

Mining half a billion topical experts across multiple social networks

Nemanja Spasojevic¹ · Prantik Bhattacharyya¹ · Adithya Rao¹

Received: 11 November 2015 / Revised: 23 June 2016 / Accepted: 24 June 2016
© Springer-Verlag Wien 2016

Abstract Mining topical experts on social media is a problem that has gained significant attention due to its wide-ranging applications. Here we present the first study that combines data from four major social networks—Twitter, Facebook, Google+ and LinkedIn—along with the Wikipedia graph and Internet webpage text and metadata, to rank topical experts across the global population of users. We perform an in-depth analysis of 37 features derived from various data sources such as message text, user lists, webpages, social graphs and Wikipedia. This large-scale study includes more than 12 billion messages over a 90-day sliding window and 58 billion social graph edges. Comparison reveals that features derived from Twitter Lists, Wikipedia, Internet webpages and Twitter Followers are especially good indicators of expertise. We train an expertise ranking model using these features on a large ground-truth dataset containing almost 90,000 labels. This model is applied within a production system that ranks over 650 million experts in more than 9000 topical domains on a daily basis. We provide results and examples on the effectiveness of our expert ranking system, along with empirical validation. Finally, we make the topical expertise data available through open REST APIs for wider use.

Keywords Social media · Online social networks · Topic expertise · Large-scale topic mining

1 Introduction

Millions of people use online social networks as platforms for communication, creating billions of interactions daily. The data from these social networks provide unique opportunities to understand the nature of social interactions, at a scale that was not previously possible. While a large portion of these interactions are casual, some interactions carry useful topical information. There has been a growing interest in identifying *topical experts* on social media using this information (Guy et al. 2013; Cheng et al. 2013; Zhao et al. 2013; Bi et al. 2014). While most studies focus on identifying the top experts in a topic, an under-explored area of research is to rank all users by their topical expertise. Identifying and ranking intermediate topical experts, in addition to the top experts, can provide great value to users via applications that involve question answering, crowd-sourced opinions, recommendation systems and influencer marketing.

Expertise may manifest itself via different data sources on different social networks. Leveraging information from multiple social networks, and combining them with webpage metadata, can lead to a more complete understanding of users' expertise, compared to using any single source. Additionally, a scalable solution to such a problem must be able to process hundreds of millions of users and identify experts for thousands of topics.

Our contributions in this study are as follows:

1. **Feature diversity** We present a comprehensive set of **37 features** that indicate topical expertise on social

✉ Prantik Bhattacharyya
prantik@klout.com

Nemanja Spasojevic
nemanja@klout.com

Adithya Rao
adithya@klout.com

¹ Lithium Technologies/Klout, San Francisco, CA, USA

networks and provide an in-depth analysis of their predictability and coverage. We derive these features from more than **12 billion message texts, 23 million Twitter Lists, 58 billion social graph edges, 1 million webpages** and **20,000 Wikipedia pages**.

2. **User and topic coverage** We rank over **650 million experts** with varying levels of topical expertise across **9000 topics**, including popular as well as niche topics.
3. **Multiple networks** We incorporate data from four major social networking platforms: Twitter (TW), Facebook (FB), Google+ (GP) and LinkedIn (LI), and combine it with data from Wikipedia (WIKI) and Internet webpage text and metadata.
4. **Evaluation** We evaluate the features on a ground-truth dataset containing almost **90,000 labels**. As far as we know, this is one of the largest datasets used for evaluating topical expertise.
5. **Open data** We make the rankings for top experts and expertise topics for a Twitter user available through **open public APIs**. We also make the topic ontology available as an open dataset.

We perform our study and analysis on a full production system available on the Klout platform. Klout¹ is a social media platform that aggregates and analyzes data from social networks like Twitter, Facebook, Google+ and LinkedIn.

2 Related work

A growing body of academic as well as industrial research has focused on mining topical experts (Popescu 2013; Campbell et al. 2003). In Guy et al. (2013), Kolari et al. (2008), Ehrlich and Shami (2008), enterprise users were studied. Among social media platforms, Twitter has been undoubtedly the most studied due to the large volume and public nature of its data. The work in Cognos Ghosh et al. (2012) presented a system utilizing only Twitter-List features, and results showed an improvement over Twitter's in-house system (Gupta et al. 2013). In our system, we generate features based on users added to lists as well as on which lists a user creates or subscribes to. Twitterrank Weng et al. (2010) presented a system that identifies topical influencers based on the link structure shared between users and on the information extracted from the 'bio' in the Twitter profile. The work in Pal and Counts (2011) presents a multi-feature approach toward expert mining. They use a set of 15 features for characterizing social media authors based on both nodal and topical properties and present results across 20 different topics.

The work by LinkedIn Ha-Thuc et al. (2015) has focused on determining topic experts based on a LinkedIn skills dataset available to the company internally. Other works have focused solely on niche topics: Bi et al. (2014) have focused on mining experts in landscape photography, and Zhang et al. (2007) analyzed Java forums to identify experts in the Java community.

Multiple works have proposed applications that utilize a topic expert list. Cognos and Twitter's Who To Follow service focused on services to recommend experts for the purpose of 'following.' Research work Wu et al. (2011) has also utilized experts as a starting set to study other social media properties like extent of homophily within categories, speed of information flow and content lifespan. Works have been presented that utilize the community around topics to understand top influences within a topic (Bi et al. 2014). Topical expertise of popular users has also been used to mine topics of interest for other users (Bhattacharya et al. 2014). To understand organizational experts, work in Fu et al. (2007) proposed using a seed of already identified top experts and then followed the network graph to identify other potential experts. Many works have focused on question answering services where an expert system provided the core platform to route questions and match experts to askers (Zhao et al. 2013; Liu et al. 2005; Jurczyk and Agichtein 2007; Adamic et al. 2008).

3 Problem setting

We identify experts in topical domains as those users who produce and share topical information which is recognized as relevant and reliable by other users in the network. We aim to capture topical interactions that indicate expertise and derive features for them. We examine how these features from different sources behave across the global population of users.

To gain insights into the performance of each feature, we use user in-degrees in the graph as the comparison metric. Since we are dealing with multiple networks, we combine degree information of a user on different graphs into a single quantity that we call *connectivity*. For a user u , we define the *connectivity*, C_u , as:

$$C_u = \|\mathbf{c}_u\| \quad (1)$$

where $\mathbf{c}_u = [c_u^1, c_u^2, \dots, c_u^n]$ is the connectivity vector and each element c_u^k is the in-degree for the user in the network k , e.g., Twitter Followers, Facebook Friends and LinkedIn Connections. The connectivity may vary from single digit edges for passive users, to hundreds of millions for celebrities, capturing the entire gamut of users.

In Fig. 1, we plot the number of users against connectivity. We observe a distribution where a large number of

¹ Klout platform is a part of Lithium Technologies, Inc.

