Utilizing Linked Data Structures for Social-aware Search Applications

André Langer, Michael Krug, Luis Moreno³ and Martin Gaedke⁴

Abstract:

Improving the user experience and conversion rate by means of personalization is of major importance for modern e-commerce applications. Several publications in the past have already dealt with the topic of adaptive search result ranking and appropriate ranking metrics. Newer approaches also took personalized ranking attributes of a connected Social Web platform into account to form so called Social Commerce Applications. However, these approaches were often limited to data silos of closed-platform data providers and none of the contributions discussed the benefits of Linked Data in building social-aware e-commerce applications so far. Therefore, we present a first formalization of a scoring model for a social-aware search approach that takes user interaction from multiple social networks into account. In contrast to other existing solutions, our approach focuses on a Linked Data information management in order to easily combine social data from different social networks. We analyze the possible influence of friend activities to the relevance of a person's search intent and how it can be combined with other ranking factors in a formalized scoring model. As a result, we implement a first demonstrator built upon RDF data to show how an application can present the user an adaptive search result list depending on the users' current social context.

Keywords: Linked Data, Social Networks, Social Commerce, Ranking Factors, Scoring Model

1 Introduction

The Social Web enables billions of users worldwide to digitally connect with other people and share personal information as well as arbitrary content that is related to their personal interests. Additionally, a personalized interaction with content becomes possible. Typical forms of these interactions are primarily *link clicks*, but also *likes*, *comments*, *shares* or simply *visual attention*.

¹ Technische Universität Chemnitz, Distributed and Self-organizing Systems Group, Str. der Nationen 62, 09111 Chemnitz, Germany, andre.langer@informatik.tu-chemnitz.de

² Technische Universität Chemnitz, Distributed and Self-organizing Systems Group, Str. der Nationen 62, 09111 Chemnitz, Germany, michael.krug@informatik.tu-chemnitz.de

³ Technische Universität Chemnitz, Distributed and Self-organizing Systems Group, Str. der Nationen 62, 09111 Chemnitz, Germany, luis-antonio.moreno-gutierrez@informatik.tu-chemnitz.de

⁴ Technische Universität Chemnitz, Distributed and Self-organizing Systems Group, Str. der Nationen 62, 09111 Chemnitz, Germany, martin.gaedke@informatik.tu-chemnitz.de

As of the end of 2016, more than 2.34 billion social network profiles existed, which represents a global penetration of 31% of all living humans [St16]. The number of different social network platforms varies continuously and includes centralized global providers such as *Facebook* with currently over 1.79 billion monthly active users⁵ as well as domain-specific social platforms of medium-size and even decentralized social networks such as semantic-enabled profile pages using WebID⁶ and FOAF⁷. This social data is a valuable asset for external application developers, because users voluntarily provide additional personal meta-information about themselves on centralized web service access points in the Social Web, which they would not necessarily enter manually on other websites, i.e. e-commerce sites. The data can be used for contextualization and personalization purposes to further improve the user experience in modern web applications. Well-known use cases are keyword auto-suggestion and -resolution, recommendations of potentially interesting items and also the usage for a search result ranking.

Search results are traditionally displayed in some kind of a list view. In general, a user has to perform a search action beforehand, normally by explicitly providing certain keywords to a search interface. The objective of automatic search result ranking is to display the corresponding search results with the highest relevance probability for the currently acting user on top of such a result list, as long as the user or system has not selected any other fixed sort criterion (such as sort by creation date, lowest price, etc.). The relevance of a result item is mainly measured by its recent performance, either through the total number of clicks (hits) by other users, the conversation rate, or based on the search history of the user. With the evolution and establishment of Social Web platforms, search providers also started to take Social Web data into consideration for result ranking purposes⁸. However, due to security and privacy concerns, common social platform providers started to limit the exposed user data for external applications. Furthermore, with the upcoming enforcement of the EU General Data Protection Regulation (GDPR)9 in 2018, access to private data will be even more restricted. Also, current trends to re-decentralize the Social Web [Ma16] require adaptions in Social Commerce applications. Mainly, because classical approaches to handle personalized data based on solely one centralized Social Web platform do not fit any longer.

