow) indicates a connection between two candidates that appear on the same list, and examples in the figure are the blue arrows in the middle between v_5 and v_6 with a list label "tech". This list peer relationship carries an important signal: a person's co-appearance with several top experts on lists further strengthens her topical authority.

Here, we have the link $e_{lp}(v_i, v_j)$ that directly connects user v_i to user v_j in a list on Twitter. We can measure the weight $w_{lp}(v_i, v_j)$ for the link $e_{lp}(v_i, v_j)$ according to the number of out links from v_i in G_{lp}. Using all the list peer relationship, we generate a social graph G_{lp} that captures the signals of expertise propagated from list peers. We can also generate the corresponding distance-weighted list peer graph $G_{lp}{}'$.

Topical Authority Score from Expertise Propagation. Given these three perspectives, we propagate expertise along these graphs through a random walk based on topic-sensitive PageRank (TSPR) [12]. Again, our intuition is that people with particular expertise have higher likelihood to be connected to other people with the same expertise. The random walk approach leverages this intuition by propagating expertise along links in the graph, and by resetting back to the candidates with high directly labeled expertise. Thus, for each particular social graph G described above (that is: G_f / $G_f{}'$, G_l / $G_l{}'$, or G_{lp} / $G_{lp}{}'$), we apply TSPR on the specific social graph to identify the most influential users for a particular query topic $t(q)$. The stabilized TSPR score for each user v_i is considered as user v_i's topical authority score $s_t(\vec{v_i}, G, t(q))$ in query topic $t(q)$. In our experiments, we explore using both the general social graph, and the distance weighted social graph to identify top local topical experts for a given query.

6. EVALUATION

In this section, we evaluate the proposed local expert finding framework. We seek answers to the following questions:

- What impact does the choice of local authority have on the quality of local expert finding in LocalRank? How much do crowdsourced geo-tagged list labels impact local authority (and ultimately the quality local expert finding)?
- Do the three types of expertise propagation over social and list-based connections of Twitter users provide strong signals of topical expertise? And if so, to what degree over directly labeled expertise?
- How well does LocalRank perform compared to alternative local expert finding approaches? Is integrating topical and local authority necessary?
- Finally, how do the approaches perform in finding top local experts for finer topics? Do we see consistent performance in comparison with more general topics?

6.1 Experimental Setup

Assessing local expertise is difficult since there is no explicit ground truth data that specifies a user's local expertise given a query (location + topic). Hence, in this section we describe how we constructed our testbed: we first describe the location + topic queries and then introduce the specific expert finding approaches we tested. We discuss how we gathered ground truth to evaluate these approaches, and how we measured approach effectiveness.

Queries. In total, we evaluate local expert finding using 56 queries (16 general topic queries and 40 finer topic queries). We consider four general query topics coupled with four locations, totaling 16 topic-location queries. Specifically, we look for local experts in the areas of "technology", "entertainment", "food", and "travel" in New York City, Houston, San Francisco, and Chicago. We also consider 10 refined topics under the general umbrella of "food" and "startup", again in the same locations, totaling 40 topic-location queries. These refined topics are "barbecue", "seafood", "pizza", "winery", and "brewery" under the "food" scenario, and "venture capital", "incubator", "founder", "entrepreneur", and "angel investor" under the "startup" scenario. By considering both general-topic and finer-topic local expertise queries, our goal is to investigate differences in local expertise finding at varying granularities of expertise.

Approaches for Finding Local Experts. In addition to the proposed local expert finding approaches presented in this paper, we consider five alternative baselines. The first considers only a candidate's topical authority (ignoring local authority):

- *Directly Labeled Expertise* (*DLE*): Rank candidates by topical authority in the query topic.

The next three consider only a candidate's local authority (ignoring topical authority):

- *Nearest* (*NE*): Rank candidates by distance to the query location.
- *Most Popular in Town by Followers Count* (*MP (follower)*): Rank candidates from the query location by follower count.
- *Most Popular in Town by Listed Count* (*MP (list)*): Rank candidates from the query location by the number of lists the candidate appears on.

