# Axiomatic Analysis of Contact Recommendation Methods in Social Networks: An IR Perspective

Javier Sanz-Cruzado[1(✉)] , Craig Macdonald[2] , Iadh Ounis[2] ,
and Pablo Castells[1]

[1] Universidad Autónoma de Madrid, Madrid, Spain
{javier.sanz-cruzado,pablo.castells}@uam.es
[2] University of Glasgow, Glasgow, UK
{craig.macdonald,iadh.ounis}@glasgow.ac.uk

**Abstract.** Contact recommendation is an important functionality in many social network scenarios including Twitter and Facebook, since they can help grow the social networks of users by suggesting, to a given user, people they might wish to follow. Recently, it has been shown that classical information retrieval (IR) weighting models – such as BM25 – can be adapted to effectively recommend new social contacts to a given user. However, the exact properties that make such adapted contact recommendation models effective at the task are as yet unknown. In this paper, inspired by new advances in the axiomatic theory of IR, we study the existing IR axioms for the contact recommendation task. Our theoretical analysis and empirical findings show that while the classical axioms related to term frequencies and term discrimination seem to have a positive impact on the recommendation effectiveness, those related to length normalization tend to be not desirable for the task.

## 1 Introduction

With the large-scale growth of social network platforms such as Twitter or Facebook, recommender systems technology that targets explicit social scenarios has seen a surge of interest [32,37]. As part of this trend, the adaptation of Information Retrieval (IR) approaches to recommend people to connect to in the network have been particularly studied [17,34]. This specific class of recommender systems has the interesting property that users play a dual role: they are the users to whom we want to provide recommendations, but they are also the items we want to recommend [32]. Recently, it has been shown that classical IR weighting models – such as BM25 – can not only be used, but are also effective and efficient for the contact recommendation task [34].

In fact, recommender systems have always had strong connections with textual information retrieval (IR), since both tasks can be considered as particular cases of information filtering [9]. These ties have been materialized in the design and development of recommendation approaches based on IR models [2,10,39].

Content-based recommender systems [2] have been the most direct realization of such ties. However, we also note the collaborative filtering methods of [10,39], which employed the vector space model or query likelihood to their advantage.

In this paper, we analyze the reasons behind the effectiveness of IR approaches for the task of recommending contacts in social networks, through an exploratory analysis of the importance and validity of the fundamental IR axioms [13]. We start our analysis by examining contact recommendation methods that directly adapt IR models [34], as they provide a bridge between existing work on axiomatic analysis in IR models, and this new task. In particular, we empirically analyze whether satisfying the IR axioms leads to an increase in the performances of the algorithms. Interestingly, we find that while this is generally true, the axioms related to length normalization negatively impact the contact recommendation performance, since they interfere with a key evolutionary principle in social networks, namely preferential attachment [8].

## 2    Related Work

By identifying the set of properties that an IR model must (at least) follow to provide effective results, axiomatic thinking as developed by Fang et al. [12] has permitted to guide the development of both sound and effective IR approaches by explaining, diagnosing and improving them. In their seminal work, Fang et al. [12] proposed several heuristics (known as axioms) addressing different properties of the models such as the frequency of the query terms in the retrieved documents, the relative discrimination between query terms, or how a model deals with long documents. They also analyzed the effect such properties had on the effectiveness of state-of-the-art models such as BM25 [29] or query likelihood [27], and found that, with minor modifications to adhere to the different proposed axioms, the modified IR models achieved an improved retrieval performance.

Since the seminal work of Fang et al., the original axioms have been refined and expanded [13,35], and other additional properties of effective IR models have been studied, such as the semantic relations between queries and documents [14] or term proximity [38]. Recently, axiomatic analysis has been applied on neural IR models: Rennings et al. [28] proposed a method for empirically checking if the learned neural models fulfil the different IR axioms, while Rosset et al. [30] used the axioms as constraints for guiding the training of neural models. Beyond IR, axiomatic analysis has also expanded to other areas such as recommender systems, where Valcarce et al. [39,40] explored the benefits of penalizing users who rate lots of items when selecting neighbors in user-based kNN approaches.

In this paper, using the IR-based contact recommendation framework proposed by Sanz-Cruzado and Castells [34] as a basis, we map the IR axioms of Fang et al. [13] into the task of recommending people in social networks, and empirically analyze how valid and meaningful each axiom is for this task.

# 3   Preliminaries

We first introduce the notations we use during the rest of the paper. Given a social network, we represent its structure as a graph $\mathcal{G} = \langle \mathcal{U}, E \rangle$, where $\mathcal{U}$ denotes the set of people in the network and $E$ is the set of relationships between users. For each user $u \in \mathcal{U}$, we denote by $\Gamma(u)$ the set of users with whom $u$ has established relationships (the neighborhood of user $u$). In directed networks, three different neighborhoods can be considered depending on the link *orientation*: users who have a link towards $u$, $\Gamma_{in}(u)$; users towards whom $u$ has a link, $\Gamma_{out}(u)$ ; and the union of both, $\Gamma_{und}(u)$. We define $\Gamma_{inv}(u)$ as the inverse neighborhood of $u$, i.e. the neighborhood $u$ would have if the orientation of the links is reversed. Weighted networks additionally include a function $w : \mathcal{U}^2 \to \mathbb{R}$, where $w(u,v) > 0 \Leftrightarrow (u,v) \in E$. Unweighted networks can be seen as a particular case where $w : \mathcal{U}^2 \to \{0, 1\}$. Then, given a target user $u$, the contact recommendation task consists of suggesting a subset of users $\hat{\Gamma}_{out}(u) \subset \mathcal{U} \setminus \Gamma_{out}(u)$ towards whom $u$ has no links but who might be of interest for $u$. We define the recommendation task as a ranking problem, in which the result set $\hat{\Gamma}_{out}(u)$ is obtained and sorted by a ranking function $f_u : \mathcal{U} \setminus \Gamma_{out}(u) \to \mathbb{R}$.