In this paper, we therefore exemplarily discuss a social-aware search application that is capable of taking ranking factors from multiple external Social Web profiles into account by handling them in a Linked Data representation. These ranking factors can be basic personal attributes describing the current user and its interests but also friend-network relationships and the activities they performed in the same web application in the past. Our main contribution is a scoring model for an adaptive result ranking algorithm taking all these social attributes, influence factors and activities into account. Additionally, we present a generic approach on how this data can be semantically managed on the basis

⁵ https://zephoria.com/top-15-valuable-facebook-statistics/

⁶ http://webid.info/spec/

⁷ http://xmlns.com/foaf/spec/

 $^{^8 \; \}texttt{https://stonetemple.com/googles-matt-cutts-understanding-social-identity-on-the-web-is-hard/}$

⁹ http://www.eugdpr.org/

of Linked Data in the form of RDF triples, which allows the combination of Social Web data from different social networks in one search activity.

The rest of the paper is structured as follows: First, Sect. 2 conceptually outlines challenges in a social-aware search application based on user data from heterogeneous, distributed social networks. In Sect. 3, we propose a first formalization of our adaptive search result ranking model that can be used for algorithmic purposes. Sect. 4 provides a proof of concept by presenting a web application that implements the presented formalization. Sect. 5 gives an overview of previous existing approaches and suggestions related to the topic of social-aware applications. Finally, Sect. 6 summarizes our proposed approach.

Conceptual Problem Analysis 2

This paper concentrates on adaptive ranking in keyword-based search applications. Our objective is to consider data from remote social networks as the basis for sophisticated appropriate ranking metrics. The proposed social-aware search (SAS) concept will not only rely on personal attributes characterizing the current user but also on the relationships in its friend network and the search actions and social activities of the user's acquaintances.

Firstly, we provide an example scenario to illustrate the challenges of social-awareness: We assume the existence of a web application WA that allows the search in a product stock ASSETSTOCK. This collection can contain arbitrary items, such as product data from a web shop or even bibliographic data from a digital library. ASSETSTOCK should be referenceable by semantic means in an RDF representation, so that each item can be addressed by a unique uniform resource identifier (URI). A user u invokes this web application and authorizes the access to one of its centralized or decentralized existing social profiles by performing an identification process. Depending on the remote Social Web platform SWP, the web application is then capable of accessing the friend network information of the current user with a set of friends F and retrieving the corresponding profile data. The data does not have to be in a Linked Data format in the retrieval step, but should be convertible into an RDF representation with a semantic concept mapping. This data is then accessed through a separate meta-data store SASSTORE. Friend entities that are already known by the application can be derived, mapped and taken into consideration for the adaptive result ranking. Since this is a one-time activity,

Based on this social data, the application can then already access information on items from the ASSETSTOCK which were searched by connected friends F in the past. After the current user u entered one or multiple keywords, the social-aware web application can retrieve all corresponding items and adaptively reorder it according to the relevance in the circle of friends of u based on an underlying scoring model SM.

There are two key requirements for this scenario to work in practice. First, the user has to social sign-in using the web application beforehand. This step is not trivial, because a user is likely to enter and search anonymously on a website, and only signs in after the article of interest was already found. As this is a user interface interaction challenge, we will not focus on it here. Furthermore, not every current Social Web platform provides all relevant social data to external applications due to privacy restrictions or scope limitations.

3 Towards a Linked Data enabled Social Scoring Model

We present a scoring model that allows the calculation of the relevance of a given (URI-addressable) item resource to the current user based on interaction data that were stored in the past, and social relationship data from remote data endpoints. It is extensible with other ranking metrics and includes features such as data aging, ranking factor weighting, and confidence/trust values, according to the current use case requirements of the application provider.

3.1 The Social-Aware Search Scoring Model

To express recent activities of acquaintances from the users' social network we model interest in the form of clicks C on specific items that are displayed in a search result list. One specific click c can be characterized by a set of meta attributes $c = (a, r, t, s) \in C$, which are at least:

- the agent a who performed the click action
- the resource r of the clicked search item with an RDF subject IRI
- the time t of the click activity
- the actual search context *s* within the web application, primarily given by the search keywords the agent entered

To consider all relevant characteristics from social profiles, we use the formalization depicted in equation 1 to calculate a RelevanceScore RS.

$$RS(u, r, t', s') = \frac{1}{|SC|} \sum_{crit \in SC} \omega_{crit} \cdot \text{MmtFn}_{crit}(u, r, t', s')$$
 (1)

with u = current user, r = search item URL, t' = search time, s' = current search context, $crit = \text{a social criterion} \in SC$ as a set of social criteria of interest, $\omega_{crit} = \text{weight of } crit$, MmtFn = a measurement function of crit