The final baseline combines simple versions of topical and local authority:

- *Most Popular in Town by Listed Count on Topic* (*MP (on-topic)*): Rank candidates from the query location by the number of on-topic lists the candidate appears on.

We compare these five baselines with the proposed LocalRank approach presented in this paper. For LocalRank, we investigate the three approaches for estimating local authority – by Candidate Proximity (CP), Spread-based Proximity (SP), and Focus-based Proximity (FP) – and the Directly Labeled Expertise (DLE) and Expertise Propagation (EP) approaches for estimating topical authority. When applying both the *Candidate Proximity*, and *Spread-based Proximity*, we preset the d_{min} to be 100 (miles), and $alpha$ to be 2.0. We calculate the local expertise score using the normalized topical authority score and the normalized local authority score.

Gathering Ground Truth. To gather ground truth, we employ human raters on Amazon Mechanical Turk. Since there are combinatorially too many *approach* + *query topic* + *query location* + *candidate expert* variations, we rely on a sampling method whereby for each experimental setting (an approach + a query topic + a query location), we retrieve the corresponding top-10 local expert candidates with the highest local expertise scores, and have human raters on Mechanical Turk label to what extent an expert candidate has local expertise in the query topic and the query location. For each expert candidate, 5 relevance assessors label the candidate's local expertise using a four-scale local expertise rating:

- Extensive Local Expertise [+2]: The candidate has extensive expertise in the query topic, and is locally well recognized in the query location for his expertise.
- Some Local Expertise [+1]: The candidate has some expertise in the query topic, and also has some influence in the query location
- No Evidence [0]: The candidate has no clear evidence to be considered as having expertise in the query topic, or influence in the query location.
- No Local Expertise [-1]: The candidate has neither any expertise in the query topic, nor influence in the query location.

For each assessment, we provide the assessor with the candidate's user profile, a word cloud generated using the labels that people used to describe the candidate, a heatmap showing the locations of the candidate's labelers, the candidate's most retweeted 5 tweets and 5 most recent tweets. To ensure the quality of these assessments, we follow the conventions suggested by Marshall and Shipman [19]. Each individual HIT (Human Intelligence Task) includes 10 query / expert candidate pairs randomly selected from all the pairs of query and expert candidate. 2 out of the 10 pairs for each HIT are manually labeled by domain experts in order to evaluate the quality of the feedback from assessors. If an assessor picks a significantly different answer comparing to ours for either one of the two particular pairs, the feedback for the HIT will be discarded. For a particular pair of query and expert candidate, we use the best judgment (i.e., the most voted rating) out of the 5 assessors as the final rating for the pair.

We investigate the inter-judge agreement using both *kappa statistic* and *Accuracy*. Since we have more than two annotators (five in our scenario) for each query-candidate pair, we adopt Fleiss' kappa [10], which ranges from 0 (when the agreement is not better than chance) to 1 (when the two annotators agree with each other perfectly). Following Brants [4] and Nowak et al. [21], we define *Accuracy* as:

$$Accuracy(Q_{pairs}) = \frac{\sum_{q_{pair} \in Q_{pairs}} \frac{\text{\# of votes for the majority rating}}{\text{\# of votes for } q_{pair}}}{|Q_{pairs}|}$$

where Q_{pairs} represents the set of query and candidate pairs, in which each pair q_{pair} includes both a query q, and an expert candidate c. An ideal *Accuracy* would be 1.0 that all the assessors pick the same local expertise rating for every particular pair of query and candidate. For example, an *Accuracy* of 0.6 indicates that for a query-candidate pair, 60% of the human raters voted for the majority choice.

Metrics. To evaluate each local expert finding approach, we measure the average *Rating@10*, *Precision@10*, and *NDCG@10* across all queries in our testbed. For the following experiments, we consider all the 0 and -1 ratings as 0s.