**Relation Between IR and Contact Recommendation.** Since we explore the importance of IR axioms for contact recommendation, we need to establish connections between both tasks. We take for this purpose the mapping proposed in [34]: we fold the three spaces in the IR task (documents, queries and terms) into a single space for people to people recommendation, namely the users in the network. We map queries and documents to the target and candidate users, respectively. We also use the neighbors of both target and candidate users as equivalent to the terms contained in the queries and documents. As proposed by Sanz-Cruzado and Castells [34], we might use different neighborhoods to represent the target and candidate users (we could take either $\Gamma_{in}, \Gamma_{out}$ or $\Gamma_{und}$ for each of them). We denote by $\Gamma^q(u)$ the neighborhood representing the target user, and by $\Gamma^d(v)$ the one for the candidate user. The frequency of a term $t$ in a document is represented as an edge weight $w^d(v, t)$ in our mapping:

$$\text{freq}(t, v) = w^d(v, t) = w(v, t) \cdot \mathbb{1}_{[\Gamma^d \neq \Gamma_{in}]} + w(t, v) \cdot \mathbb{1}_{[\Gamma^d \neq \Gamma_{out}]} \qquad (1)$$

where $\mathbb{1}_x$ is equal to one when the condition $x$ is true, or 0 otherwise.

In textual IR, the frequency is the basis to establish a measure of how important a term is for a document, and it is always positive. Therefore, we assume that $w^d \geq 0$, and $w^d(v, t) = 0$ if and only if $t \notin \Gamma^d(v)$. The higher the importance of the link $(v, t)$, the higher the weight $w^d(v, t)$ should be. In our experiments (described in Sect. 6), we use the number of interactions (i.e. retweets, mentions) between users as an example definition of $w^d(v, t)$. In those network datasets where this type of information is not available, we simply use binary weights.

Finally, the document length is mapped to the sum of the weights of the neighborhood of the target user: $\text{len}(v) = \sum_{t \in \Gamma^l(v)} w^l(v, t)$, which can be seen as a generalized notion of vertex degree in the social graph. For some methods (such as BM25 [29]), we may consider a different neighborhood orientation when

computing the user "size"; This explains the different symbols $\Gamma^l, w^l$ (not necessarily equal to $\Gamma^d, w^d$) in the definition of $\text{len}(v)$. In this framework, as the IR models rely on common neighbors between the target and the candidate user, they can only recommend people at distance $2$[1]. Table 1 summarizes the relation between the IR and contact recommendation tasks. Further details about the mapping are described in [34].

**Table 1.** Relation between the IR and contact recommendation tasks.

| Information retrieval | Contact recommendation |
|---|---|
| Document collection, $D$ | Set of users, $\mathcal{U}$ |
| Query, $q$ | Target user's neighborhood, $\Gamma^q(u)$ |
| Document, $d$ | Candidate user's neighborhood, $\Gamma^d(u)$ |
| Term $t \in q/d$ | Neighbor user $t \in \Gamma^q(u)/\Gamma^d(v)$ |
| Documents containing a term, $D_t$ | User's inverse neighborhood, $\Gamma^d_{inv}(t)$ |
| Frequency of a term, $\text{freq}(t, d)$ | Weight of a link, $w^d(v, t)$ |
| Document length, $|d'|$ | Length of the user, $\text{len}(v)$ |

## 4    IR Axioms in Contact Recommendation

Before analyzing the importance of the IR axioms in the recommendation task, we first recall the IR axioms, and reformulate them using the mapping from IR to contact recommendation. In the remainder of this section, we take the seven axioms proposed by Fang et al. [13], divided into four categories, and analyze them.

### 4.1    Term Frequency Constraints (TFC)

The first family of axioms analyzes the role of the frequency of the query terms in the retrieved documents. Since term frequencies are represented as edge weights in our framework, we rename them as "edge weight constraints" (EWC) in our reformulation. The first constraint, TFC1, establishes that if the only difference between two documents is the frequency of a query term, then, the document with the higher term frequency should be ranked atop of the other. The intuition behind this axiom is naturally translated to contact recommendation by considering the "common friends" principle in social bonding: all things being equal, you are more likely to connect to people who have stronger bonds to common friends. This principle can be expressed as follows:

---

[1]    Distance is the minimum number of links you need to traverse from the target user to the candidate user, regardless of the *orientation* (direction) of the link.

**EWC1:** If the target user $u$ has a single neighbor $\Gamma^q(u) = \{t\}$, and we have two different candidate users $v_1, v_2$ such that $\text{len}(v_1) = \text{len}(v_2)$, and $w^d(v_1, t) > w^d(v_2, t)$, then we should have $f_u(v_1) > f_u(v_2)$.

