# Ranking Given Names
## Algorithms and Evaluation Paradigms

Folke Mitzlaff

Knowledge and Data Engineering Group (KDE)
University of Kassel
Wilhelmshöher Allee 73, D-34121 Kassel, Germany
Email: mitzlaff@cs.uni-kassel.de

Gerd Stumme

Knowledge and Data Engineering Group (KDE)
University of Kassel
Wilhelmshöher Allee 73, D-34121 Kassel, Germany
Email: stumme@cs.uni-kassel.de

*Abstract*— As a result of the author's need for help in finding a given name for the unborn baby, the Nameling[1], a search engine for given names, based on data from the "Social Web" was born. Within less than six months, more than 35,000 users accessed Nameling with more than 300,000 search requests, underpinning the relevance of the underlying research questions.

The present work compares different metrics for calculating similarities among given names, based on co-occurrences within Wikipedia. In particular, the task of finding relevant names for a given search query is considered as a ranking task and the performance of different statical measures of relatedness among given names are evaluated with respect to Nameling's actual usage data. By publishing the considered usage data, the research community is stipulated for developing advanced recommendation systems and analyzing influencing factors for the choice of a given name.

## I. INTRODUCTION

Whoever was in the need for a given name knows how difficult it is to choose a suitable name which meets (in the best case) all constraints. The choice of a given name is not only influenced by personal taste. Many influencing factors have to be considered, such as cultural background, language dependent pronunciation, current trends and, last but not least, public perception or even prejudices towards given names.

Though there is a constant demand for finding a suitable name, little aid exists, beside alphabetically ordered lists of names and simple filtering techniques. From a scientist's perspective, the task of ranking and recommending given names is challenging and can be tackled from many different disciplines, such as social network analysis and data mining, in our case.

In the present work, we consider the task, where a user searches for suitable names based on a given set of query names. The user thus can, e. g., search for names which are related to his own name, the names of all family members together or just for names which are similar to a name he or she likes.

The proposed approach for assessing similarity among given names is based on co-occurrences within Wikipedia,[2] but can as well be applied to other data sources, such as micro blogging systems, news paper archives or collections of eBooks. Manual inspection of a first implementation of cosine similarity (cf. Section III) already showed promising results, but as there is a variety of metrics for calculating similarity in such co-occurrence graphs, the question arises, which one gives raise to the most "meaningfull" notion of relatedness.

For assessing the performance of the different similarity metrics, we firstly compare the obtained rankings with an external reference ranking. Secondly, we evaluate the ranking performance with respect to actual usage data from Nameling, which comprises more than 35,000 users with more than 300,000 search queries during the time period of evaluation. To promote other researchers' efforts in developing new ranking and recommendation systems, all considered data is made publicly available.[3]

The rest of the work is structured as follows: Section II presents related work. Section III summarizes basic notions and concepts. Section IV briefly describes Nameling and Section V presents results on the semantics of the similarity metrics under consideration. In Section VI, the actual usage data of Nameling is introduced and analyzed, which is then used in Section VII for comparing the performance of different similarity metrics. Finally, in Section VIII, the present work's contributions are summarized and forthcoming work is presented.

## II. RELATED WORK

The present work tackles the new problem of discovering and assessing relatedness of given names based on data from the social web for building search and recommendation systems which aid (not only) future parents in finding and choosing a suitable name.

Major parts of the underlying research questions are closely related to work on *link prediction* in the context of social networks as well as *distributional similarity*, where, more generally, semantic relations among named entities are investigated. However, this work focuses on the evaluation of similarity metrics relative to actual user preferences as expressed by interactions with names within a live system. This is related to the evaluation of *recommender systems* which exist for many application contexts, such as movies [7], tags [11] and products [18].

---

[1]http://nameling.net
[2]http://www.wikipedia.org
[3]http://www.kde.cs.uni-kassel.de/nameling/dumps

*a) Distributional Similarity & Semantic Relatedness:*
The field of distributional similarity and semantic relatedness has attracted a lot of attention in literature during the past decades (see [4] for a review). Several statistical measures for assessing the similarity of words are proposed, as for example in [3], [8], [10], [15], [25]. Notably, first approaches for using Wikipedia as a source for discovering relatedness of concepts can be found in [2], [23], [6].

*b) Vertex Similarity & Link Prediction:* In the context of social networks, the task of predicting (future) links is especially relevant for online social networks, where social interaction is significantly stimulated by suggesting people as contacts which the user might know. From a methodological point of view, most approaches build on different similarity metrics on pairs of nodes within weighted or unweighted graphs [12], [16], [19], [20]. A good comparative evaluation of different similarity metrics is presented in [17].

Nevertheless, usage data of systems such as Nameling is, to the best of our knowledge, new and was not available before. The present work combines approaches from the link prediction and recommendation tasks with a focus on the performance of structural similarity metrics based on co-occurrence networks obtained from Wikipedia.