The formula calculates a continuous, normalized score value between 0.0 and 1.0 as a linear combination of all social criteria of interest, representing the relevance of the current search item to the user and the current search context. Each search item r, the user u and its acquaintances as well as the corresponding metadata properties can be directly described and

referenced by Linked Data resources. The measurement functions are individually modeled for each criterion, thus may have different dimensions. The result of the measurement function of each social criterion can be calculated independently and returns a value between 0.0 and 1.0 stating the partial relevance of this specific criterion. The influence of each criterion can be weighted according to the specific search application domain using ω_{crit} . Equation 2 instantiates the general formulation of the scoring model for an example use case, where the user u has relevant properties like e.g. gender, age and its friend network.

$$RS(u, r, t', s') = 1/4 \cdot \\ \omega_{GenderEquality} \cdot \text{MmtFn}_{GenderEquality}(u.Gender, r, s') + \\ \omega_{AgeReq} \cdot \text{MmtFn}_{AgeReq}(u.Age, r, s') + \\ \omega_{OwnInterestHist} \cdot \text{MmtFn}_{OwnInterestHist}(u.OwnInterestHist, r, t', s') + \\ \omega_{FriendInterest} \cdot \text{MmtFn}_{FriendInterest}(u.FriendNetwork, r, t', s')$$

As an example, the *FriendInterest* metric *FI* can be implemented as shown in equation 3 and interprets clicks on a search result as a signal for interest.

$$\operatorname{MmtFn}_{FI}(u.FN, r, t', s') = \frac{1}{|C|} \cdot \sum_{f: (finu.FN), (f, r, t, s) \in C} \frac{\operatorname{strength}(u, f)}{\operatorname{distance}(u, f)} \cdot \operatorname{aging}(t, t')$$
(3)

with s == s' and t < t', where FI = FriendInterest, FN = FriendNetwork, f = a friend of the user's friend network, strength = a network cluster function to describe the trust relationship between u and f between 0.0 and 1.0, distance = the hops in the friend network between u and f (for direct contacts equal to 1), aging = an aging modification function

We base our work on the assumption that there is an interest influence probability between related users which was also shown by other authors like [AKM08] or [GBL10] as statistically significant (cf. Sect. 5) and that certain search actions of a user's friend circle in the past are also of personal interest for him/herself in the present. We included a network distance concept between the current user and a friend, as well as a trust/confidence strength factor in the formalization (is it a good friend in the same knowledge area, or a loosely known acquaintance). The particular relevance calculation can then be based on the ratio of reoccurring clicks from the user's social network to the total number of clicks. We also identified that the age of the click action is of particular interest for its relevance, so the equation also contains an additional term with an aging function (clicks within the last hours might be of higher relevance than performed actions of a friend one year ago).

3.2 Key Benefits from using Linked Data

The presented scoring model for a relevance-based resource ordering based on the linear combination of multiple metrics of interest could be implemented built upon conventional data representations. But a Linked Data approach is more appropriate that it directly fits to the representation of search items as resources that are consumed by specific, explicitly describable people, both representable by a URL.

The Linked Data approach leads to several benefits:

- Uniform representation: The concrete application domain of the searchable items
 as well as the particular user's social network is irrelevant for the Scoring Model
 in the Social Commerce application because both are represented through uniform
 concepts. Moreover, we can uniformly define ranking metrics on all available data,
 optionally by using well-known query languages such as SPARQL.
- *Mapping of users:* By representing users from a particular social network through a specific URL, it is not only possible to retrieve their direct contacts but also relationships over multiple persons if several of them use the same social-aware web application. By using background knowledge, it is even possible to map users and friend interactions between multiple social networks by means of *owl:sameAs*.
- Deduction and inference operations: If additional properties are available describing a search item resource or the user from a social network in a Linked Data fashion, we can also use these properties to rank search items based on criteria that are not explicitly stated (such as spatial constraints, but also social concepts such as Likes on a Fan Page of a company or institution producing a particular product).