The second term frequency constraint (TFC2) establishes that the ranking score increment produced by increasing term frequency should decrease with the frequency (i.e. ranking scores should have a dampened growth on term frequency, as in a diminishing returns pattern). This also has a direct meaning in the contact recommendation space: the difference in scores between two candidate contacts should decrease with the weights of their common friends with the target user. Formally, this constraint is expressed as:

**EWC2:** For a target user $u$ with a single neighbor $\Gamma^q(u) = \{t\}$, and three candidate users $v_1, v_2, v_3$ such that $\text{len}(v_1) = \text{len}(v_2) = \text{len}(v_3)$, and $w^d(v_3, t) = w^d(v_2, t) + 1$ and $w^d(v_2, t) = w^d(v_1, t) + 1$, then $f_u(v_2) - f_u(v_1) > f_u(v_3) - f_u(v_2)$.

Finally, the third axiom reflects the following property: occurrence frequencies and discriminative power being equal, the document that covers more distinct query terms should attain a higher score. In people recommendation, this translates to the triadic closure principle [25,26]: all other things being equal, the more common friends a candidate contact has with the target user, the higher the chance that a new link between them exists. Formally:

**EWC3:** Let $\{t_1, t_2\} \subset \Gamma^q(u)$ be two neighbors of target user $u$, with $\text{td}(t_1) = \text{td}(t_2)$. Given two candidate users $v_1, v_2$ with $\text{len}(v_1) = \text{len}(v_2)$, if $w^d(v_1, t_1) = w^d(v_2, t_1) + w^d(v_2, t_2)$, $t_2 \notin \Gamma^d(v_1)$, and $\{t_1, t_2\} \subset \Gamma^d(v_2)$, then $f_u(v_1) < f_u(v_2)$. where $\text{td}(t)$ is a measure of the informativeness of the common neighbors of the target and candidate users, as can be obtained from an IDF measure.

These three axioms are interdependent: if we take $\Gamma^q(u) = \{t\}$ and we fix the values for $\text{td}(t)$ and $\text{len}(v)$, we could rewrite $f_u(v)$ as a function of the document weight, $f_u(w^d(v, t))$. If $f_u(w^d(v, t))$ is positive, it is easy to see that EWC1 $\Leftrightarrow$ $f_u(w^d(v, t))$ is an increasing function, EWC2 $\Leftrightarrow$ $f_u(w^d(v, t))$ is strictly concave, and EWC3 $\Leftrightarrow$ $f_u(w^d(v, t))$ is strictly subadditive. Given a function $g$, $g$ positive and concave $\Rightarrow$ $g$ is increasing and subadditive. Therefore, for such functions (as is the case for most of the classic IR functions), EWC2 $\Rightarrow$ EWC1 $\wedge$ EWC3. However, if EWC2 is not satisfied, either EWC1 or EWC3 could still be satisfied.

## 4.2   Term Discrimination Constraint (TDC)

The term discrimination constraint is an axiom that formalizes the intuition that penalizing popular words in the collection (such as stopwords) and assigning higher weights to more discriminative query terms should produce better search results. This principle makes sense in contact recommendation: sharing a very popular and highly connected friend (e.g. two people following Katy Perry on Twitter) may be a rather weak signal to infer that these two people would relate to each other. A less social common friend, however, may suggest the two people may indeed have more interests in common. This idea is in fact reflected in some contact recommendation algorithms such as Adamic-Adar [1,22].

Hence, we rename the axiom as "neighbor discrimination constraint" (NDC), and we adapt the version of the axiom proposed by Shi et al. [35], which simplifies the translation to our domain, as follows:

**NDC:** Let $u$ be the target user, with $\Gamma^q(u) = \{t_1, t_2\}$. Given two candidate users $v_1, v_2$ where where $\text{len}(v_1) = \text{len}(v_2)$, and $w^d(v_1, t_1) = w^d(v_2, t_2)$ and $w^d(v_1, t_2) = w^d(v_2, t_1)$, if $w^d(v_1, t_1) > w^d(v_1, t_2)$ and $\text{td}(t_1) > \text{td}(t_2)$, then $f_u(v_1) > f_u(v_2)$.

### 4.3   Length Normalization Constraints (LNC)

The third family of IR axioms studies how algorithms should deal with the length of the documents. As defined in Sect. 3, in our mapping, the length of the document is translated to the sum of the edge weights between the candidate user and its neighbors: $\text{len}(v)$. As we only study the length of the candidate user, we will rename this family of constraints as "candidate length normalization constraints" (CLNC). Fang et al. [13] proposed two different LNCs.

The first axiom states that for two documents with the same query term occurrence frequency, we should choose the shorter one, since it contains the least amount of query-unrelated information. In contact recommendation, this means penalizing popular, highly connected candidate users with many neighbors not shared with the target user. We hence reformulate this axiom as:

**CLNC1:** Given a target user $u$ and two candidate users $v_1, v_2$, if $w^d(v_2, t) > w^d(v_1, t)$ for some user $t \notin \Gamma^q(u)$, but $w^d(v_1, x) = w^d(v_2, x)$ for any other user $x \neq t$, then $f_u(v_1) > f_u(v_2)$.