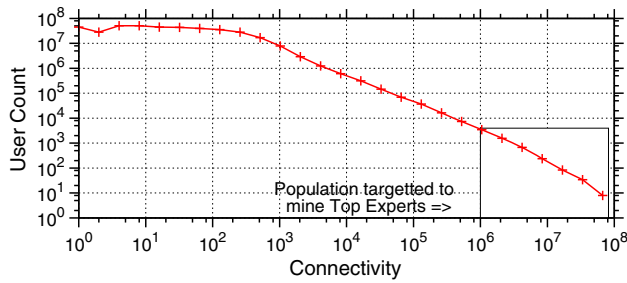


Fig. 1 User count distribution versus connectivity

users have small connectivity and only a small fraction of users have very large values of connectivity. Most results from previous works like Ghosh et al. (2012), Pal and Counts (2011) that find top experts are mainly applicable to the head of this distribution. Here we instead aim to rank experts over the full distribution, i.e., on the head as well as the long tail of expertise.

3.1 Problem statement

At Klout, topics are represented as entries in a hierarchical ontology tree, \mathcal{T} . The tree structure has three levels: *super*, *sub* and *entity*, each with 15, 602 and 8551 topics, respectively. More details are available in our earlier work (Spasojevic et al. 2014).

We wish to compute an expert score for each user-topic pair. To begin, we define a feature vector $\mathcal{F}(u, t_i)$ for a user u and a topic t_i as:

$$\mathcal{F}(u, t_i) = [f_1(u, t_i), f_2(u, t_i), \dots, f_m(u, t_i)] \quad (2)$$

where $f_k(u, t_i)$ is the feature value associated with a specific feature f_k . The normalized feature values are denoted by $\hat{f}_k(u, t_i)$ and the normalized feature vector is represented as: $\hat{\mathcal{F}}(u, t_i) = [\hat{f}_1(u, t_i), \hat{f}_2(u, t_i), \dots, \hat{f}_m(u, t_i)]$.

The expert score for a user-topic pair denoted by $\mathcal{E}(u, t_i)$ is computed as the result of the dot product of a weight vector \mathbf{w} and the normalized feature vector, $\hat{\mathcal{F}}(u, t_i)$.

$$\mathcal{E}(u, t_i) = \mathbf{w} \cdot \hat{\mathcal{F}}(u, t_i) \quad (3)$$

The weight vector is computed with supervised learning techniques, using labeled ground-truth data.

Our ground-truth data collection system generates user-pair labels across multiple topics. For a topic t_i , the labeled data between two users are defined as:

$$\text{label}(u_1, u_2, t_i) = \begin{cases} +1 & \text{if } u_1 \text{ is voted up,} \\ -1 & \text{if } u_2 \text{ is voted up.} \end{cases} \quad (4)$$

For a topic t_i and two users u_1 and u_2 , we define the feature difference vector as:

$$\hat{\mathcal{F}}_{\Delta}(u_1, u_2, t_i) = \hat{\mathcal{F}}(u_1, t_i) - \hat{\mathcal{F}}(u_2, t_i) \quad (5)$$

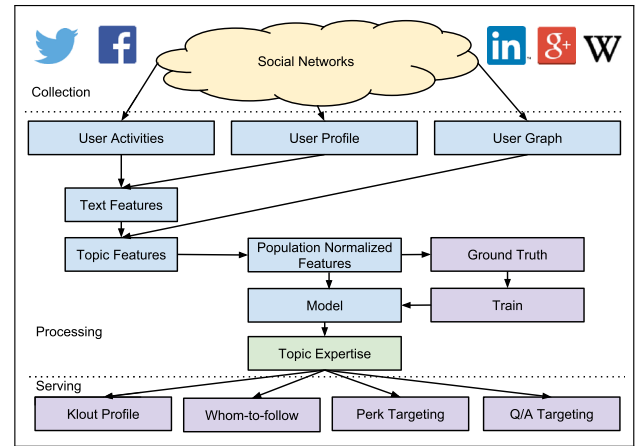


Fig. 2 Topic expertise pipeline

For a feature f_k , the corresponding element in the vector $\hat{\mathcal{F}}_{\Delta}(u_1, u_2, t_i)$ is represented as $\hat{f}_{k\Delta}(u_1, u_2, t_i)$. We can thus compare the expertise of a user pair by operating on their feature difference vector $\hat{\mathcal{F}}_{\Delta}(u_1, u_2, t_i)$.

3.2 System details

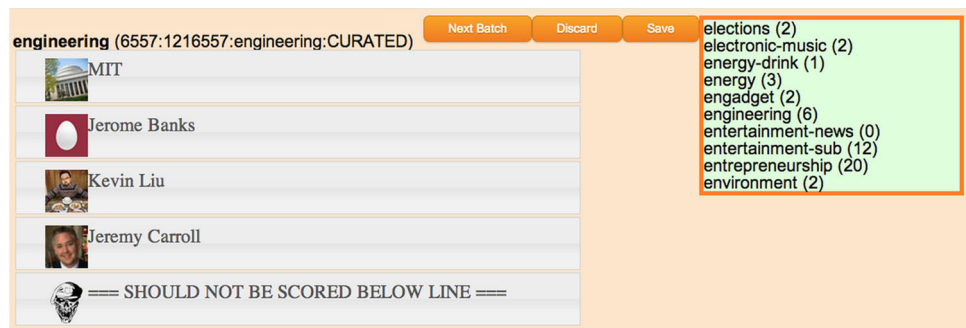
Figure 2 presents an overview of our production system. When a user registers on Klout, he connects one or more social networks and grants permission to Klout to collect and analyze his data. We use OAuth tokens provided by registered users to collect data from Facebook, LinkedIn and Google+. We also collect data about users' Twitter graph and list-mentions using OAuth tokens. Klout partners with GNIP to collect public data available via the Twitter Mention Stream.² We do not distinguish between human, organizational and non-human social media accounts and refer to all accounts as users in the rest of the paper.

We collect, parse, extract and normalize features for hundreds of millions of users daily. We use scalable infrastructure in the form of Hadoop MapReduce and Hive to bulk process large amounts of data. The daily resource usage for uncompressed HDFS data reads and writes during feature generation is: 55.42 CPU days, 6.66 TB reads, 2.33 TB writes.

4 Ground truth

Ground-truth data were collected in a controlled experiment via trusted internal evaluators. The data collection tool web UI is shown in Fig. 3.

² <http://support.gnip.com/sources/twitter/overview.html>.