4 Prototype

To evaluate the integration of social attributes and social network friend relationships into a conventional search experience, we implemented a first social-aware search demonstrator. The underlying data model was separated into two databases: an *ASSETSTOCK* to store information about specific products provided by the web application, and an *SASSTORE* to represent the metadata from the Social Web describing specific people attributes and interactions with the search interface of the web applications. Both data stores are graph databases containing structured Linked Data sets.

The entire system architecture of the prototype is depicted in Fig. 1. An external authentication middleware module¹⁰ is used to authenticate the current user and access its social profile (allowing centralized social networks such as *Facebook* as well as decentralized approaches such as *WebID*).

¹⁰ http://passportjs.org/

After entering some keywords in a text field of a *WebUI*, the keywords are first *processed* and then used to define a search context by a *contextualization* module. A *result retrieval* module retrieves corresponding potential result items from the *ASSETSTOCK* data base. A *result processing* module then applies the approach presented in Sect. 3 and uses the available metadata from the *SASSTORE* to adaptively rank the available results based on the users' social data and its friends recent interaction data. In particular, it uses the resource URL from the RDF subject of a result item as an identifier to retrieve all the clicks of other known users that were performed in the same context on these result items in the past.

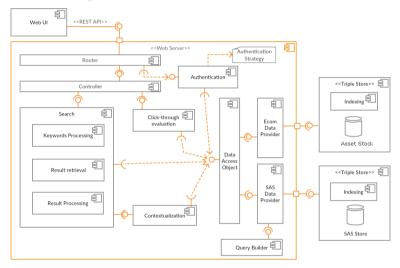


Fig. 1: Initial prototype architecture

We use bibliographic data of recent publications in the *ASSETSTOCK* as a basic data set example, that is primarily based on the schema.org vocabulary. Clicks on the displayed search results are written back in a semantic fashion to the *SASSTORE*, especially the resource URL of the selected item, a time and the keyword context the user entered (see List. 1).

```
sas:searchRecord1484823549222 rdfs:type schema:searchAction; schema:agent <a href="http://example.org/sas/users/user190">http://example.org/sas/users/user190</a>; schema:startTime "1484823549222"; schema:query "Semantic Web"; schema:result <a href="http://example.org/keseda/publication127">http://example.org/keseda/publication127</a>.
```

List. 1: Exemplary search activity for a user who clicked on a result item after searching for 'Semantic Web'

Basically, the SAS demonstrator can be used anonymously without a social login as a conventional search interface by ranking the search results according to their relevance based on a click-through rate (CTR). As depicted in Fig. 2, the first two items in the result list were selected twice in total in the past and are then ranked alphabet-

ically, whereas the third item was only clicked once by users performing a search for publications dealing with 'Semantic Web' in the past.

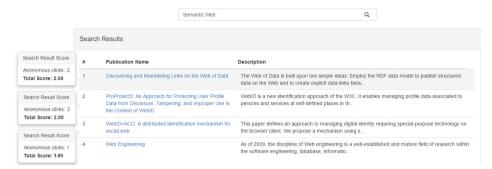


Fig. 2: Default result ranking without social information

If a user now alternatively social signs-in, the system will then retrieve the exposed social user data from a Social Network and store it in the SASSTORE graph store. For evaluation purposes, we used the Facebook OpenGraph API to retrieve basic personal attributes such as gender, age and location of the current user as well as its direct friend relationships to other Facebook users. The exposed data was converted into a Linked Data representation, primarily by using the FOAF ontology. However, in order to represent all characteristics of social network relationships, FOAF turned out to be not sufficient. Therefore, an extension with the Social Relationships Ontology (SORON) was necessary (see List. 2).

List. 2: Exemplary social meta-data for user190

After performing the search action, the retrieved result item RDF triples are now further processed. For each result item, the set of agents (particular the agent resource URLs who clicked on this item in the past from the *SASSTORE*) are compared with the social-network acquaintances of the current user. If other acquaintances of the user have already searched for these items within the same web application, a set of relevant characteristics is used such as the distance, relationship strength and click time to calculate a global relevance score value for each result item of the executed search. The search items are then listed accordingly to their relevance score.

An example GUI of the demonstrator with enabled social-aware search functionality is illustrated in Fig. 3. For demonstration purposes, a graph visualization component¹¹ was added to display the influences of the current social friend network on the result ranking decision.