## III. PRELIMINARIES

In this chapter, we want to familiarize the reader with the basic concepts and notations used throughout this paper.

### A. Graph & Network Basics

A *graph* $G = (V, E)$ is an ordered pair, consisting of a finite set $V$ of *vertices* or *nodes*, and a set $E$ of *edges*, which are two-element subsets of $V$. A *directed graph* is defined accordingly: $E$ denotes a subset of $V \times V$. For simplicity, we write $(u, v) \in E$ in both cases for an edge belonging to $E$ and freely use the term *network* as synonym for a graph. In a *weighted graph*, each edge $l \in E$ is given an edge weight $w(l)$ by some weighting function $w \colon E \to \mathbb{R}$. The *density* of a graph denotes the fraction of realized links, i. e., $\frac{2m}{n(n-1)}$ for undirected graphs and $\frac{m}{n(n-1)}$ for directed graphs (excluding self loops). The *neighborhood* $\Gamma$ of a node $u \in V$ is the set $\{v \in V \mid (u, v) \in E\}$ of *adjacent* nodes. The *degree* of a node in a network measures the number of connections it has to other nodes. For the *adjacency matrix* $A \in \mathbb{R}^{n \times n}$ with $n = |V|$ holds $A_{ij} = 1$ ($A_{ij} = w(i, j)$) iff $(i, j) \in E$ for any nodes $i, j$ in $V$ (assuming some bijective mapping from $1, \ldots, n$ to $V$). We represent a graph by its according adjacency matrix where appropriate.

A *path* $v_0 \to_G v_n$ of *length* $n$ in a graph $G$ is a sequence $v_0, \ldots, v_n$ of nodes with $n \geq 0$ and $(v_i, v_{i+1}) \in E$ for $i = 0, \ldots, n-1$. A *shortest path* between nodes $u$ and $v$ is a path $u \to_G v$ of minimal length. The *transitive closure* of a graph $G = (V, E)$ is given by $G^* = (V, E^*)$ with $(u, v) \in E^*$ iff there exists a path $u \to_G v$. A *strongly connected component (scc)* of $G$ is a subset $U \subseteq V$, such that $u \to_{G^*} v$ exists for every $u, v \in U$. A *(weakly) connected component (wcc)* is defined accordingly, ignoring the direction of edges $(u, v) \in E$.

### B. Vertex Similarities

Similarity scores for pairs of vertices based only on the surrounding network structure have a broad range of applications, especially for the link prediction task [17]. In the following we present all considered similarity functions, following the presentation given in [5] which builds on the extensions of standard similarity functions for weighted networks from [22].

The *Jaccard coefficient* measures the fraction of common neighbors:

$$JAC(x, y) := \frac{|\Gamma(x) \cap \Gamma(y)|}{|\Gamma(x) \cup \Gamma(y)|}$$

The Jaccard coefficient is broadly applicable and commonly used for various data mining tasks. For weighted networks the Jaccard coefficient becomes:

$$\widetilde{JAC}(x, y) := \sum_{z \in \Gamma(x) \cap \Gamma(y)} \frac{w(x, z) + w(y, z)}{\sum_{a \in \Gamma(x)} w(a, x) + \sum_{b \in \Gamma(y)} w(b, y)}$$

The *cosine similarity* measures the cosine of the angle between the corresponding rows of the adjacency matrix, which for a unweighted graph can be expressed as

$$COS(x, y) := \frac{|\Gamma(x) \cap \Gamma(y)|}{\sqrt{|\Gamma(x)|} \cdot \sqrt{|\Gamma(y)|}},$$

and for a weighted graph is given by

$$\widetilde{COS}(x, y) := \sum_{z \in \Gamma(x) \cap \Gamma(y)} \frac{w(x, z) w(y, z)}{\sqrt{\sum_{a \in \Gamma(x)} w(x, a)^2} \cdot \sqrt{\sum_{b \in \Gamma(y)} w(y, b)^2}}.$$

The *preferential PageRank* similarity is based on the well known PageRank $^{TM}$[1] algorithm. For a column stochastic adjacency matrix $A$ and damping factor $\alpha$, the *global* PageRank vector $\vec{w}$ with uniform *preference vector* $\vec{p}$ is given as the fixpoint of the following equation:

$$\vec{w} = \alpha A \vec{w} + (1 - \alpha) \vec{p}$$

In case of the *preferential PageRank* for a given node $i$, only the corresponding component of the preference vector is set. For vertices $x, y$ we set accordingly

$$PPR(x, y) := \vec{w}_{(x)}[y],$$

that is, we compute the preferential PageRank vector $\vec{w}_{(x)}$ for node $x$ and take its $y$'th component. We also calculate an adapted preferential PageRank score by adopting the idea presented in [9], where the global PageRank score $PR$ is subtracted from the preferential PageRank score in order to reduce frequency effects and set