Fig. 3 Ground-truth tool web UI**Table 1** Ground-truth statistics

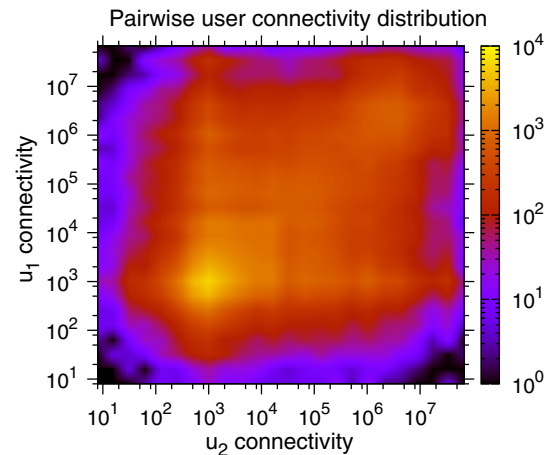
Labels for (u_i, u_j, t_k)	88,995
Unique triplets (u_i, u_j, t_k)	84,964
Unique user pairs (u_i, u_j)	66,656
Unique users	4448
Average list length	8.19
Unique topics	746
Number of evaluators	38

All evaluators were given guidelines on how to use the tool, but no specifics on how to interpret expertise. For a wide range of topics, evaluators were shown a user list asked to sort the users in order of expertise. Evaluators also had the option to mark a user as un-sortable or irrelevant to the topic. To reduce ambiguity in the data, an evaluator only judged users whom he or she was familiar with, and a personalized evaluation dataset was created for each evaluator. The dataset included selected users from an evaluator's outgoing social graph such as Facebook Friends or Twitter Following; or users with whom the evaluator had interacted through Facebook comments, Twitter retweets and so on. Users from the dataset were placed in an evaluator's topic list if at least one of their features had a nonzero value for the given topic. Table 1 presents details about the collected dataset.

For each topic t and evaluator e with input user list $U_{in}(e, t)$, we obtain the sorted list $U_{sorted}(e, t)$, which excludes the set of un-sortable users. The sorted list is represented as: $U_{sorted}(e, t) = [u_1^s, u_2^s, \dots, u_N^s]$. Thus, as per the evaluator, $\mathcal{E}(u_i^s, t) > \mathcal{E}(u_j^s, t)$, for $i, j \in (1..N), i > j$.

Unique user pairs (u_i, u_j) are created from the quadratic explosion of U_{sorted} , with $i, j \in (1..N), i > j$. A label (+1, -1) is generated for the pair (u_i, u_j) as described in Eq. 4, depending on the users' relative position on the list. The training and data evaluation are performed on $\hat{\mathcal{F}}_{\Delta}(u_i, u_j)$ for each (u_i, u_j) .

Since our goal is to study the feature distribution for the entire user population, it is important we collect ground truth for users with a wide range of connectivity. In Fig. 4,

**Fig. 4** Ground-truth connectivity distribution

we plot the connectivity distribution for users in the ground-truth dataset. The heatmap represents the numbers of users for a given pair of connectivity. For users with connectivity between 10^2 and 10^7 , we observe that we have a good coverage in the number of evaluations, with the most evaluations for users between 10^2 and 10^4 .

Since expertise is somewhat subjective based on evaluators' perception, a strict definition of expertise was not provided to evaluators. We show the label consensus among the human evaluators within the ground truth in Fig. 5a. We define label consensus as the number of voted up labels for a pair $P(u_i^s, u_j^s)$ divided by total votes casted for the pair and topic. One can observe that when two unique evaluators are asked to order a user pair, in 84 % of cases, evaluators agree on the ordering. As the number of casted votes grows, consensus drops to the minimum of 80 % for 5 unique evaluators, after which consensus grows to about 90 % for 8 evaluators. We can conclude that humans do not fully agree on exact expertise ordering, thereby imposing fundamental limitations on supervised machine learned models.

Finally, we present the dependency of connectivity difference between evaluated user pair on consensus in Fig. 5b. Consensus exhibits an increasing trend from 82 to

Fig. 5 Label consensus distribution. **a** Number of votes casted. **b** Connectivity difference

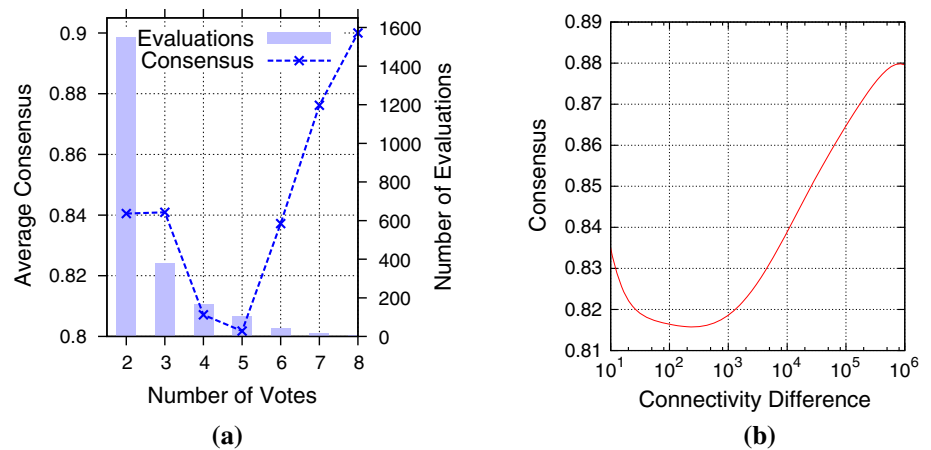


Table 2 Feature attribution nomenclature

	Attribution
GENERATED	Originally generated or authored content by the user, including posts, tweets, and profiles
REACTED	Content generated by another user (actor), but as a reaction to content originally authored by the user under consideration. This includes comments, retweets and replies
CREDITED	Text that is associated with the user without direct involvement from the user. Examples include tags and lists
GRAPH	Topics aggregated and derived from a user's friends, followers and following users

88 % as connectivity difference increases. This is expected as users with higher connectivities might be more easily recognized as experts.

5 Feature analysis

We present 37 features here that capture topical expertise for users from various sources. The textual inputs from each source are mapped to bags of phrases by matching against a dictionary of approximately 2 million phrases. Bags of phrases are then mapped to a topic ontology to create bags of topics, reducing the dimensionality of the text from 2 million to more than 9000 topics. These bags of topics are exploded, and for each user-topic pair (u, t_i) , we build the feature vector $\mathcal{F}(u, t_i)$. A more detailed discussion about feature generation is provided in our earlier work (Spasojevic et al. 2014).

Each element in the feature vector, $f_k(u, t_i)$, has a naming convention:

<Network>_<Source>_<Attribution>, which encodes the following characteristics: (a) the social network of origin, (b) the source data type and (c) the attribution relation of a feature to the user. The networks we consider include Twitter (TW), Facebook (FB), Google+ (GP), LinkedIn (LI) and Wikipedia (WIKI). Table 2 summarizes the attribution relations and their descriptions.

Below we describe the different source data types considered:

Message text We derive features as the frequency of occurrence of topics in text of messages posted by users, including original posts, comments and replies. These data sources are named under ‘MSG TEXT’ and provide useful topical information for all users who are actively posting and reacting to messages on the network. In the case of FB Fan Pages, we call this data source ‘PAGE TEXT,’ to differentiate from message text in personal FB pages. The feature vectors are thus named as TW_MSG_TEXT_GENERATED, TW_MSG_TEXT_REACTED, TW_MSG_TEXT_CREDITED and so on.

We present the number of message texts processed for each network in Table 3. Since we have to access to the entirety of Twitter’s public data, the volume of messages processed from Twitter is significantly more than the number of messages from Facebook.