Rating@10 measures the average local expertise ratings by the human-raters for the top 10 ranked local experts across all the queries:

$$Rating@10 = \frac{\sum_{q \in Q_{pairs}} (\sum_{i=1}^{10} rating(c_i, q)/10)}{|Q_{pairs}|}$$

where Q_{pairs} represents the set of all query pairs, and $rating(c_i, q)$ denotes the most voted local expertise rating for candidate c_i in query q. $Rating@10$ ranges between 0 to 2, and an ideal approach will have a $Rating@10$ value 2, which all identifies local experts with extensive local expertise in the query topics and locations. Conversely, the worst performing approach will have a $Rating@10$ value 0, indicating that the approach only identifies local experts as those with no local expertise or no evidence.

Precision@10 measures the average percentage of candidates that are relevant to the query topic and query location in the top 10 candidates across all the queries. It is defined as:

$$Precision@10 = \frac{\sum_{q \in Q_{pairs}} \frac{|\{c_i | rating(c_i, q) > 0\}|}{10}}{|Q_{pairs}|}$$

In this paper, we consider expert candidates with both "extensive local expertise", and "some local expertise" as relevant, while we consider both "no local expertise" and "no evidence" as irrelevant. A perfect local expertise estimator has a $Precision@10$ value of 1.0.

NDCG@10 (Normalized Discounted Cumulative Gain@10) measures how well the predicted local expert rank order is compared to the ideal rank order (i.e., candidates are ranked according to their actual local expertise) for the top 10 results across all the query pairs. $NDCG@10$ ranges between 0 and 1, and a higher value indicates an approach that generates better rank orders.

6.2 Agreement of Local Expertise

Overall, we have 11,285 individual judgments made by the human raters. How consistent and reliable are these judgments? We report the kappa (κ) and *Accuracy* values in Table 3. When considering 3 rating categories for each pair (2: "extensive local expertise", 1: "some local expertise", and 0: either "no local expertise" or "no evidence"), the overall *Accuracy* for agreement is 0.716, indicating that for a pair of query and candidate, on average 71.6% of the human raters voted for the majority vote. This demonstrates good user agreement and is significantly higher than accuracy by chance (33.3% for three categories). When considering only 2 rating categories (2 and 1 as relevant, and 0 as irrelevant), the overall *Accuracy* increases to 82.2%, which is also much higher than the accuracy by chance (50% for two categories). For kappa, we see that the overall value is 0.280 in the 3 rating category case. For the binary rating case, the overall kappa value is 0.397. Both kappa statistics are typically considered "fair" inter-judge agreements. Together, these kappa and *Accuracy* values suggest that these human raters have a fairly reasonable agreement. And we observe both much higher *Accuracy* and kappa value for binary rating categories, which indicates that raters find it easier to decide whether a candidate has local expertise or not, rather than determining the extent of a candidate's local expertise.

6.3 Comparing Local Expert Finding Approaches

In this subsection, we seek answers for the questions brought up in the beginning of this section, with four set of experiments: (i)

Table 3: User agreement for overall judgments

	3 Rating Categories		2 Rating Categories	
Overall	Accuracy	κ	Accuracy	κ
	0.716	0.280	0.822	0.397
General Topics	Accuracy	κ	Accuracy	κ
	0.715	0.279	0.818	0.393
Finer Topics	Accuracy	κ	Accuracy	κ
	0.717	0.281	0.825	0.401

Table 4: LocalRank: Evaluating the three local authority approaches

Local Authority	$Rating@10$	$Precision@10$	$NDCG@10$
CP	0.952	0.553	0.685
SP	1.330	0.830	**0.903**
FP	**1.334**	**0.842**	0.896

evaluating the performance of local authority metrics; (ii) studying the impacts of expertise propagation; (iii) comparing the performance of baseline approaches and the LocalRank approaches; and (iv) evaluating the performance of expert finding via finer topics.