$$PPR+(x, y) := PPR(x, y) - PR(x, y).$$

### C. Evaluation Metrics

Several metrics for assessing the perfomance of recommendation systems exists. We apply the *mean average precision* for obtaining a single value performance score for a set $Q$ of ranked predicted recommendations $P_i$ with relevant documents $R_i$:

$$\mathrm{MAP}(Q) := \frac{1}{|Q|} \sum_{i=1}^{|Q|} \mathrm{AveP}(P_i, R_i)$$

where the *average precision* is given by

$$\text{AveP}(P_i, R_i) := \frac{1}{|R_i|} \sum_{k=1}^{|R_i|} [\text{Prec}(P_i, k) \cdot \delta(P_i(k), R_i)]$$

and $\text{Prec}(P_i, k)$ is the *precision* for all predicted elements up to rank $k$ and $\delta(P_i(k), R_i) = 1$, iff the predicted element at rank $k$ is a relevant document ($P_i(k) \in R_i$). Refer to [26] for more details.

## IV. NAMELING– A SEARCH ENGINE FOR GIVEN NAMES

Nameling is designed as a search engine and recommendation system for given names. The basic principle is simple: The user enters a given name and gets a browsable list of "relevant" names, called "*namelings*". Figure 1a exemplarily shows namelings for the classical masculine German given name "Oskar".

The list of namelings in this example ("Rudolf", "Hermann", "Egon", …) exclusively contains classical German masculine given names as well. Whenever an according article in Wikipedia exists, categories for the respective given name are displayed, as, e. g., "*Masculine given names*" and "*Place names*" for the given name "Egon". Via hyperlinks, the user can browse for namelings of each listed name or get a list of all names linked to a certain category in Wikipedia. Further background information for the query name is summarized in a corresponding details view, where, among others, popularity of the name in different language editions of Wikipedia as well as in Twitter is shown. As depicted in Fig. 1b, the user may also explore the "neighborhood" of a given name, i. e., names which co-occur often with the query name.

From a user's perspective, the Nameling is a tool for finding a suitable given name. Accordingly, names can easily be added to a personal list of favorite names. The list of favorite names is shown on every page in the Nameling and can be shared with a friend, for collaboratively finding a given name.

The Nameling is based on a comprehensive list of given names, which was initially manually collected, but then populated by user suggestions. It currently covers more then 35,000 names from a broad range of cultural contexts. For different use cases, three different data sources are respectively used, as depicted in Fig. 2.
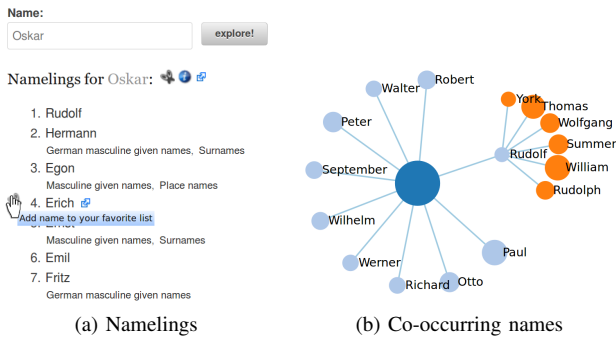


(a) Namelings  (b) Co-occurring names

Fig. 1: The user queries for the classical German given name "Oskar".
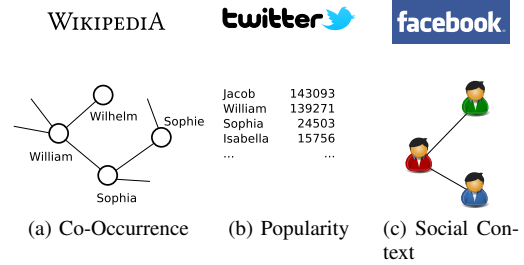


(a) Co-Occurrence  (b) Popularity  (c) Social Context

Fig. 2: Nameling determines similarities among given names based on co-occurrence networks from Wikipedia, popularity of given names via Twitter and social context of the querying user via facebook.

*a) Wikipedia:* As basis for discovering relations among given names, a co-occurrence graph is generated for each language edition of Wikipedia separately. That is, for each language, a corresponding data set is downloaded from the Wikimedia Foundation. Afterwards, for any pair of given names, the number of sentences where they jointly occur is determined. Thus, for every language, an undirected graph is obtained, where two names are adjacent, if they occur together in at least one sentence within any of the articles and the edge's weight is given by the number of such sentences.

Relations among given names are established by calculating a vertex similarity score between the corresponding nodes in the co-occurrence graph. Currently, namelings are calculated based on cosine similarity (cf. Section III).