User lists One of the most important data sources related to topical expertise is user lists on Twitter, as previously explored in Ghosh et al. (2012), Cheng et al. (2013). A user may be added to topical lists by other users, thereby marking him or her as an expert. We refer to this data source simply as ‘LIST.’ In addition, we also derive other features from the lists a user creates or subscribes to. The list-based features are derived from over 23 million lists corresponding to over 7.5 million unique users.

Table 3 Message text processed

Network	Number of message texts
Twitter	11 Billion
Facebook	696 Million
Facebook Pages	1.47 Billion
Google+	22 Million

Since only a subset of lists per user is available for collection from Twitter, we estimate this feature as $f_k(u, t_i) \approx L_c(u, t_i) \times \frac{L(u)}{L_c(u)}$ where $L_c(u, t_i)$ is the number of collected lists for the user on topic t_i , $L_c(u)$ is the total number of collected lists, and $L(u)$ is the true number of lists for the user, retrieved from the user's Twitter profile.

User profile We derive features based on the user-listed information available in user profiles. For a given user, we extract the skill and industry information from LinkedIn profiles to derive topic signals. We assign the number of followers of a company normalized over the number of followers for the particular industry as the feature value. Feature vectors are thus named as LI_SKILLS_GENERATED, LI_INDUSTRY_GENERATED.

Social graphs We leverage graph and peer-based information to derive topical signals from a user's connections. For a given user and topic, we aggregate the topic strength across all of his connections and scale this with the aggregated strength for the global population. Apart from 'FOLLOWERS,' these features may come from other graph sets such as 'FOLLOWING' and 'FRIENDS,' depending on the networks. This feature is especially important for the set of users who may not explicitly talk about certain topics through messages, but may yet be recognized as experts in their fields. For example, if a large number of experts in *finance* follow Warren Buffett on Twitter, then he can be identified as an expert in the topic, though he may not post messages related to *finance*. The features are derived from more than 56 billion follower and 2.7 billion following edges from Twitter and 1.69 billion friend edges from Facebook.

Webpage text and metadata Users often share and react to URLs on social media that are related to their topics of expertise. We extract and create features from the content body of such shared URLs and also from metadata such as categories associated with the shared URLs. These data sources, named 'URL' and 'URL META,' respectively, provide rich contextual information about the associated social media message. Inversely, articles that are published on online publishing platforms often contain useful metadata, including social identity information for the authors through the Open Graph Protocol and Twitter Creator cards.³ Such URLs and published articles provide a means

of bridging social network information with Internet webpage data, thereby enabling a richer understanding of content as it relates to users. We process 1 million such webpages that were the most shared or reacted over a 90-day rolling window, extracting features for over 50,000 users based on the popularity of the page. In this case, we derive the feature value as:

$$f_k(u, t_i) = \sum_{j=1}^N tf(t_i, BT_j^{SWWW}) \times n_j$$

where N is the total number of documents attributed to the user, $tf(t_i, BT_j^{SWWW})$ is the topic frequency in the bag of topics derived for the j th document, and n_j is the number of times the j th document was reacted upon in the network. The features from this data source, which we call 'SOCIAL WWW,' are especially important to attribute expertise to journalists and bloggers who write long-form articles and can be recognized as topical experts due to their authorship.

In addition to social networks and webpage text and metadata, we also consider Wikipedia data⁴ that provides information about a user's expertise. We manually identify mapping for over 20,000 users from Wikipedia page to the user's social network identity. The existence of a Wikipedia page for a user itself may be a strong signal that he is recognized for his expertise in some topic. However, in order to eliminate spurious pages, we do not consider merely the existence of a page to be a signal. Instead, we compute the inlink-to-outlink ratio for a Wikipedia page, which indicates the authority of the user's page. Using pagerank gives similar results, so we only present results for the inlink-to-outlink feature. The final feature values are computed as:

$$f_k(u, t_i) = tf(t_i, BT^{WIKI}(u)) \times \frac{L_{in}}{L_{out}}$$

where $tf(t_i, BT^{WIKI}(u))$ is the topic frequency in the bag of topics derived from the Wikipedia page, and L_{in} , L_{out} are the number of inlinks and outlinks to the page, respectively, in the full Wikipedia graph.

Note that the features described above are applicable for mining topical interests as well, but here we are focused on topical expertise. The major difference between the two problems is in feature normalization, where topical expertise features are scaled against the global population, and topical interest features are scaled with respect to the user. The problem of expertise has to scale to millions of users per topic, while that of interests has to scale to hundreds of topics per user.

³ <http://ogp.me>, <https://dev.twitter.com/cards/markup>.

⁴ <https://dumps.wikimedia.org/>.

5.1 Feature distribution

In this section, we describe four ways we analyze and visualize the coverage and predictability of each feature.

5.1.1 Feature distribution

The first column in Fig. 6 shows the population feature distribution for selected features. The distributions are plotted on the log-log scale, where the number of users is plotted on the y-axis against the raw feature values on the x-axis. Features such as Wikipedia and Google+ URLs are present for less than 10^5 users, while the other two features are present for a much greater number of users. We observe that when plotted on the log-log scale, the number of users has an almost linear relationship to the feature values. The plots suggest that most features could be modeled as power law distributions over the population of users under consideration. We therefore rescale the features by transforming them as follows:

$$\hat{f}_k(u, t_i) = \frac{\log(f_k(u, t_i))}{\max_{u_i \in U} \log(f_k(u_i, t_i))} \quad (6)$$

5.1.2 Feature connectivity distribution

The second column in Fig. 6 plots the number of users who possess the specified feature against the connectivity of the users. The Twitter List feature is present over the entire range of users, but Wikipedia is present only for users with high connectivity. The URL feature for Google+ on the other hand is present for only a small number of users with low-to-medium connectivity. These differences in coverage highlight the need to have features that can capture information for different sections of the population.

5.1.3 Ground-truth distribution

Figure 6 also shows the ground-truth distribution for the features in the third column. In this case, the number of labels in the ground truth is plotted against the normalized feature difference values between user pairs. The labels plotted had at least one user in the pair possessing a nonzero feature value. For a user pair (u_1, u_2) , we plot $\hat{f}_{k\Delta}(u_1, u_2, t_i)$ and $\hat{f}_{k\Delta}(u_2, u_1, t_i)$, allowing us to symmetrically visualize the differentiating nature of the feature over the ground truth. We observe that for Wikipedia,