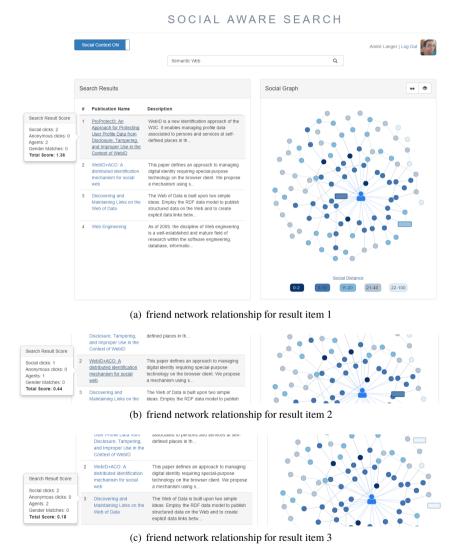


Fig. 3: Adaptive ranking of search results in our social-aware search prototype

The scenario uses the same data as already discussed in the conventional use case in Fig. 2, but this time the adaptive ranking can benefit from the knowledge of existing social

¹¹ http://visjs.org/docs/network/index.html

relationships to other known users. The presented search result item order is different from the anonymous use case. The secondly ranked result item from the anonymous usage scenario is now the top-ranked result item in Fig. 3(a), because the item was recently also relevant for direct acquaintances of the current user. A different item is also placed on the second position of the result list; interestingly with less absolute conversion clicks than the third-ranked item. Obviously, because its clicks are classified as higher relevant because they derive from an acquaintance in the direct contact circle whereas the clicks of the third item origin from indirect contacts with a higher relationship distance.

The quality of the adaptive social-aware result ranking is of course dependent from the number of known users who already performed similar interactions in the past, and know each other. Nevertheless, the demonstrator results confirm the assumptions that users in a social network tend to influence each other and that interactions of a particular user from the past can be used as relevant meta-data for a social aware adaptive search result ranking of a related user in the present.

Additionally, the demonstrator shows that the Linked Data approach easily allows obtaining advanced social relationship knowledge by using shattered friend network data of multiple users. Whereas the *Facebook OpenGraph API* permits only to retrieve contact information of direct friends, the web application can subsume the information of multiple friends to build a multi-level social friend network representation. This contact and social relationship information does not necessarily have to be from the same social network; it can also be derived from internal background knowledge or even by synthetically generated.

5 Related Work

Adaptive result ranking is of major interest for data scientists as well as for industrial purposes. It is not only relevant for traditional search engine providers, but also for application providers with sophisticated search interfaces in general. Traditional ranking factors and metrics are frequently analyzed and published, e.g. in [Se16]. Newer approaches also focus on social attributes that can be used for a personalized result ranking. Authors analyzed influence probabilities among friend networks within the Social Web, e.g. [AKM08] and showed a qualitative indication of the existence of influence and correlation in social networks, whereas [GBL10] focused on influence probabilities. [Wu14] provided examples for a top-k search approach, which is extended by others to a personalized, network-aware search [Ca09, MC13]; common scenarios are a news search or image tagging. [Kö16] emphasized the trust aspect and generation through social recommendations, especially in mobile commerce. However, none of these approaches used Linked Data structures for social-information filtering in the past. Our approach uses RDF structures for representing item resources as well as social interaction meta information. We used the results of these recent contributions and built a scoring model that is capable of taking interaction influence between related users from different social networks into account.

The Linked Data approach to semantically annotate, access, combine and interpret information as RDF triples in a unique way is already well understood [BHB09, WT13]. It is also used in various ways. However, concepts combining social web data with Linked Data to uniformly access and process the information from social networks for industrial purposes are to the best of our knowledge still rare; e. g. [PV13] use Linked Open Data to improve recommendations in the e-commerce domain. [GNP05] used a similar idea to define and exchange social context information through FOAF graphs but they applied a PageRank formula with a random surfer model to calculate a ranking score whereas our scoring model allows the linear combination of arbitrary metrics on concrete result resource items.

6 Conclusion and Future Work

In this paper, we described the idea of a social-aware search (SAS) that takes multiple social attributes and friend network relationships into consideration to achieve an adaptive result ranking. In contrast to several other existing suggestions for ranking models based on data from the Social Web, our approach is built upon Linked Data principles and allows the interoperability and combination of user behavior data from different social networks without relying on application-specific user profiles. A SAS enabled Web application can access user data from Social Web platforms and use this information for improving search relevance and experience.