The second constraint aims to avoid over-penalizing long documents: it states that if a document is concatenated to itself multiple times, the resulting document should not get a lower score than the original. In contact recommendation, this means that, if we multiply all the edge weights of a candidate user by a positive number, the score for the candidate user should not decrease. Formally:

**CLNC2:** If two candidate users $v_1, v_2$ are such that $w^d(v_1, x) = k \cdot w^d(v_2, x)$ for all users $x$ and some constant $k > 1$, and $w^d(v_1, t) > 0$ for some neighbor $t \in \Gamma^q(u)$ of the target user $u$, then we have $f_u(v_1) \geq f_u(v_2)$.

### 4.4   Term Frequency – Length Normalization Constraint (TF-LNC)

The last heuristic aims to provide a balance between query term frequency in documents and length normalization. The axiom states that if we add more occurrences of a query term to a document, its retrieval score should increase. For contact recommendation, the intuition is similar: if the link weight between two users $v$ and $t$ increases, then $v$'s score as a candidate for target users having $t$ in their neighborhood should increase. This axiom is then expressed as follows:

**EW-CLNC:** Given a target user $u$ with a single neighbor $\Gamma^q(u) = \{t\}$, if two candidates $v_1$ and $v_2$ are such that $w^d(v_1, t) > w^d(v_2, t)$ and $\text{len}(v_1) = \text{len}(v_2) + w^d(v_1, t) - w^d(v_2, t)$, then $f_u(v_1) > f_u(v_2)$.

# 5   Theoretical Analysis

The first step to undertake an analysis of the IR axioms in contact recommendation is to determine the set of algorithms for which the different axioms are applicable, and, for those, to identify which constraints they satisfy and under which conditions. In this section, we provide an overview of different contact recommendation methods and their relation with the axioms.

We divide the approaches into two groups: friends of friends approaches, which only recommend people at network distance 2 from the target user, and methods which might recommend more distant users. The first group includes all IR models, as well as other approaches such as the most common neighbors (MCN) and Adamic-Adar's approach [22], whereas the second group includes matrix factorization [18,21], random walk-based methods [16,41] and kNN [2].

The proposed set of constraints is not applicable to the algorithms in the second group, since the constraints are based on the idea that the weighting functions depend on the common users between the target and the candidate users. Therefore, in the rest of the article, we focus on the algorithms in the first family. As future work, we envisage the formulation of new constraints tailored for algorithms that recommend users at distance greater than 2, possibly as a generalization of the set of constraints we study in this paper (see e.g. the formal analysis of pseudo-relevance feedback by Clinchant and Gaussier [11], which in our mapping would correspond to distance greater than 2).

We start analyzing the friends of friends methods by studying the IR models. In the adaptation of these models by Sanz-Cruzado and Castells [34], the components of the ranking functions (frequency/weight, discriminative power functions, document/user length) maintain the basic properties on which the formal analysis by Fang et al. [12,13] has relied. Therefore, the adapted methods satisfy the same constraints in the social network as those satisfied in the text IR space, and, if they are only satisfied under certain conditions, we can find the new conditions just by adapting them for the contact recommendation task. Then, models like PL2 [3,7], the pivoted normalization vector space model (VSM) [36] query likelihood with Dirichlet (QLD) [42] or Jelinek-Mercer smoothing (QLJM) [27] keep their original properties in this new space.

We find however one point of difference related to a possibility considered by Sanz-Cruzado and Castells in the definition of the candidate user length; namely, that we can define the length of the candidate users by selecting a different neighborhood $\Gamma^l(v)$ than the one used for defining the candidate user, $\Gamma^d(v)$, as explained in Sect. 3. As the only difference between the original and the version of BM25 defined by Sanz-Cruzado and Castells is just the definition of the candidate length, it is straightforward to prove that all edge weight constraints and NDC are satisfied in the same way as they are for textual IR: NDC is unconditionally true, whereas all EWC axioms depend just on the condition:

$$C_1 : |\Gamma^d_{inv}(t)| < |\mathcal{U}|/2 \tag{2}$$

which, in contact recommendation, is likely to be true – indeed, as of 2019, Twitter has >300 M users, and, the most followed user has just 107 M followers.

On the other hand, differences arise when we study the constraints involving length normalization: CLNCs and EW-CLNC. If we keep the same orientation for the user length and neighborhood selection for the candidate user, the mapping maintains the same components as the original ranking function, and, consequently, the condition for satisfying the three axioms is the same as the original: satisfying condition $C_1$. However, if the orientation for the length is changed, it is easy to show that, for CLNC1, BM25 satisfies the axiom if both conditions $C_1$ and $C_2$ are true, or both are false, where:

$$C_2 : \left( \Gamma^l := \Gamma^d \right) \vee \left( \text{len}(v_2) > \text{len}(v_1) \right) \tag{3}$$

and, for the EW-CLNC, the constraint is kept if conditions $C_1$ and $C_3$ are met, or none of them are, where:

$$C_3 : \left( \Gamma^l := \Gamma^d \right) \vee \left( \Gamma^l := \Gamma_{und} \right) \vee \left( \frac{1-b}{b} \operatorname*{avg}_{v'} \text{len}(v') > w^d(t, v_2) - \text{len}(v_2) \right) \tag{4}$$