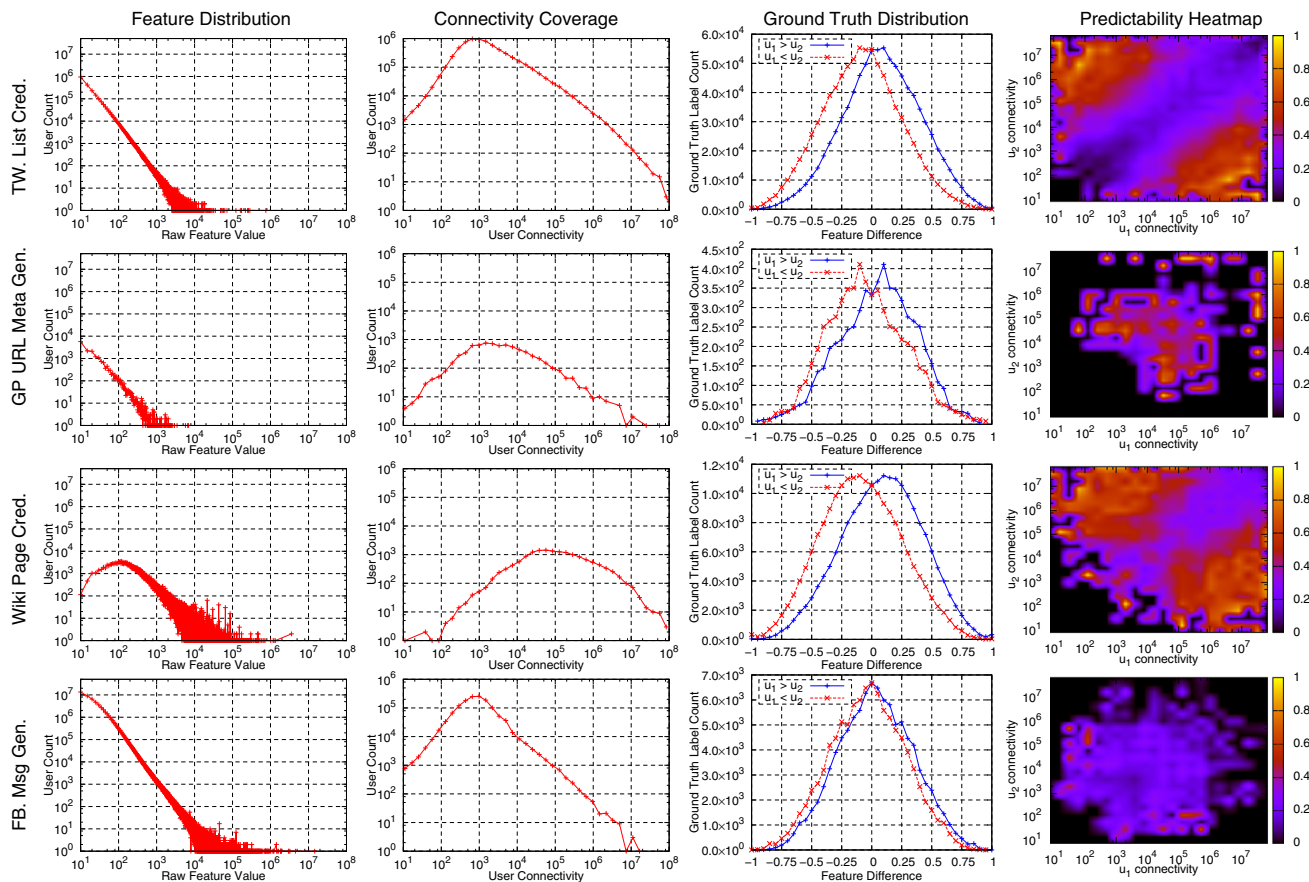


Fig. 6 Feature distribution (FD), connectivity coverage (CC), ground-truth distribution (GTD), predictability heatmap (PH) for selected features

Table 4 Feature catalog

Precision (P), recall(R), F1 score, coverage [% of corpus] (C), feature distribution (FD), connectivity coverage (CC), ground-truth distribution (GTD), predictability heatmap (.). The axes for feature distribution, connectivity coverage, ground-truth distribution and predictability heatmap are equivalent to the axes in Fig. 6

Attr.	Source	P	R	F1	C	FD	CC	GTD	.
Twitter									
GEN.	MSG TEXT	0.67	0.24	0.68	6.43				
	URL	0.66	0.29	0.67	1.99				
	URL META	0.70	0.13	0.71	1.46				
	LIST CREATOR	0.73	0.00	0.75	0.49				
	SOCIAL WWW	0.71	0.08	0.72	0.02				
REACT.	MSG TEXT	0.64	0.13	0.66	5.03				
	URL	0.61	0.20	0.63	1.92				
	URL META	0.72	0.16	0.72	0.24				
CRED.	MSG TEXT	0.71	0.33	0.71	4.23				
	LIST	0.70	0.36	0.72	0.59				
	URL	0.69	0.28	0.70	0.41				
	URL META	0.72	0.16	0.72	0.24				
	MSG HASHTAG	0.72	0.18	0.72	4.03				
GRAPH	FOLLOWERS	0.67	0.57	0.68	12.96				
	FOLLOWING	0.58	0.49	0.61	50.45				
Facebook									
GEN.	MSG TEXT	0.54	0.03	0.58	4.51				
	URL	0.52	0.03	0.57	0.23				
	URL META	0.54	0.01	0.61	0.14				
	SOCIAL WWW	0.73	0.01	0.69	0.00				
	PAGE TEXT	0.73	0.03	0.77	0.89				
REACT.	MSG TEXT	0.55	0.06	0.59	20.4				
	URL	0.51	0.04	0.55	1.49				
	URL META	0.58	0.01	0.62	0.86				
	PAGE TEXT	0.77	0.01	0.79	21.13				
CRED.	MSG TEXT	0.53	0.02	0.58	0.64				
	PAGE TEXT	0.74	0.01	0.78	0.02				
GRAPH	FRIENDS	0.53	0.18	0.58	0.31				
Google+									
GEN.	MSG TEXT	0.72	0.01	0.77	0.02				
	URL	0.69	0.01	0.74	0.00				
	URL META	0.71	0.00	0.76	0.00				
REACT.	MSG TEXT	0.69	0.00	0.69	0.20				
	URL	0.64	0.00	0.64	0.02				
	URL META	0.91	0.00	0.82	0.01				
CRED.	MSG TEXT	0.67	0.01	0.72	0.00				
LinkedIn									
GEN.	SKILLS	0.54	0.07	0.57	0.15				
	INDUSTRY	0.56	0.05	0.58	0.13				
Wikipedia									
CREDITED	WIKI PAGE	0.73	0.08	0.72	0.00				
Combined									
EXPERTISE SCORE		0.68	0.72	0.70	74.86				

Twitter Lists and Google+ URLs, the curves show greater separation compared to Facebook message text, which has almost overlapping curves. This shows that

the former features have a higher ability to identify experts compared to the features derived from Facebook message text.

5.1.4 Predictability heatmap

In the final column of Fig. 6, we observe feature predictability with respect to user connectivity, visualized as a heatmap. For the labeled pairs in the ground truth, we plot the average feature value difference for the instances when the feature correctly predicted the greater expert in the pair. The coordinates of each point are the connectivities of the first and the second user in a pair, and the brightness of the point indicates the average absolute value of the feature difference.

We observe that the Wikipedia and Twitter List features show a high ability to predict when the connectivity difference between the users is high. For Twitter Lists, connectivity coverage in the second plot of the same row shows that a majority of users have connectivity of less than 10^5 , but the heatmap shows low predictability for this cohort of users. Twitter Lists thus have weaker predictability in the long tail of the global population distribution, but still provide significant coverage to scale our systems and are good predictors for top experts.