6.3.1 LocalRank: Evaluating Local Authority

To begin with, we seek to understand the impact of the local authority approach on the quality of local expert finding in LocalRank. Specifically, we fix the LocalRank topical authority as the Directly Labeled Expertise, while we vary the local authority across the three approaches presented in Section 4: Candidate Proximity (CP), Spread-based Proximity (SP), and Focus-based Proximity (FP). Our goal is to understand to what degree the local authority affects local expert finding, and to assess if (and how much) the crowdsourced geo-tagged list labels impact local authority.

We present in Table 4 the $Rating@10$, $Precision@10$, and $NDCG@10$ for each of the three local authority approaches. We observe that both of the approaches (SP and FP) that utilize the locations from the candidates' core audience significantly improve the performance of local expert finding in comparison with the candidate proximity approach (CP) that only takes the candidate's physical location into consideration. Using candidate proximity (CP), the LocalRank approach only identifies true local expert 55% of the time on average among the top 10 candidates. Similarly, we see comparatively low values of $Rating@10$ as 0.952, and $NDCG@10$ as 0.685. In contrast, the Spread-based Proximity (SP) and Focus-based Proximity (FP) approaches reach $Precision$ $@10$ of almost 85%, $Rating@10$ over 1.33, and $NDCG@10$ of 0.90. This indicates the core audience for an expert candidate is crucial to estimating a candidate's local authority. And in absolute terms, the rating scores for both approaches range between "some local expertise" (1) and "extensive local expertise" (2), indicating that these approaches can identify candidates who are actually local experts. Interestingly, we see for this evaluation framework that the two approaches perform nearly equally well, although they capture two different perspectives on local authority (recall that SP considers the average distance of labelers, whereas FP considers the fraction of labelers within a radius).

6.3.2 LocalRank: Impact of Expertise Propagation

Given these results for local authority, we next consider the impact of expertise propagation on the topical authority (and ultimately on the quality of local expert finding). As described in Section 3.2, we explore whether the three types of social and list-based connections of Twitter users do indeed provide strong signals of expertise. We consider the (i) friendship graph, (ii) list-labeling relationship graph, and (iii) list-peer relationship graph. For each graph (both with and without distance-weighted edges), we apply the topic-sensitive PageRank algorithm to propagate expertise. For each particular graph as well as a particular type of edge weight, we iterate the damping factor from 0.10 to 0.30 to 0.50 to study how the damping factor affects the task of finding top local experts. A

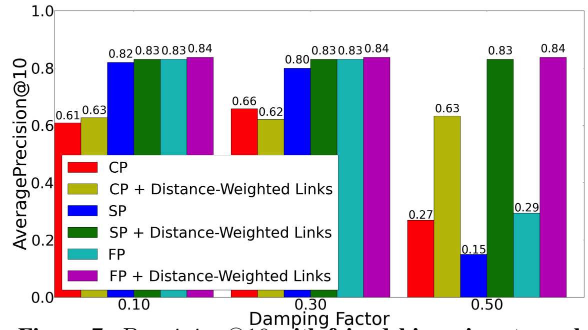


Figure 7: $Precision@10$ with friendship as input graph

smaller damping factor indicates less score propagation and more random walking among more topic-relevant nodes in the graph. We find that the conventional damping factor value (0.85 or 0.90) finds only national celebrities like @JimmyFallon (Jimmy Fallon, host of talk show Late Night with Jimmy Fallon), @TheEllenShow (Ellen Degeneres, host of the Ellen Degeneres Show), and @Jack (Jack Dorsey, Twitter and Square co-founder) no matter what the query topic is. With a smaller value of damping factor, we hope to identify more topical relevant local experts.