*b) Twitter:* For assessing up-to-date popularity of given names, a random sample of tweets in Twitter is constantly processed via the Twitter streaming api[4]. For each name, the number of tweets mentioning it is counted.

*c) facebook:* Optionally a user may connect the Nameling with facebook[5]. If the user allows the Nameling to access his or her profile information, the given names of all contacts in facebook are collected anonymously. Thus, a "social context" for the user's given name is recorded. Currently, the social context graph is too small for implementing features based on it, but it will be a valuable source for discovering and evaluating relations among given names.

## V. SEMANTICS OF SOCIAL CO-OCCURRENCES

The basic idea behind the Nameling was to find relations among given names based on user-generated content in the social web. The most basic relation among such entities can be observed when they occur together within a given atomic context. In case of Wikipedia, such *co-occurrences* were counted based on sentences.

Considering the German and English Wikipedia separately, undirected weighted graphs $\text{Wiki}^{\text{DE}}$ and $\text{Wiki}^{\text{EN}}$ are obtained, where name nodes $u$ and $v$ are connected and labeled with weight $c$, if $u$ and $v$ co-occurred in exactly $c$ sentences. For

---

[4]https://dev.twitter.com/docs/api/1/get/statuses/sample
[5]http://www.facebook.com

example, the given names "*Peter*" and "*Paul*" co-occurred in 30,565 sentences within the English Wikipedia. Accordingly, there is an edge (Peter, Paul) in Wiki$^{EN}$ with a corresponding edge weight. For the present analysis, the co-occurrence networks are derived from the official Wikipedia data dumps, which are freely available for download,[6] whereby the English dump was dated *2012-01-05* and the German dump *2011-12-12*. The following experiments focus on the English Wikipedia unless explicitly stated otherwise.

When Nameling was implemented, the choice of the applied similarity metric on those co-occurrence networks was inspired by previous work on emergent semantics of social tagging systems [21], but only based on manual inspection of a (non representative) sample. This section aims at grounding the obtained notions of relatedness among given names by comparing the different similarity metrics with an external reference similarity.

We built such a reference relation on the set of given names from the on-line dictionary Wiktionary[7] by extracting all category assignments from pages corresponding to given names. Thus, for each of 10,938 given names a respective binary vector was built, where each component indicates whether the corresponding category was assigned to it (in total 7,923 different categories and 80,726 non-zero entries). These vectors were then used for assessing relatedness among given names, based on a categorization which was explicitly established by the authors of the corresponding entries in Wiktionary.

In detail: For any pair $(u, v)$ of names in the co-occurrence network which have a category assignment, we calculated the cosine similarity $COS(u, v)$ based on the respective category assignment vectors as well as any of the similarity metrics $s(u, v)$ on the co-occurrence graph as described in section III. As the number of data points $(COS(u, v), s(u, v))$ grows quadratically with the number of names, we grouped the co-occurrence based similarity scores in 1,000 equidistant bins (in case of $COS$ and JC) or logarithmic bins (in case of $PPR$ and $PPR+$). For each bin, the average cosine similarity based on category assignments was calculated, as shown in Figure 3. Notably, all considered similarity metrics capture a positive correlation between similarity in the co-occurrence network and similarity between category assignments to names. But significant differences between the applied similarity functions can be observed. The weighted cosine similarity performs very well, firstly in showing a steep slope and secondly in exhibiting a stable monotonous curve progression. The unweighted Jaccard coefficient shows an even more pronounced linear progression, but is less stable for higher similarity scores whereas the weighted Jaccard coefficient shows a higher correlation with the reference similarity for high similarity scores. It is worth noting, that the cosine similarity is only marginally effected by the edge weights of the co-occurrence graph, whereas the weighted Jaccard coefficient significantly

---

[6]http://dumps.wikimedia.org/backup-index.html
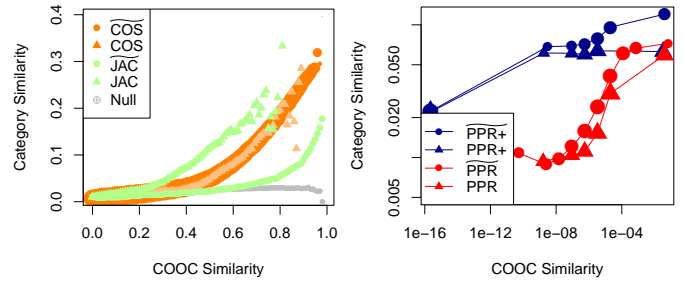[7]http://www.wiktionary.org



Fig. 3: Similarity based on name categories in Wiktionary vs. vertex similarity in the co-occurrence-networks (weighted and unweighted) whereby the null model was obtained by shuffling the mapping of names to categories.

differs from its unweighted variant. In case of the PageRank based similarity metrics, the weighted variants consistently outperform the corresponding unweighted variants and the adapted preference PageRank function *PPR+* outperforms the plain preference PageRank.