For moderately connected users and users with similar connectivities, Google+ URLs feature is a good predictor and behaves in a complementary manner to Twitter Lists and Wikipedia. Facebook messages show poor predictability, which may be because Facebook messages are more conversational and not very indicative of topical expertise. Note that other Facebook features do have better predictability, but here we highlight this feature for contrast and comparison.

5.2 Feature catalog

Table 4 shows the precision, recall, F1 scores and the population coverage percentage for the full list of features used. The features are evaluated over the ground-truth data, where the prediction by a feature f_k is correct when the following relation holds:

$$\text{label}(u_1, u_2, t_i) \cdot \text{sgn}(\hat{f}_{k\Delta}(u_1, u_2, t_i)) > 0, \quad (7)$$

where $\text{sgn}(x)$ is the signum function.

We observe from Table 4 that the Twitter List-based features have some of the highest precision values on the ground-truth data. This corroborates the approach used in Ghosh et al. (2012), where Twitter Lists are used to identify experts. The SOCIAL WWW and Facebook Pages features show similar behavior to Twitter Lists, with high precision and F1 scores, but low coverage and recall values. Other features such as those derived from ‘Credited’ message texts and hashtags, and those derived from the followers of a user, prove to have high precision, F1 score and population coverage.

Features based on URLs, URL META and message text from Google+ have higher precision and F1 scores than Facebook and some Twitter features. One reason for this could be that users tend to post messages of greater length on Google+ compared to other networks (Spasojevic et al. 2014). For LinkedIn, though the precision and recall values are low, we observe from the heatmaps that they have high predictability for users with low connectivity, indicated by the bright spots near the lower left corners. Therefore, they are still useful in identifying experts in the long tail. Finally, since well-known personalities with Wikipedia pages are already recognized as experts in their domains, Wikipedia features have high precision and F1 scores.

To conclude, the best features for topical expertise in terms of F1 scores are Twitter Lists, Wikipedia pages, Social WWW, Facebook Fan Page text and URL metadata features. Graph and text-based features provide long tail coverage, though they are less predictive.

5.3 Connectivity analysis

As described in Sect. 4, our ground-truth dataset contains pairwise comparisons for users with varying amounts of connectivity. To study feature behavior across the spectrum of users, we plot the average feature difference in a user pair against the connectivity difference, for a few selected features in Fig. 7. This is effectively a lower-dimensional version of the predictability heatmaps that reveal new insights.

We observe that a few features shown have high predictability when the connectivity difference is low. This indicates that even when user network sizes are similar, their expertise can be differentiated with these features. The predictability for Twitter follower feature increases monotonically with connectivity difference. Finally, the SOCIAL WWW and Wikipedia features suffer from significant dips in the mid-range and perform the best when comparing users with network size difference of 10^5 and above.

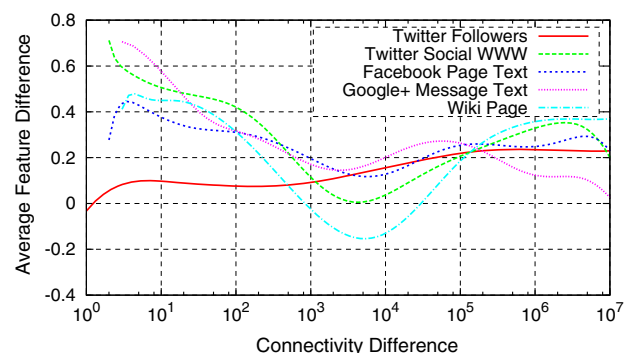


Fig. 7 Feature versus connectivity

Table 5 Super-topic user distribution across different networks

Topic	Combined	TW	FB	LI	GP	WIKI
technology	17.053	19.089	12.234	30.273	19.628	8.439
business	21.726	12.419	7.132	47.517	11.74	12.089
entertainment	18.231	16.616	16.728	3.22	12.597	31.658
sports-and-recreation	10.123	7.764	8.92	0.841	5.662	9.212
arts-and-humanities	4.859	7.387	6.516	5.404	9.653	9.581
lifestyle	5.253	6.779	10.859	0.926	8.766	6.059
science-and-nature	2.66	4.364	4.177	1.043	5.138	2.805
government-and-politics	4.227	6.009	5.592	2.421	3.888	5.726
education	4.802	2.853	4.568	2.791	2.72	4.705
food-and-drink	3.308	6.854	11.426	0.908	8.075	2.114
health-and-wellness	3.463	4.587	5.838	1.717	5.167	2.524
news-and-media	1.423	1.831	1.169	1.964	0.932	2.457
fashion	1.456	2.138	2.625	0.574	3.46	1.886
travel-and-tourism	0.641	0.744	1.532	0.306	1.341	0.425
hobbies	0.775	0.567	0.685	0.095	1.233	0.32

6 Expertise score

The features described above are combined into a feature vector for each user. The ground-truth dataset is split into training and test sets. The labels for the pairs in the training set are fit against the feature difference vectors $\hat{\mathcal{F}}_{\Delta}(u_1, u_2, t_i)$ using nonnegative least squares (NNLS) regression. We constrain the weights to be nonnegative because the features contributing to the score are designed to be directly proportional to expertise.

Model building is performed in a two-step process. In the first step, we build network-level models, based on the disjoint sets of features for each given network. In the second step we build a global model treating the network expertise scores obtained in the first step as features. Weighted network models are generated by multiplying the network-level weights with their corresponding weight from the global model. These weighted network models are then combined by concatenation into a single weight vector \mathbf{w} . This two-step process enables representation of sparse features derived from low population networks. The final expertise score for a user is computed by applying the weight vector to the user's normalized feature vector $\hat{\mathcal{F}}(u, t_i)$ as shown in Eq. 3.

The F measure for the expertise score is 0.70, covering 75 % of users across all networks. Table 4 shows additional performance metrics. Given that the human evaluator consensus is close to 84 %, this F measure is reasonably high.

6.1 Super-topics distribution

In Table 5, we examine the characteristics of topic expertise aggregated across users on different networks. We roll

up entities and sub-topics to super-topics to aid the visualization and reduce the topic dimension space from 9000 topical domains to 15. We plot percentage breakdown of super-topics on each network for the users on that network and also across all networks combined. We observe that users in each network have distinct topical expertise distribution. On Twitter and Google+, the super-topic 'technology' is the most represented one, whereas 'entertainment' is the most represented super-topic on Facebook. Facebook users are also experts in 'technology,' 'lifestyle' and 'food-and-drink.' On LinkedIn, most users are expert in 'business' and 'technology' and the representation of other super-topics drops off significantly. This is expected as LinkedIn is a professional networking platform. The leftmost column shows the distribution of expertise when all the network features are combined to build the combined expertise for users. The most represented topics are 'business,' 'entertainment' and 'technology.' We also observe that due to our multi-network approach, topical distribution across the users is not as skewed as the distribution for individual networks.