We present in Figure 7 the local expertise results for expertise propagation using the Friendship graph as input, coupled with corresponding parameter settings. We vary the choice of local authority (CP, SP, and FP), the use of distance-weighted links or not, as well as the choice of damping factor. This figure focuses on $Precision@10$, while the subsequent Table 6 includes $Rating@10$, $Precision@10$, and $NDCG@10$ for all graph types. First, in terms of the damping factors, we see that across all settings (0.10, 0.30, and 0.50), that the best performing result is comparable. However, we do observe a significant performance drop for damping factor 0.50 using regular edge weight that does not consider distance between the nodes as a factor. Upon investigation into the top local expert candidate under this setting, we observe that many of the top local candidates are national celebrities (e.g., @JimmyFallon, @TheEllenShow, and @Jack), compared to the candidates retrieved using a damping factor of 0.10 or 0.30. We attribute this result to the higher weight on score propagation through general friendship edges. On the other hand, for a damping factor 0.10 or 0.30, most of the scores are propagated through topic-relevant nodes via random walking.

Second, we observe a slight improvement for distance-based edge weight when using a damping factor of 0.10 or 0.30 rather than using the regular edge weight. And we observe a dramatic improvement of performance for distance-weighted edge weight using a damping factor of 0.50 than the alternative version. This indicates that giving local friends more credit (in terms of expertise propagation flowing more strongly to nearby friends than far away ones) does help improve the likelihood to find better top local experts.

Third, in terms of the choice of location authority metric, we observe a similar result to what we observed in the previous section – that the approaches (SP and FP) that utilize the locations from the candidates' core audience significantly improve the performance of local expert finding.

Finally, compared to the simpler approach of not propagating expertise at all, but just using the directly labeled expertise, we see that the results are quite similar (with $Precision@10$ near 0.84). Given this result, we compared the lists of top-10 local experts returned by LocalRank using directly labeled expertise versus LocalRank using each one of the expertise propagation approaches. While the overall precision is similar, the experts that each approach finds are different: we find an average Jaccard coefficient between local expert lists of around 60 to 80%. In other words, on average, 20 to 40% of the top-10 local experts for the same query are different, when we compare the directly labeled expertise approach versus a particular expertise propagation approach.

This indicates that the expertise propagation approaches are bringing in new signals of local expertise from the social and link-based connections of users; in our continuing work we are investigating methods to integrate these two types of topical authority by finding more diverse experts from each of these alternative approaches.

Table 5: Comparing LocalRank to five alternatives

Approach	$Rating$@10	$Precision$@10	$NDCG$@10
DLE	0.225	0.088	0.199
NE	0.141	0.114	0.487
MP (followers)	0.058	0.031	0.234
MP (list)	0.070	0.038	0.301
MP (on-topic)	1.059	0.628	0.750
LR: SP + DLE	1.334	**0.842**	**0.896**
LR: SP + EP + Friendship	**1.354**	0.838	0.884

6.3.3 Comparing LocalRank versus Alternatives

So far we have investigated the impact of local authority and the impact of topical authority on the quality of local experts found by the LocalRank framework. In this section, we compare LocalRank to the five alternative local expert finding approaches described in the experimental setup over the set of 10 general topics.

We first report the results for the five baselines in Table 5. We see that relying solely on topical authority – Directly Labeled Expertise (DLE) – with no notion of localness, results in a very low $Rating$@10, $Precision$@10, and $NDCG$@10. Similarly, relying solely on local authority – Nearest (NE), Most Popular in Town by Followers Count (MP followers), and Most Popular in Town by Listed Count (MP list) – with no notion of topical authority also leads to very poor results. Since local experts are defined both by their localness and their on-topic expertise, these results confirm our intuition driving the LocalRank approach to combine both factors. The baseline that does incorporate both factors – Most Popular in Town by Listed Count on Topic (MP (on-topic)) – captures this notion of local expertise by rewarding candidates who have been listed on many Twitter lists on the topic of interest within a particular location. We see in the table that this approach significantly outperforms the single factor alternatives ($Rating$@10 of 1.059, $Precision$@10 of 0.628, and $NDCG$@10 of 0.750).