The only length normalization-related constraint that is satisfied under the same conditions as the original BM25 model is the CLNC2 constraint, since it does not really depend on the definition of user length. Table 3 shows the differences between the original version and this adaptation of the BM25 model for contact recommendation. Hence, we introduce a new IR-based approach, namely the Extreme BM25 (EBM25) method, a variant of BM25 where we make the $k$ parameter tend to infinity. In comparison with BM25, all constraints are satisfied under the conditions specified for BM25, except EWC2 and EWC3, which are not satisfied at all for EBM25. In the BM25 model, under the conditions of EWC2, the $k$ parameter establishes how $f_u(v)$ grows as a function of the weight of the only common neighbor between the target and candidate users. The greater the value of $k$, the more the growth function approximates a linear function. When $k \to \infty$, the growth becomes linear, and as a consequence, the model does not meet the EWC2 constraint. A similar issue occurs with EWC3.

Beyond the IR models, other approaches such as Adamic-Adar or MCN do operate at distance 2. In the particular case of these methods, they consider neither weights nor any means of normalization; only EWC3 and CLNC2 are applicable here: under the conditions of EWC3, both methods just measure the number of common neighbors, satisfying the constraint. For CLNC2, if we multiply all the weights of the link for a candidate by any number $k \neq 0$, the score of the functions would not vary (and, consequently, they meet the axiom).

We summarize this analysis in Table 2, where we identify whether a method satisfies (fully or conditionally) or not the different axioms. In the case of the models not described in this section (pivoted normalization VSM, PL2, QLD), we refer to the article by Fang et al. [13] for further information on the conditions to satisfy the axioms. Next, we empirically analyze whether satisfying the axioms leads to an improvement of the performance of such algorithms.

**Table 2.** Constraint satisfaction for different contact recommendation algorithms.

| Algorithm | EWC1 | EWC2 | EWC3 | NDC | CLNC1 | CLNC2 | EW-CLNC |
|---|---|---|---|---|---|---|---|
| BM25 | Cond. | Cond. | Cond. | Yes | Cond. | Cond. | Cond. |
| EBM25 | Cond. | No | No | Yes | Cond. | Cond. | Cond. |
| Pivoted | Yes | Yes | Yes | Yes | Yes | Cond. | Cond. |
| PL2 | Cond. | Cond. | Cond. | Cond. | Cond. | Cond. | Cond. |
| QLD | Yes | Yes | Yes | Yes | Yes | Cond. | Yes |
| QLJM | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| MCN | No | No | Yes | No | No | Yes | No |
| Adamic-Adar | No | No | Yes | No | No | Yes | No |

**Table 3.** Constraint analysis results for BM25. By the equivalence notation e.g. $C_1 \equiv C_2$ we mean that $C_1$ and $C_2$ can only be either both true or both false.

| | TFC/EWC | | | TDC/NDC | LNC/CULNC | | TF-LNC/EW-CULNC |
|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | | 1 | 2 | |
| Text IR | $C_1$ | $C_1$ | $C_1$ | Yes | $C_1$ | $C_1$ | $C_1$ |
| Contact rec. | $C_1$ | $C_1$ | $C_1$ | Yes | $C_1 \equiv C_2$ | $C_1$ | $C_1 \equiv C_3$ |

## 6    Empirical Analysis

Prior work on axiomatic thinking [12,13] has analyzed to which extent the satisfaction of a suitable set of constraints correlates with effectiveness. This is also a mechanism to validate such constraints, showing that it is useful to predict, explain or diagnose why an IR system is working well or badly. Taking up this perspective, we undertake next such an empirical analysis of constraints in the contact recommendation setting, using a set of friends-of-friends algorithms.

### 6.1    Experimental Setup

**Data:** We use different network samples from Twitter and Facebook: the ego-Facebook network released in the Stanford Large Network Dataset collection [24], and two Twitter data downloads described in [34] as 1-month and 200-tweets. The Twitter downloads include each two different sets of edges for the same set of users: the follow network (where $(u, v) \in E$ if $u$ follows $v$), and the interaction network (where $(u, v) \in E$ if $u$ retweeted or mentioned $v$). The datasets are described in more detail in [32–34].

For evaluation purposes, we partition each network into a training graph that is supplied as input to the recommendation algorithms, and a test graph that is held out for evaluation. Using the test graph, IR metrics such as precision, recall or nDCG can be computed, as well as other accuracy metrics such as AUC [15], by considering test edges as binary relevance judgements: a user $v$ is relevant to

**Table 4.** Dataset statistics

| | Twitter 1-month | | Twitter 200-tweets | | Facebook |
|---|---|---|---|---|---|
| | Interactions | Follows | Interactions | Follows | |
| Directed | Yes | Yes | Yes | Yes | No |
| Users with links | 9,528 | 9,770 | 9,985 | 9,964 | 4,039 |
| Training edges | 170,425 | 645,022 | 137,850 | 475,730 | 70,566 |
| Validation edges | 33,867 | 46,628 | 29,131 | 46,760 | 14,100 |
| Test edges | 54,335 | 81,110 | 21,598 | 98,519 | 17,643 |

a user $u$ if – and only if – the edge $(u, v)$ appears in the test graph. We further divide the training graph into a smaller training graph and a validation graph for parameter tuning. Table 4 shows the size of the different resulting subgraphs.