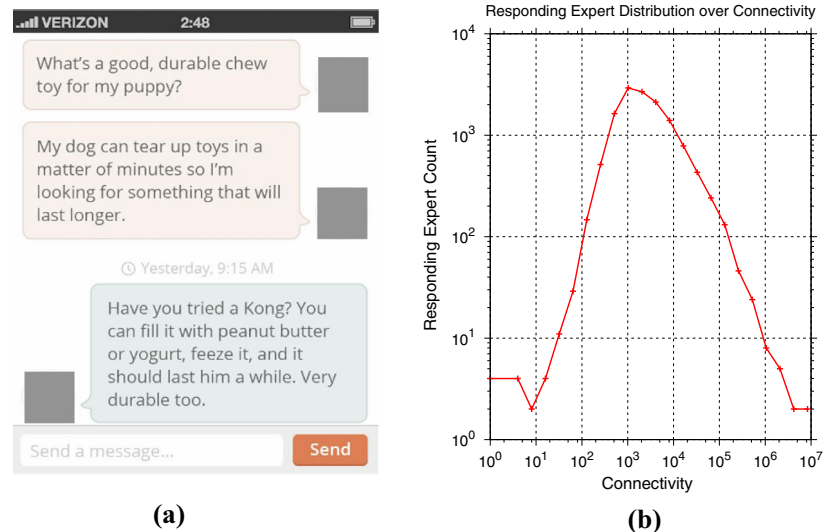
7 Applications

7.1 Application validation

To validate the expertise system, we developed an application⁵ where users could ask questions pertaining to their topics of interest. The questions were then routed to topical experts who could presumably answer the questions. The topical experts then could choose to respond to the questions posed to them, through a direct conversation with the

⁵ Cinch App from Klout, <http://tnw.to/t3FtH>.

Fig. 8 Q & A experts. **a** Q & A application. **b** Distribution of responding experts



user asking the question. An example of such an interaction in the application is shown in Fig. 8a.

The application generated over 30,000 such conversations, with more than 13,000 topical experts responding to questions asked. Figure 8b shows the connectivity distribution for the responding experts. We observe that a majority of experts who responded fell within the connectivity range of 500 to 5000, whereas top experts with connectivity values of more than 1,000,000 were rarely responsive. This shows the value of ranking users who are not necessarily the top experts in their fields while building applications used by the average population.

7.2 Expert examples

In Table 6, we present user examples with feature values for selected Twitter features and topics, along with their user population counts. For a topic such as ‘technology,’ in a typical *top-k* approach, users like *Google* and *TechReview* (MIT Technology Review) may be regarded as the top experts. But by ranking the full population, users like *Tim O'Reilly* and *Arrington* with lower feature values can also be identified as experts in the topic.

The effectiveness of a multi-source approach is evident in the examples of ‘politics’ and ‘machine learning.’ For ‘politics,’ *BarackObama* has the highest feature values for the Twitter follower graph feature, due to the large number of users who follow President Obama on Twitter. However,

Politico and *NYTimes* have high feature values for other features such as URL and SOCIAL WWW, enabling their identification as experts. As a niche topic, ‘machine learning’ attracts a relatively small community of users on the social platforms. For the most socially engaging accounts like *KDnuggets* and *Kaggle*, we observe that many nonzero features exist and it is relatively simple to identify these accounts as top experts. However, for passive users such as *ylecun*, the small individual feature values add up to recognize him as an expert.

In terms of coverage, the system scores 224K users for a niche topic like ‘machine learning,’ and approximately 3 million users for ‘San Francisco,’ a number that is close to half of the bay area’s population. Finally, Fig. 9 shows a screen shot of the top ranked users for a few selected topics.

8 Open API

We have opened the results of this paper via two REST API endpoints. The first endpoint returns the top experts for a specified topic. The second returns the expertise topics for a specified user. The documentation for using the REST APIs to retrieve these results is available at <http://klout.com/s/developers/research>. The results are provided in JSON format. The ontology of the 9000 topics used is available at <https://github.com/klout/opensdata>.

Listing 1: Ranked experts for 'politics'

```

{
  "topicSlug": "politics",
  "experts": [
    {
      "twitterUsername": "BarackObama",
      "rank": 1
    },
    {
      "twitterUsername": "washingtonpost",
      "rank": 2
    }
  ]
}

```

Table 6 Examples for Twitter Normalized Features

Attr.	Source	User with Normalized Feature Value					
		technology - 20M					
		google	techreview	abc7news-BayArea	tim-oreilly	arrington	# of Users
CRED.	LIST	0.80	0.97	0.26	0.76	0.72	246K
GEN.	S. WWW	-	0.79	-	-	-	10K
	URL	0.64	0.83	0.39	0.53	0.25	579K
	MSG	0.63	0.60	0.38	-	-	117K
	URL META	0.45	0.66	0.34	0.40	0.17	50K
GRAPH	FOLLOWER	0.96	0.77	0.62	0.86	0.77	9M
	FOLLOWING	0.46	0.56	0.73	0.51	0.52	13M
		machine-learning - 224K					
		kdnuggets	kaggle	jay kreps	andrewyng	ylecun	# of Users
CRED.	LIST	0.74	0.71	0.78	0.77	0.72	3K
GEN.	S. WWW	-	-	-	-	-	53
	URL	1.00	0.20	0.55	0.62	-	9K
	MSG	0.89	0.37	0.45	-	0.55	6K
	URL META	1.00	0.48	-	0.51	-	1K
GRAPH	FOLLOWER	0.97	0.96	0.85	0.96	0.86	151K
	FOLLOWING	0.49	0.44	0.61	-	0.21	87K
		politics - 114M					
		Barack-Obama	politico	nytimes	Huffington Post	white house	# of Users
CRED.	LIST	0.84	0.90	0.54	0.66	0.90	401K
GEN.	S. WWW	-	0.91	0.18	-	-	17K
	URL	0.82	0.92	0.80	0.84	0.87	3M
	MSG	0.69	0.82	0.79	0.75	0.75	1M
	URL META	0.75	0.91	0.53	0.54	0.52	392K
GRAPH	FOLLOWER	1.00	0.82	0.93	0.87	0.89	22M
	FOLLOWING	0.91	0.57	0.46	0.66	0.38	98M
		san-francisco - 3M					
		sfgov	abc7news-BayArea	only-insf	sf weekly	mayor ed lee	# of Users
CRED.	LIST	0.99	0.78	0.97	0.91	0.84	29K
GEN.	S. WWW	-	1.00	-	-	-	3K
	URL	0.31	0.82	0.77	0.77	0.27	421K
	MSG	0.45	0.64	0.53	0.55	-	80K
	URL META	0.22	0.89	0.90	0.73	-	24K
GRAPH	FOLLOWER	0.83	0.88	0.86	0.86	0.78	1M
	FOLLOWING	0.47	0.98	0.69	0.53	0.53	1M

Listing 2: Topical expertise percentiles for 'washingtonpost'

```

{
  "twitterUsername": "washingtonpost",
  "topicSetType": "expertise",
  "topicSet": [
    {
      "topicId": "7192018630863038134",
      "topicSlug": "journalism",
      "topicDisplayName": "Journalism",
      "topicScore": 0.9999999874118743
    },
    {
      "topicId": "8119426466902417284",
      "topicSlug": "politics",
      "topicDisplayName": "Politics",
      "topicScore": 0.9999999230887799
    }
  ]
}