We particularly focused on friend-network relationships and adapted the idea of collaborative filtering. Therefore, we extended the traditional way to retrieve Social Web data by means of Linked Data, so that operations such as inference can be performed based on specific ontologies that operate on RDF data.

After presenting a first working prototype, the next step will be to conduct user studies to evaluate the advantages of a SAS and how users respond to it. Additionally, we want to particularly use the inference capabilities of the Linked Data approach to derive advanced ranking factors and measurements based on ontological relationships between involved social web platform entities. This will include the extension of our scoring model with factors like shares, comments and likes.

Acknowledgment. This work was supported by the grant from the German Federal Ministry of Education and Research (BMBF) for the LEDS Project under grant agreement No 03WKCG11D.

References

[AKM08] Anagnostopoulos, Aris; Kumar, Ravi; Mahdian, Mohammad: Influence and Correlation in Social Networks. In: Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. KDD '08, ACM, New York, NY, USA, pp. 7–15, 2008. http://doi.acm.org/10.1145/1401890.1401897.

- [BHB09] Bizer, Christian; Heath, Tom; Berners-Lee, Tim: Linked Data The Story So Far. International Journal on Semantic Web and Information Systems, 5(3):1–22, 2009. http://dx.doi.org/10.4018/iswis.2009081901.
- [Ca09] Carmel, David et al.: Personalized Social Search Based on the User's Social Network. In: Proceedings of the 18th ACM Conference on Information and Knowledge Management. CIKM '09, ACM, New York, NY, USA, pp. 1227–1236, 2009. http://doi.acm.org/10.1145/1645953.1646109.
- [GBL10] Goyal, Amit; Bonchi, Francesco; Lakshmanan, Laks V.S.: Learning Influence Probabilities in Social Networks. In: Proceedings of the Third ACM International Conference on Web Search and Data Mining. WSDM '10, ACM, New York, NY, USA, pp. 241–250, 2010. http://doi.acm.org/10.1145/1718487.1718518.
- [GNP05] Ghita, Stefania; Nejdl, Wolfgang; Paiu, Raluca: Semantically rich recommendations in social networks for sharing, exchanging and ranking semantic context. Lecture Notes in Computer Science, 3729 LNCS:293–307, 2005.
- [Kö16] Köster, Antonia; Matt, Christian; Torremante, Larissa; Hess, Thomas; Heinze, Jörg: Vertrauensbildung durch soziale Empfehlungen: Zur Rolle von Empfehlungssystemen im M-Commerce. Informatik-Spektrum, 2016.
- [Ma16] Mansour, Essam; Sambra, Andrei Vlad; Hawke, Sandro; Zereba, Maged; Capadisli, Sarven; Ghanem, Abdurrahman; Aboulnaga, Ashraf; Berners-Lee, Tim: A Demonstration of the Solid Platform for Social Web Applications. Proceedings of the 25th International Conference Companion on World Wide Web, pp. 223–226, 2016.
- [MC13] Maniu, Silviu; Cautis, Bogdan: Network-aware Search in Social Tagging Applications: Instance Optimality Versus Efficiency. In: Proceedings of the 22Nd ACM International Conference on Information & Knowledge Management. CIKM '13, ACM, New York, NY, USA, pp. 939–948, 2013. http://doi.acm.org/10.1145/2505515.2505760.
- [PV13] Peska, Ladislav; Vojtas, Peter: Using Linked Open Data to Improve Recommending on E-Commerce. In: Proceedings of the 2nd International Workshop on Recommender Systems meet Big Data & Semantic Technologies. 2013.
- [Se16] Searchmetrics: Rebooting Ranking-Faktoren. Technical report, 2016. http://www.searchmetrics.com/de/knowledge-base/ranking-faktoren, Accessed 2017-01-17.
- [St16] Statista: , Statistics and facts about social media usage, 2016. https://www.statista.com/topics/1164/social-networks/, Accessed 2017-01-16.
- [WT13] Weaver, Jesse; Tarjan, Paul: Facebook Linked Data via the Graph API. Semantic Web, 4(3):245–250, July 2013. http://dl.acm.org/citation.cfm?id=2786071.2786074.
- [Wu14] Wu, Dingming; Li, Yafei; Choi, Byron; Xu, Jianliang: Social-Aware Top-k Spatial Keyword Search. In: 2014 IEEE 15th International Conference on Mobile Data Management. volume 1, pp. 235–244, 2014.