For all Twitter networks, temporal splits are applied: the training data includes edges created before a given time, and the test set includes links created afterwards. Edges appearing in both sides of the split are removed from the test network. For the interaction network, two different temporal points are selected to generate the split: July $5^{th}$ and July $12^{th}$ in the 1-month dataset, and July $24^{th}$ and July $29^{th}$ in 200-tweets. Weights for the training graphs were computed by counting the number of interactions before the splits.

For the follow networks, the edges between the users of the interaction network were downloaded three times: the first download is used as training graph for parameter tuning; the new links in the second snapshot (not present in the initial one), downloaded four months later, are used as the validation set; the complete second snapshot is given as input to the recommendation algorithms under evaluation; finally, the new edges in the third download (not present in the second), obtained two years afterwards, are used as the test data for evaluation.

For the Facebook data, since temporal information is not available, we apply a simple random split: 80% of links are sampled as training and 20% as test; within the training data, we use 25% of the edges as the validation subset.

**Algorithms:** We focus on contact recommendation approaches that recommend users at distance 2. From that set, as representative IR models, we include adaptations for the pivoted normalization vector space model [36]; BIR and BM25 [29] as probabilistic models based on the probability ranking principle; query likelihood [27] with Jelinek-Mercer [20], Dirichlet [23] and Laplace [39] smoothing as language models; and PL2 [3,7], DFRee, DFReeKLIM [6], DPH [4] and DLH [5] as divergence from randomness approaches. In addition, we include adaptations of a number of link prediction methods [22] (following [34]): Adamic-Adar [1], Jaccard [19], most common neighbors [22] and cosine similarity [31].[2]

---

[2] Code and additional details about the experimental configuration are available at https://github.com/ir-uam/contact-rec-axioms.

## 6.2   Experiments and Results

**Edge Weight Constraints (EWCs):** We start by analyzing the edge weight constraints. Since weights are binary in the Twitter follow graphs and Facebook, we focus here on interaction graphs, where the interaction frequency provides a natural basis for edge weighting.

**Table 5.** Average AUC values for the most common neighbors algorithm for the different datasets, using $\Gamma^q := \Gamma_{und}$ and $\Gamma^d := \Gamma_{in}$ in the directed networks.

| Twitter 1-month | | Twitter 200-tweets | | Facebook |
|---|---|---|---|---|
| Interactions | Follows | Interactions | Follows | |
| 0.7545 | 0.8327 | 0.7064 | 0.7951 | 0.9218 |

A first natural question that arises when we study these axioms is whether the weights are useful or not for providing good recommendations. This is equivalent to test the importance of the first axiom for the contact recommendation task. To answer that question, we compare the two options (binarized vs. not binarized weights) in all algorithms which make use of weights: cosine similarity between users and all the IR models except BIR. We show the results in Fig. 1(a), where each dot represents a different approach. In the $x$ axis, we show the nDCG@10 value for the unweighted approaches, whereas the $y$ axis shows nDCG@10 for the weighted ones. We can see that using weights results in an inferior performance in all algorithms except for BM25 and the simple cosine similarity. These observations suggest that EWC1 does not appear to be a reliable heuristic for contact recommendation in networks.

However, once the weight is important for a model (and, therefore, EWC1 is important) does satisfying the rest of the edge weight constraints provide more accurate recommendations? To check that, similarly to Fang et al. [12,13], we compare an algorithm that satisfies all three EWCs (and benefits from weights) with another one that does not satisfy EWC2 and EWC3: we compare BM25 vs. EBM25. Fixing the $k$ parameter for the BM25 model (using the optimal configuration from our experiments), we compare different parameter configurations for BM25 and EBM25. Results are shown in Fig. 1(b), where every dot in the plot corresponds to a different model configuration, the $x$ axis represents the nDCG@10 values for BM25, and the $y$ axis those of the EBM25 model. As it can be observed, EBM25 does not improve over BM25 for almost every configuration (dots are all below the $y = x$ plane), thus showing that, as long as EWC1 is important for the model, both EWC2 and EWC3 are relevant.

As explained in Sect. 4, EWC3 can also be satisfied independently of EWC1 and EWC2, so we finally check its importance. For that purpose, we address the following question: for any friends-of-friends algorithm, such as Adamic-Adar [1] or the IR models, is it beneficial to reward the number of common users between the target and the candidate users? To analyze this, we compare

the MCN approach (which satisfies the constraint) with a binarized version of MCN which returns all people at distance 2 regardless of the common neighbor count. Restricting the test set to people at distance 2, Table 5 shows the resulting AUC [15] of the MCN algorithm, averaged over users on each network. Under these conditions, the binarized version would have an AUC value of 0.5. Hence, our results show that the number of common neighbors seem to be a strong signal for providing accurate recommendations (and, therefore, EWC3 seems to be important on its own for the contact recommendation task).