```

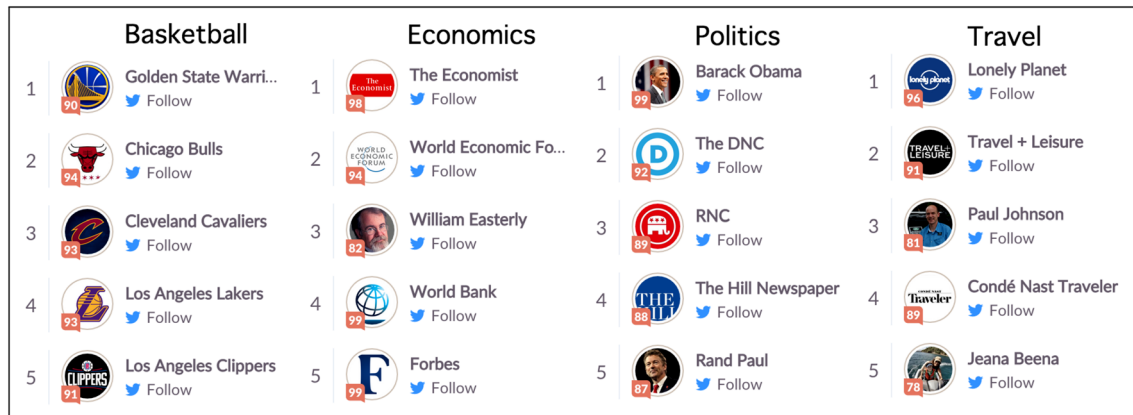


Fig. 9 Top ranked experts by topic; Snapshot on July 12, 2015

An example of the API response for top experts in 'politics' is provided in Listing 1.

Listing 2 presents an example response for the expertise topics for '@washingtonpost.' The score associated with a topic in the result is the percentile based on the expertise scores within the topic.

9 Conclusion and future work

In this study, we derive and examine a variety of features that predict online topical expertise for the full spectrum of users on multiple social networks. We evaluate these features on a large ground-truth dataset containing almost 90,000 labels. We train models and derive an expertise scores for over 650 million users, and make the lists of top experts available via APIs for more than 9000 topics.

We find that features that are derived from Twitter Lists, Facebook Fan Pages, Wikipedia and webpage text and

metadata are able to predict expertise very well. Other features with higher coverage such as those derived from Facebook Message Text and the Twitter Follower graph enable us to find experts in the long tail. We also found that combining social network information with Wikipedia and webpage data can prove to be very valuable for expert mining. Thus, a combination of multiple features that complement each other in terms of predictability and coverage yields the best results.

Further studies in this direction could unify cross-platform online expertise information, in addition to other data sources such as Freebase and IMDB. Another area that could be explored in the future is the overlap and differences in the dual problems of topical expertise and topical interest mining. To conclude, we provide an in-depth comparison of topical data sources and features in this study, which we hope will prove valuable to the community when building comprehensive expert systems.

Acknowledgments We thank Sarah Ellinger and Tyler Singletary for their valuable contributions toward this study.

References

- Adamic LA, Zhang J, Bakshy E, Ackerman MS (2008) Knowledge sharing and yahoo answers: everyone knows something. In: Proceedings of ACM conference on world wide web (WWW)
- Bhattacharya P, Zafar MB, Ganguly N, Ghosh S, Gummadi KP (2014) Inferring user interests in the twitter social network. In: Proceedings of the 8th ACM conference on recommender systems. ACM, pp 357–360
- Bi B, Tian Y, Sismanis Y, Balmin A, Cho J (2014) Scalable topic-specific influence analysis on microblogs. In: Proceedings of ACM conference on web search and data mining (WSDM)
- Bi B, Kao B, Wan C, Cho J (2014) Who are experts specializing in landscape photography? Analyzing topic-specific authority on content sharing services. In: Proceedings of ACM conference on knowledge discovery and data mining (KDD)
- Campbell CS, Maglio PP, Cozzi A, Dom B (2003) Expertise identification using email communications. In: Proceedings of ACM conference on information and knowledge management (CIKM)
- Cheng A, Bansal N, Koudas N (2013) Peckalytics: analyzing experts and interests on twitter. In: Proceedings of the 2013 international conference on management of data
- Ehrlich K, Shami NS (2008) Searching for expertise. In: Proceedings of the SIGCHI conference on human factors in computing systems. ACM, pp 1093–1096
- Fu Y, Xiang R, Liu Y, Zhang M, Ma S (2007) Finding experts using social network analysis. In: Web intelligence
- Ghosh S, Sharma N, Benevenuto F, Ganguly N, Gummadi K (2012) Cognos: crowdsourcing search for topic experts in microblogs. In: SIGIR
- Gupta P, Goel A, Lin J, Sharma A, Wang D, Zadeh R (2013) Wtf: the who to follow service at twitter. In: Proceedings of ACM conference on world wide web (WWW)
- Ha-Thuc V, Venkataraman G, Rodriguez M, Sinha S, Sundaram S, Guo L (2015) Personalized expertise search at linkedin. In: Proceedings of the IEEE international conference on big data. IEEE, pp 1238–1247
- Ido G, Avraham U, Carmel D, Ur S, Jacovi M, Ronen I (2013) Mining expertise and interests from social media. In: Proceedings of ACM conference on world wide web (WWW)
- Jurczyk P, Agichtein E (2007) Discovering authorities in question answer communities by using link analysis. In: Proceedings of ACM conference on information and knowledge management (CIKM)
- Kolari P, Finin T, Lyons K, Yesha Y (2008) Expert search using internal corporate blogs. In: Workshop on future challenges in expertise retrieval, SIGIR, vol 8
- Liu X, Croft WB, Koll M (2005) Finding experts in community-based question-answering services. In: Proceedings of ACM conference on information and knowledge management (CIKM)
- Pal A, Counts S (2011) Identifying topical authorities in microblogs. In: Proceedings of ACM conference on web search and data mining (WSDM)
- Popescu A-M, Kamath KY, Caverlee J (2013) Mining potential domain expertise in pinterest. In: UMAP workshops
- Spasojevic N, Yan J, Rao A, Bhattacharyya P (2014) Lasta: large scale topic assignment on multiple social networks. In: Proceedings of ACM conference on knowledge discovery and data mining (KDD), KDD'14
- Weng J, Lim E-P, Jiang J, He Q (2010) Twitterrank: finding topic-sensitive influential twitterers. In: WSDM'10
- Wu S, Hofman JM, Mason WA, Watts DJ (2011) Who says what to whom on twitter. In: Proceedings of ACM conference on world wide web (WWW)
- Zhang J, Ackerman MS, Adamic L (2007) Expertise networks in online communities: structure and algorithms. In: Proceedings of ACM conference on world wide web (WWW)
- Zhao T, Bian N, Li C, Li M (2013) Topic-level expert modeling in community question answering. In: SDM