We conclude that all considered similarity metrics on the co-occurrence networks obtained from Wikipedia capture a notion of relatedness which correlates to an external semantically motivated notion of relatedness among given names.

## VI. USAGE DATA

The results presented in the previous section indicate correlations between semantic relatedness among given names and structural similarity within co-occurrence networks obtained from Wikipedia.

This section aims at assessing the performance of the different structural similarity metrics with respect to actual interactions of users within Nameling. For this purpose, we considered the Nameling's activity log entries within the time range *2012-03-06* until *2012-08-10*. In the following, we firstly describe the collected usage data, analyze properties of emerging network structures among names and users and finally compare interrelations between the different networks.

In total, 38,404 users issued 342,979 search requests. Subsequently, we differentiate between the following activities:

- *"Enter":* A user manually entered a given name into search mask.
- *"Click":* A user followed a link to a name within a result list.
- *"Favorite":* A user added a given name to his/her list of. favorite names
- *"Nameling":* All search requests together

Table I summarizes high level statistics for these activity classes, showing, e.g., that $35,684$ users entered $16,498$ different given names. For analyzing how different users contribute to the Nameling's activities, Figure 4 shows the distribution of activities over the set of users, separately for *Enter*, *Click* and *Favorite* requests. Clearly, all activities' distributions exhibit long tailed distributions, that is, most users entered less than 20 names but there are also users with more than 200 requests.
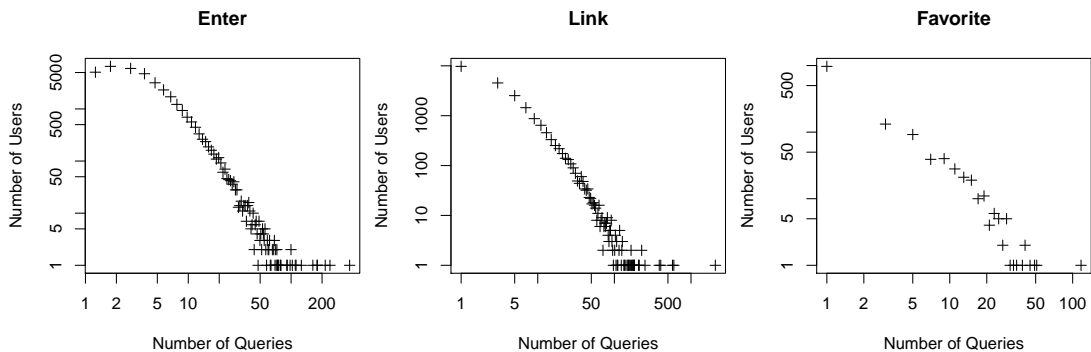
Fig. 4: Number of users per query count for *Enter*, *Click* and *Favorite* activities.

TABLE II: Most popular names separately for *Enter*,*Click* and *Favorite* activities, as well as global popularity and popularity in Wikipedia.

| Enter | | Click | | Favorite | | Nameling (global) | | Wikipedia (EN) | | Wikipedia (DE) | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Emma | 1,433 | Emma | 1,461 | Emma | 62 | Emma | 3,073 | John | 2,474,408 | Friedrich | 356,636 |
| Anna | 1,125 | Jonas | 1,010 | Lina | 49 | Anna | 2,178 | David | 1,130,766 | Karl | 348,879 |
| Paul | 1,073 | Emil | 965 | Ida | 40 | Paul | 1,918 | William | 1,100,253 | Hans | 309,734 |
| Julia | 942 | Anna | 947 | Jakob | 39 | Jonas | 1,856 | James | 1,051,219 | Peter | 304,516 |
| Greta | 895 | Alexander | 851 | Felix | 37 | Emil | 1,816 | George | 973,210 | Johann | 298,575 |
| Michael | 878 | Daniel | 819 | Oskar | 36 | Michael | 1,771 | Robert | 871,368 | John | 295,563 |

TABLE I: Basic statistics for the different activities within the Nameling.

| | #Users | #Names |
|---|---|---|
| *Enter* | 35,684 | 16,498 |
| *Click* | 22,339 | 10,028 |
| *Favorite* | 1,396 | 1,558 |

To get a glimpse at the actual usage patterns within the Nameling and the distribution of names within Wikipedia, Table II exemplarily shows most popular names for the different activity classes as well as the considered data sets derived from Wikipedia. Thereby we assess a name's "popularity" by its frequency within the corresponding Wikipedia dump or the number of search queries for the name within Nameling. Firstly, it is worth noting that indeed in the German Wikipedia German given names are dominant, whereas in the English Wikipedia, accordingly, English given names are the most popular. Secondly, both language editions of Wikipedia are dominated by male given names. As for the Nameling, users (still) mostly originate in Germany and accordingly the corresponding query logs are dominated by search requests for German given names, both male and female.