**Neighbor Discrimination Constraint (NDC):** As previously explained, this constraint suggests penalizing highly popular common neighbors. In IR approaches, this constraint is satisfied or not depending on the presence or absence of a term discrimination element (such as the Robertson-Spärck-Jones in BM25/EBM25 or the $p_c(t)$ term in query likelihood approaches). Therefore, to check the effectiveness benefit of this axiom, we compare – in terms of nDCG@10 – the BM25, EBM25, QLD, QLJM and the pivoted normalization VSM models with variants of them that lack term discrimination.
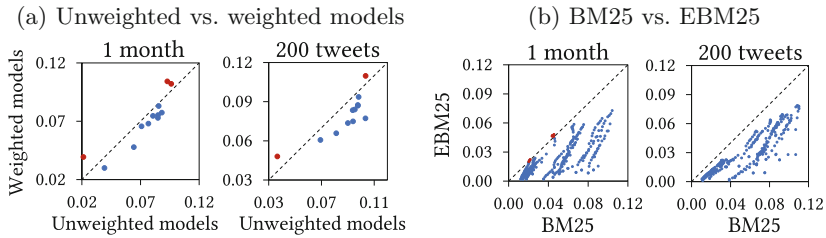


**Fig. 1.** For the Twitter interaction datasets: (a) nDCG@10 comparison between the weighted (y axis) and unweighted (x axis) versions of different contact recommendation algorithms. (b) nDCG@10 comparison between weighted versions of BM25 (x axis) and EBM25 (y axis). In both graphs, red dots represent those elements such that the value of nDCG@10 is greater for the y axis than for the x axis. (Color figure online)
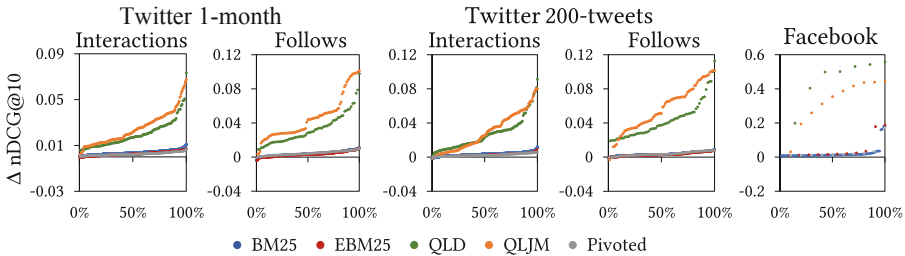


**Fig. 2.** Difference in nDCG with and without term discrimination for different configurations of IR-based algorithms, sorted by difference value. Each dot represents a different configuration of the corresponding algorithm. A positive value indicates that the variant with term discrimination is more effective.

Figure 2 shows the difference between different variants of each model. In the figure, a positive value indicates that the original version (with term discrimination) performs better. We observe that in an overwhelming majority of points the original versions achieve a better accuracy, hence NDC appears to be key to providing good contact recommendations. This confirms the hypothesis in many recommendation approaches that using high-degree users to discriminate which users are recommended does not seem to be a good idea [1,43].

**Length Normalization Constraints (CLNCs and EW-CLNC):** Finally, we study the effect of normalizing by candidate user length. For that purpose, similarly to the previous section, we compare the BM25, EBM25, QLJM, QLD and the pivoted normalization VSM models with versions of the models lacking the normalization by the candidate user length (which do not satisfy CLNC1 and EW-CLNC) using nDCG@10. We show a graph showing the differences in accuracy between different variants of the algorithms in Fig. 3(a). Since there are few differences between datasets, we only show results for the interactions network of the Twitter 1-month dataset. In the figure, we observe an opposite trend to what was expected: instead of performing worse, the algorithms without normalization do improve the results. Therefore, it seems that the different length normalization constraints are not useful for contact recommendation.
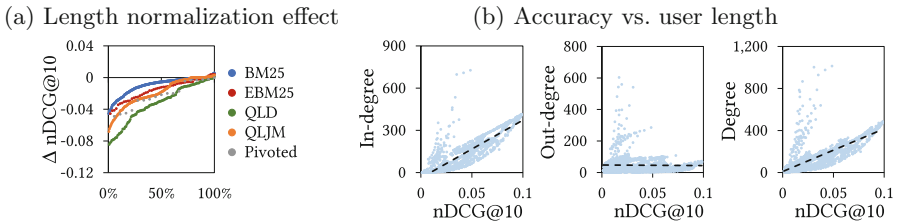


**Fig. 3.** For the Twitter 1-month interaction network: (a) Difference in nDCG with and without length normalization for different configurations of IR-based algorithms, sorted by difference value. A positive value indicates that the variant with length normalization is more effective. (b) Comparison between nDCG@10 and the average in-degree and out-degree of the recommended users.

These observations are consistent with the preferential attachment phenomenon in social networks [8], whereby high-degree users are more likely to receive new links than long-tail degree users. As an example, we check this in Fig. 3(b), where we compare the performances of the recommendation approaches listed in Section 6.1 with the average in-degree, out-degree and (undirected) degree of the recommended people. We observe that, in general, in-degree and degree are clearly correlated with the performances of the methods, as the principle indicates. With out-degree this is not so clear though. This explains the few configurations in Fig. 3(a) that do not improve when we remove the normalization: all of them normalize by the sum of the weights of the outgoing links of the candidate users. Similar trends are observed in other networks.