We compare all five of these baselines to two versions of LocalRank. Both consider local authority based on Spread-based Proximity (SP); one uses directly labeled expertise (SP + DLE), while the other uses expertise propagation (SP + EP + Friendship) over the friendship graph. We see similar qualitative results when evaluating Focus Proximity (FP) and alternative expertise propagation approaches. Both approaches significantly outperform the four single factor baselines, as well as significantly outperforming the best alternative incorporating both local and topical authority, MP (on-topic). We see for LocalRank (SP + DLE) a $Rating$@10 of 1.334, $Precision$@10 of 0.842, and $NDCG$@10 of 0.896. For LocalRank (SP + EP + Friendship), we have $Rating$@10 of 1.354, $Precision$@10 of 0.838, and $NDCG$@10 of 0.884. These results confirm the effectiveness of the LocalRank approach and the importance of carefully leveraging the large-scale geo-tagged list relationships on Twitter.

Continuing this investigation, we report the results of the different LocalRank approaches versus the best performing baseline in in Table 6. We see that the Expertise Propagation approaches generally perform slightly better than the Directly Labeled Expertise approach in terms of $Rating$@10 and $Precision$@10. This suggests that adding in social connections bring in more signals to identify top local experts. In particular, LocalRank with expertise propagation coupled with the social graph of list-labeling relationships generates the best performance, with $Rating$@10 of 1.354 (an improvement of 27.6% over MP (on-topic)), $Precision$@10 of 0.847 (an improvement of 34.9%), and $NDCG$@10 of 0.886 (an improvement of 18.1%). However, in terms of $NDCG$@10, we see that the simpler DLE approach performs slightly better. But in all cases, the LocalRank approach outperforms the alternative.

Table 7: Comparing LocalRank versus the best performing alternative over finer topics

Approach	$Rating$@10	$Precision$@10	$NDCG$@10
MP (on-topic)	0.782	0.526	0.707
LR: SP + DLE	**0.924**	**0.583**	**0.851**
LR: SP + EP + Friendship	0.871	0.538	0.846
LR: SP + EP + List-labeling	0.868	0.535	0.837
LR: SP + EP + List-peer	0.865	0.533	0.844

6.3.4 LocalRank: Local Experts Over Finer Topics

Finally, we drill down from general topics to more fine-grained topics, to investigate the ability of local expertise finding approaches to handle these more specific cases. Here we evaluate the proposed LocalRank approaches via the refined topics under the "food", and "startup" scenarios. We report the performance using the best parameter settings for each of the proposed approaches. In this experiment, we set local authority as using Spread Proximity and expertise propagation relies on a damping factor of 0.30.

Table 8: How well does LocalRank perform on finer topics?

Query Topic	$Rating$@10	$Precision$@10	$NDCG$@10
barbecue	0.631	0.404	0.787
seafood	0.825	0.525	0.868
pizza	0.775	0.425	0.712
brewery	**1.178**	**0.738**	**0.928**
winery	0.763	0.475	0.744
entrepreneur	**1.248**	**0.800**	**0.921**
venture capital	1.180	0.663	0.956
angel investor	0.923	0.638	0.846
incubator	0.660	0.413	0.732
founder	0.995	0.688	0.786

Table 7 presents the local expert finding results for the four types of LocalRank versus the best performing alternative (MP (on-topic)). We observe that once again the LocalRank approaches outperform the best-performing alternative in all cases. However, we notice that the performance for these finer topics is worse than what we observed for the more general topics. For example, LocalRank with Directly Labeled Expertise performs the best with $Rating$@10 of 0.924, $Precision$@10 of 0.583, and $NDCG$@10 of 0.851 over these finer topics. But the same approach over the more general topics results in an average $Rating$@10 of nearly 0.4 points higher. Similarly, we see improved performance over the other metrics in the general topic case. We believe these results reflect two challenges: (i) First, it is fundamentally more challenging to identify local experts for more refined topics. For example, it may be easier to assess whether someone is a "food" expert, rather than that they are an expert in a specific topic like "barbecue". (ii) Second, there is inherent data sparsity at the level of these finer topics. The number of candidates for a finer topic in a query location is much smaller compared to the number of candidates for a general topic in the same query location. For example, we observe that the approaches consider the probable No. 1 barbecue expert in Texas – Daniel Vaughn – as a local expert for barbecue for query locations of Chicago and San Francisco, in addition to his natural expertise in Houston. For these two distant locations, Vaughn is often a top choice since there are few barbecue candidates recognized in the location.