For a more formal analysis of the relationship between the popularity induced by search queries in the Nameling and corresponding frequencies within a Wikipedia corpus, we calculated Kendall's $\tau$ rank correlation coefficient [14] pairwise on the common set of names for all considered activity classes and both language editions of Wikipedia, as shown in Table III. Firstly, we note that the most pronounced correlation is indicated for pairs of popularity rankings within a system. Nevertheless, rankings induced by search queries in the Nameling and those induced by frequency within Wikipedia

TABLE III: Kendall's $\tau$ rank correlation coefficient, pairwise for the popularity induced rankings in the different systems.

| | Enter | Click | Favorite | Nameling | Wikipedia (DE) | Wikipedia (EN) |
|---|---|---|---|---|---|---|
| *Enter* | — | | | | | |
| *Click* | 0.62 | — | | | | |
| *Favorite* | 0.54 | 0.60 | — | | | |
| Nameling | 0.86 | 0.79 | 0.59 | — | | |
| Wikipedia (DE) | 0.35 | 0.35 | 0.33 | 0.40 | — | |
| Wikipedia (EN) | 0.28 | 0.26 | 0.25 | 0.33 | 0.69 | — |

are also assessed, which is slightly more pronounced for the German Wikipedia (e.g., $\tau = 0.40$ for the global ranking over all activity classes in the Nameling and the German Wikipedia versus $\tau = 0.33$ for the English Wikipedia).

## VII. RANKING PERFORMANCE OF STRUCTURAL SIMILARITY METRICS

Up to now, the performance of the various structural similarity metrics presented in III are evaluated statically with respect to an external notion of semantic relatedness, as summarized in Section V. This section presents a comparative evaluation of the considered similarity metrics' performance with respect to actual name preferences of users as indicated by the usage data presented in Section VI.

For this purpose, an according ranking scenario is formulated, where for a given name, a ranking on all names is obtained by calculating pairwise similarity with all other names based on the considered similarity metrics with respect to the co-occurrence networks obtained from Wikipedia. Then,

an established evaluation metric is applied for assessing the relative quality of the different similarity functions.

Firstly, we need a "ground truth" as a reference for evaluating the performance of a given ranking on the set of given names. Considering the usage data presented in Section VI, a first natural choice would be to predict for each search request of a user the names which will be added as favorites. But thereby, the evaluation would be biased towards the similarity metric which was implemented in Nameling during the period of evaluation. Thus, these evaluation target is only valid in a live setting, where different ranking systems are comparatively presented to a user.

We therefore consider for each user $u$ the corresponding
- set of names which were directly entered ($Enter_u$),
- the set of names the user clicked on ($Click_u$) and
- the set of $u$'s favorite names ($Favorite_u$).

We argue, that by directly entering a name into the search mask, a user already expresses interest in the corresponding name. Furthermore, a name which a user entered into the search mask is not directly influenced by the names which were presented based on the similarity function implemented in Nameling during the period of evaluation.

We evaluate the performance of a structural similarity metric $SIM$, by randomly selecting just one name $i \in Enter_u$ and calculating the *average precision* score $\text{AveP}(SIM(i, *), Enter_u)$ for every user $u$ and then calculating the *mean average precision* MAP (cf. Section III). For reference, we also computed as a baseline the MAP score for randomly ordered given names (labeled with "RND" in the corresponding figures). All results in this section are averaged over five repetitions of the corresponding random experiments.

Figure 5 shows the obtained results for each considered similarity metric in its weighted and unweighted variant, separately evaluated on the *Enter*, *Click* and *Favorite* sets. First of all we have to note, that all MAP scores are very low. Partly, this is due to the uncleaned evaluation data set which also contains search queries for names, which are not contained in the list of known names and therefore could not be listed by the considered similarity metrics. Also we retained from requiring a minimum number of search queries per user (despite that there must be at least one to predict). This renders the task of predicting "missing" names even more hard, as the profile of a user who searched for more than 20 names is expected to be more consistent than the profile of a user who just entered two search queries (depending on the similarity metric, the results are improved by $40\%$ to $150\%$ if only users with more than 10 search queries are considered). We argue that the usage data which is taken as a reference should be applied without further adjustments, as other ranking algorithms may overcome some of the mentioned factors. Such improvements can then be assessed by applying the same evaluation on the new system and comparing the gain in MAP over the present approaches.