In our continuing work, we are investigating the contours of expertise across the country, so that topics with a strong regional factor (like Barbecue, with its traditional centers in Texas, North Carolina, and the Midwest) can be balanced with topics of expertise

Table 6: The impact of expertise propagation on LocalRank versus the best performing alternative

Approach	$Rating@10$	% of Improvement	$Precision@10$	% of Improvement	$NDCG@10$	% of Improvement
MP (on-topic)	1.059	–	0.628	–	0.750	–
LR: DLE + Local Authority	1.334	26.0%	0.841	33.9%	**0.897**	**19.6%**
LR: EP + Friendship Graph	1.354	27.6%	0.838	33.4%	0.884	17.9%
LR: EP + List-labeling Graph	**1.354**	**27.6%**	**0.847**	**34.9%**	0.886	18.1%
LR: EP + List-peer Graph	1.345	27.0%	0.844	34.4%	0.887	18.3%

that are found nearly everywhere (e.g., the more general "foodies"). Along these lines, we show in Table 8 the results of LocalRank (SP + DLE) for each of the fine-grained topics. As we observed in our original investigation of Twitter lists, where we observed topics like "food" being more local than topics like "technology", here we see great variation in local expertise finding across these different subtopics.

7. CONCLUSION

The exponential growth in social media over the past decade has recently been joined by the rise of location as a central organizing theme of how users engage with online information services and with each other. Enabled by the widespread adoption of GPS-enabled smartphones, users are now forming a comprehensive geo-social overlay of the physical environment of the planet. In this paper, we have argued for leveraging these geo-spatial clues embedded in Twitter lists to power new local expert finding approaches. We have proposed and evaluated the LocalRank framework for finding local experts, by integrating both a candidate's local authority and topical authority. We have seen that assessing local authority based on the spread and focus-based proximity of a candidate's "core audience" – that is, the users who have labeled him – can lead to good estimates of local authority and ultimately to high-quality local expert finding. Through an investigation of 56 queries coupled with over 11,000 individual judgments from Amazon Mechanical Turk, we have seen high average precision, rating, and NDCG in comparison with alternatives. In our continuing work, we are interested to (i) further investigate the borders of "localness" by investigating when an expert is considered a local expert versus a regional expert; (ii) enhance our current LocalRank approach with temporal signals to capture expertise evolution; and (iii) incorporate the detected local experts into a prototype system that can direct information needs to local experts who are considered authoritative and responsive on the local topic of interest.

8. ACKNOWLEDGEMENTS

This work was supported in part by NSF grant IIS-1149383. Any opinions, findings and conclusions or recommendations expressed in this material are the author(s) and do not necessarily reflect those of the sponsors.

9. REFERENCES

[1] J. Antin, M. de Sa, and E. F. Churchill. Local experts and online review sites. In *CSCW 2012*.

[2] L. Backstrom, E. Sun, and C. Marlow. Find me if you can: improving geographical prediction with social and spatial proximity. In *WWW 2010*.

[3] K. Balog, L. Azzopardi, and M. De Rijke. Formal models for expert finding in enterprise corpora. In *SIGIR 2006*.

[4] T. Brants. Inter-annotator agreement for a german newspaper corpus. In *Proceeding of the 2nd International Conference on Language Resources and Evaluation*, 2010.

[5] C. S. Campbell, P. P. Maglio, A. Cozzi, and B. Dom. Expertise identification using email communications. In *CIKM 2003*.

[6] Z. Cheng, J. Caverlee, H. Barthwal, and V. Bachani. Finding local experts on twitter. In *WWW 2014*.

[7] Z. Cheng, J. Caverlee, and K. Lee. You are where you tweet: a content-based approach to geo-locating twitter users. In *CIKM 2010*.