In any case, all considered similarity metrics show significant better performance than the random baseline. We firstly consider the *Enter* sets. Notably, all but $PPR+$ are negatively

effected by the weights in the co-occurrence graph. On the other hand, the performance of $PPR+$ drops to the level of the considered baseline for the unweighted case. The dependence of $PPR+$ on the edge weights is in line with the motivation of its design, as the global PageRank score is subtracted to reduce the impact of global frequencies, which are absent in the unweighted case. Please note that the influence of weights is discussed in the related field of predicting links in social networks (cf. [20], [5]). Altogether, the PageRank based similarity metrics outperform the other metrics.

As for the *Click* set, the weighted cosine similarity $COS$ significantly outperforms all other metrics. This is in line with the bias towards the implemented similarity metric within Nameling, as all result lists were ordered according to the weighted cosine similarity during the time period of evaluation. Considering the *Favorite* set, the unweighted Jaccard coefficient $JAC$ performs surprisingly well, outperforming even the weighted cosine similarity.

Summing up, the results presented in this chapter indicate that the cosine similarity and the adjusted preference PageRank are candidates for recommending similar names, based on co-occurrences of given names within Wikipedia. For recommendations based on a single query term it yielded better results than the weighted cosine similarity, though slightly worse then the plain preference PageRank on the unweighted co-occurrence graphs. But it showed consistent well performance scores even in the evaluation sets *Click* and *Favorite* which are strongly biased towards the weighted cosine similarity. Nevertheless, the much simpler weighted cosine similarity showed comparable results.

Of course, the usage data can be used for personalizing search results or implementing recommendation systems based on the own name preferences and those of similar users. Even the very simple baseline recommender which just recommends the most popular names yields results which significantly outperform all statical name rankings considered in this paper. Building recommendation systems which apply collaborative filtering techniques [24] are expected to increase the prediction performance significantly. A thorough discussion and evaluation of according recommendation systems are subject to future research.

## VIII. CONCLUSION & FUTURE WORK

The present work introduces the research task of discovering relations among given names. A new approach for ranking names based on co-occurrence graphs is proposed and evaluated. For this, usage data from the running system Nameling is firstly introduced and thoroughly analyzed and then used for setting up an experimental framework for evaluating the performance of name rankings. The results presented in Section VII and V form a basis for deciding which similarity metric can be used for ranking names relative to a given set of query terms.

By making all considered usage data publicly available, other researchers are invited to build and evaluate new ranking and recommendation systems. The success of the Nameling
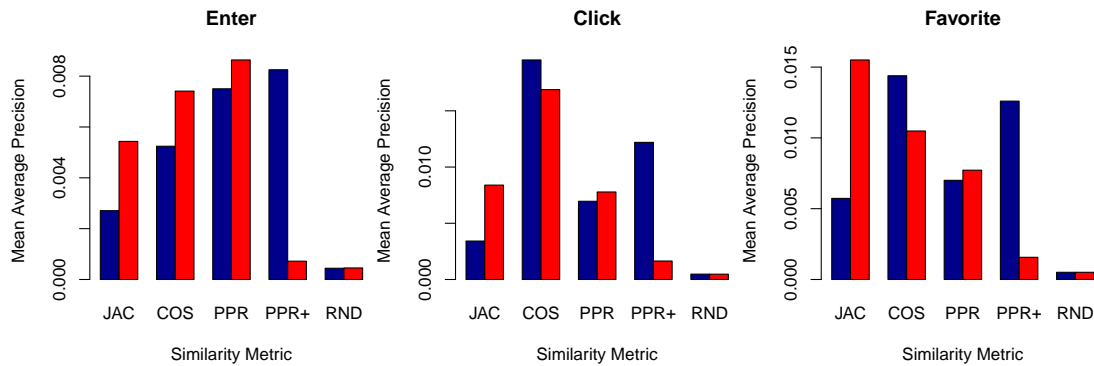
Fig. 5: Mean average precision for all considered structural similarity metrics and activity classes in the weighted (blue) and unweighted (red) case.

indicates that there is a need for such recommendation systems and the inherent interdisciplinary research questions render the task of discovering relations among given names fascinating and challenging to tackle.

For future work we plan to implement personalized recommendation and ranking systems, thereby incorporating further influencing factors, such as, e.g., the geographic distance among users. Additionally we plan to implement an open and flexible recommendation framework based on BibSonomy tag recommendation framework[13] which will allow other research to directly integrate and evaluate their approaches within the running system.