[8] E. H. Chi. Who knows?: searching for expertise on the social web: technical perspective. *Commun. ACM*, 55(4), Apr. 2012.

[9] H. Cramer, M. Rost, and L. E. Holmquist. Performing a check-in: emerging practices, norms and'conflicts' in location-sharing using foursquare. In *Proceedings of the 13th International Conference on Human Computer Interaction with Mobile Devices and Services*, MobileHCI '11, 2011.

[10] J. L. Fleiss, J. Cohen, and B. Everitt. Large sample standard errors of kappa and weighted kappa. *Psychological Bulletin*, 72(5), 1969.

[11] S. Ghosh, N. Sharma, F. Benevenuto, N. Ganguly, and K. Gummadi. Cognos: crowdsourcing search for topic experts in microblogs. In *SIGIR 2012*.

[12] T. H. Haveliwala. Topic-sensitive pagerank. In *WWW 2002*.

[13] B. Hecht, L. Hong, B. Suh, and E. H. Chi. Tweets from justin bieber's heart: the dynamics of the location field in user profiles. In *SIGCHI 2011*.

[14] K. Kamath, J. Caverlee, K. Lee, and Z. Cheng. Spatio-temporal dynamics of online memes: A study of geo-tagged tweets. In *WWW 2013*.

[15] H. Kwak, C. Lee, H. Park, and S. Moon. What is twitter, a social network or a news media? In *WWW 2010*.

[16] T. Lappas, K. Liu, and E. Terzi. A survey of algorithms and systems for expert location in social networks. In *Social Network Data Analytics*. Springer, 2011.

[17] W. Li, C. Eickhoff, and A. P. de Vries. Geo-spatial domain expertise in microblogs. In *ECIR 2014*.

[18] X. Liu, W. B. Croft, and M. Koll. Finding experts in community-based question-answering services. In *CIKM 2005*.

[19] C. Marshall and F. Shipman. Experiences surveying the crowd: Reflections on methods, participation, and reliability. In *ACM Web Science 2013*.

[20] M. Naaman, Y. J. Song, A. Paepcke, and H. Garcia-Molina. Automatic organization for digital photographs with geographic coordinates. In *JCDL 2004*.

[21] S. Nowak and S. Rüger. How reliable are annotations via crowdsourcing: a study about inter-annotator agreement for multi-label image annotation. In *Proceedings of the international conference on Multimedia information retrieval*, 2010.

[22] A. Pal and S. Counts. Identifying topical authorities in microblogs. In *WSDM 2011*.

[23] S. Scellato, A. Noulas, R. Lambiotte, and C. Mascolo. Socio-spatial properties of online location-based social networks. In *ICWSM 2011*.

[24] W. R. Tobler. A computer movie simulating urban growth in the detroit region. *Economic Geography*, 46:pp. 234–240, 1970.

[25] J. Weng, E. P. Lim, J. Jiang, and Q. He. TwitterRank: finding topic-sensitive influential twitterers. In *WSDM 2010*.

[26] M. Ye, D. Shou, W.-C. Lee, P. Yin, and K. Janowicz. On the semantic annotation of places in location-based social networks. In *SIGKDD 2011*.

[27] M. Ye, P. Yin, W.-C. Lee, and D.-L. Lee. Exploiting geographical influence for collaborative point-of-interest recommendation. In *SIGIR 2011*.

[28] M. Yuan, L. Chen, and P. S. Yu. Personalized privacy protection in social networks. *PVLDB*, 4(2):141–150, 2010.

[29] J. Zhang, M. S. Ackerman, and L. Adamic. Expertise networks in online communities: structure and algorithms. In *WWW 2007*.

[30] J. Zhang, J. Tang, and J. Li. Expert finding in a social network. In *Advances in Databases: Concepts, Systems and Applications*, volume 4443 of *Lecture Notes in Computer Science*. Springer Berlin Heidelberg, 2007.

[31] V. W. Zheng, Y. Zheng, X. Xie, and Q. Yang. Collaborative location and activity recommendations with gps history data. In *WWW 2010*.