### REFERENCES

[1] S. Brin and L. Page. The anatomy of a large-scale hypertextual web search engine. In *Computer Networks and ISDN Systems*, pages 107–117, 1998.
[2] R. Bunescu and M. Pasca. Using encyclopedic knowledge for named entity disambiguation. In *Proc. of EACL*, volume 6, pages 9–16, 2006.
[3] C. Cattuto, D. Benz, A. Hotho, and G. Stumme. Semantic grounding of tag relatedness in social bookmarking systems. *The Semantic Web-ISWC 2008*, pages 615–631, 2008.
[4] T. Cohen and D. Widdows. Empirical distributional semantics: methods and biomedical applications. *J Biomed Inform*, 42(2):390–405, Apr. 2009.
[5] H. de Sá and R. Prudencio. Supervised link prediction in weighted networks. In *Neural Networks (IJCNN), The 2011 Int. Joint Conference on*, pages 2281–2288. IEEE, 2011.
[6] E. Gabrilovich and S. Markovitch. Computing semantic relatedness using wikipedia-based explicit semantic analysis. In *Proc. of the 20th int. joint conference on artificial intelligence*, volume 6, page 12. Morgan Kaufmann Publishers Inc., 2007.
[7] J. Golbeck and J. Hendler. Filmtrust: Movie recommendations using trust in web-based social networks. In *Proceedings of the IEEE Consumer communications and networking conference*, volume 96. Citeseer, 2006.
[8] G. Grefenstette. Finding semantic similarity in raw text: The deese antonyms. In *Fall Symposium Series, Working Notes, Probabilistic Approaches to Natural Language*, pages 61–65, 1992.
[9] A. Hotho, R. JÃd'schke, C. Schmitz, and G. Stumme. Information retrieval in folksonomies: Search and ranking. In Y. Sure and J. Domingue, editors, *The Semantic Web: Research and Applications*, volume 4011 of *Lecture Notes in Computer Science*, pages 411–426, Heidelberg, June 2006. Springer.

[10] A. Islam and D. Inkpen. Second order co-occurrence pmi for determining the semantic similarity of words. In *Proc. of the Int. Conference on Language Resources and Evaluation (LREC 2006)*, pages 1033–1038, 2006.
[11] R. Jäschke, L. Marinho, A. Hotho, L. Schmidt-Thieme, and G. Stumme. Tag recommendations in social bookmarking systems. *Ai Communications*, 21(4):231–247, 2008.
[12] G. Jeh and J. Widom. Simrank: a measure of structural-context similarity. In *Proc. of the eighth ACM SIGKDD int. conference on Knowledge discovery and data mining*, KDD '02, pages 538–543, New York, NY, USA, 2002. ACM.
[13] R. JÃd'schke, F. Eisterlehner, A. Hotho, and G. Stumme. Testing and evaluating tag recommenders in a live system. In D. Benz and F. Janssen, editors, *Workshop on Knowledge Discovery, Data Mining, and Machine Learning*, pages 44 –51, Sept. 2009.
[14] M. G. Kendall. A new measure of rank correlation. *Biometrika*, 30(1/2):81–93, 1938.
[15] T. Landauer and S. Dumais. A solution to plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review*, 104(2):211, 1997.
[16] E. A. Leicht, P. Holme, and M. E. J. Newman. Vertex similarity in networks, 2005. cite arxiv:physics/0510143.
[17] D. Liben-Nowell and J. Kleinberg. The link-prediction problem for social networks. *J. of the American society for inf. science and technology*, 58(7):1019–1031, 2007.
[18] G. Linden, B. Smith, and J. York. Amazon. com recommendations: Item-to-item collaborative filtering. *Internet Computing, IEEE*, 7(1):76–80, 2003.
[19] L. Lü, C. Jin, and T. Zhou. Similarity index based on local paths for link prediction of complex networks. *Physical Review E*, 80(4):046122, 2009.
[20] L. Lü and T. Zhou. Link prediction in weighted networks: The role of weak ties. *EPL (Europhysics Letters)*, 89:18001, 2010.
[21] B. Markines, C. Cattuto, F. Menczer, D. Benz, A. Hotho, and G. Stumme. Evaluating similarity measures for emergent semantics of social tagging. In *18th Int'l WWW Conference*, pages 641–641, 2009.
[22] T. Murata and S. Moriyasu. Link prediction of social networks based on weighted proximity measures. In *Web Intelligence, IEEE/WIC/ACM Int. Conference on*, pages 85–88. IEEE, 2007.
[23] M. Strube and S. P. Ponzetto. WikiRelate! Computing semantic relatedness using Wikipedia. In *proc. of the 21st national conference on Artificial intelligence - Volume 2*, AAAI'06, pages 1419–1424. AAAI Press, 2006.
[24] X. Su and T. Khoshgoftaar. A survey of collaborative filtering techniques. *Advances in Artificial Intelligence*, 2009:4, 2009.
[25] P. Turney. Mining the web for synonyms: Pmi-ir versus lsa on toefl. In *Proc. of the 12th European Conference on Machine Learning*, pages 491–502. Springer-Verlag, 2001.
[26] E. Voorhees, D. Harman, N. I. of Standards, and T. (US). *TREC: Experiment and evaluation in information retrieval*, volume 63. MIT press Cambridge, 2005.