## 7    Conclusions

We have theoretically and empirically analyzed the importance of the fundamental IR axioms for the contact recommendation task in social networks. Theoretically, we have translated the different axioms proposed in [13] to the contact recommendation task, and we have checked whether the mapping introduced in [34] is sound and complete. We have found that, in general, the properties of the IR models are held in the recommendation task when we apply this mapping, unless we use a different definition for the document length from the usual. Empirically, we have conducted several experiments over various Twitter and Facebook networks to check if those axioms have any positive effect on the accuracy of the recommenders. We showed that satisfying the constraints related to term frequencies and term discrimination have a positive impact on the accuracy. However, those related to length normalization tend to have the opposite effect, as they interfere with a basic evolutionary principle of social networks, namely preferential attachment [8].

## References

1. Adamic, L.A., Adar, E.: Friends and neighbors on the Web. Soc. Netw. **25**(3), 211–230 (2003)
2. Adomavicius, G., Tuzhilin, A.: Toward the next generation of recommender systems: a survey of the state-of-the-art and possible extensions. IEEE Trans. Knowl. Data Eng. **17**(6), 734–749 (2005)
3. Amati, G.: Probability information models for retrieval based on divergence from randomness. Ph.D. thesis, University of Glasgow (2003)
4. Amati, G.: Frequentist and Bayesian approach to information retrieval. In: Lalmas, M., et al. (eds.) ECIR 2006. LNCS, vol. 3936, pp. 13–24. Springer, Heidelberg (2006). https://doi.org/10.1007/11735106_3
5. Amati, G., Ambrosi, E., Bianchi, M., Gaibisso, C., Gambosi, G.: FUB, IASI-CNR and University of Tor Vergata at TREC 2007 blog track. In: Proceedings of the 16th Text REtrieval Conference (TREC 2007). NIST (2007)
6. Amati, G., et al.: FUB, IASI-CNR, UNIVAQ at TREC 2011 microblog track. In: Proceedings of the 20th Text REtrieval Conference (TREC 2011). NIST (2011)
7. Amati, G., Van Rijsbergen, C.J.: Probabilistic models of information retrieval based on measuring the divergence from randomness. ACM Trans. Inf. Syst. **20**(4), 357–389 (2002)
8. Barabàsi, A.L., Albert, R.: Emergence of scaling in random networks. Science **286**(5439), 509–512 (1999)
9. Belkin, N.J., Croft, W.B.: Information filtering and information retrieval: two sides of the same coin? Commun. ACM **35**(12), 29–38 (1992)
10. Bellogín, A., Wang, J., Castells, P.: Bridging memory-based collaborative filtering and text retrieval. Inf. Retrieval **16**(6), 697–724 (2013)

11. Clinchant, S., Gaussier, E.: A theoretical analysis of pseudo-relevance feedback models. In: Proceedings of the 2013 Conference on the Theory of Information Retrieval (ICTIR 2013), pp. 6–13. ACM (2013)
12. Fang, H., Tao, T., Zhai, C.: A formal study of information retrieval heuristics. In: Proceedings of the 27th annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR 2004), pp. 49–56. ACM (2004)
13. Fang, H., Tao, T., Zhai, C.: Diagnostic evaluation of information retrieval models. ACM Trans. Inf. Syst. **29**(2), 1–42 (2011)
14. Fang, H., Zhai, C.: Semantic term matching in axiomatic approaches to information retrieval. In: Proceedings of the 29th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR 2006), pp. 115–122. ACM (2006)
15. Fawcett, T.: An introduction to ROC analysis. Pattern Recogn. Lett. **27**(8), 861–874 (2006)
16. Goel, A., Gupta, P., Sirois, J., Wang, D., Sharma, A., Gurumurthy, S.: The who-to-follow system at Twitter: strategy, algorithms, and revenue impact. Interfaces **45**(1), 98–107 (2015)
17. Hannon, J., Bennett, M., Smyth, B.: Recommending Twitter users to follow using content and collaborative filtering approaches. In: Proceedings of the 4th ACM Conference on Recommender Systems (RecSys 2010), pp. 199–206. ACM (2010)
18. Hu, Y., Koren, Y., Volinsky, C.: Collaborative filtering for implicit feedback datasets. In: Proceedings of the 8th IEEE International Conference on Data Mining (ICDM 2008), pp. 263–272. IEEE (2008)
19. Jaccard, P.: Étude comparative de la distribution florale dans une portion des Alpes et des Jura. Bulletin de la Société Vaudoise des Sciences Naturelles **37**(142), 547–579 (1901)
20. Jelinek, F., Mercer, R.: Interpolated estimation of Markov source parameters from sparse data. In: Gelsema, E.S., Kanal, L.N. (eds.) Pattern Recognition in Practice, pp. 381–402. North-Holland (1980)
21. Koren, Y., Bell, R., Volinsky, C.: Matrix factorization techniques for recommender systems. Computer **42**(8), 30–37 (2009)
22. Liben-Nowell, D., Kleinberg, J.: The link-prediction problem for social networks. J. Am. Soc. Inf. Sci. Technol. **58**(7), 1019–1031 (2007)
23. MacKay, D.J.C., Peto, L.C.B.: A hierarchical Dirichlet language model. Nat. Lang. Eng. **1**(3), 289–307 (1995)
24. McAuley, J., Leskovec, J.: Learning to discover social circles in ego networks. In: Proceedings of the 25th International Conference on Neural Information Processing Systems (NIPS 2012), pp. 539–547. Curran Associates Inc. (2012)
25. Newman, M.E.J.: Clustering and preferential attachment in growing networks. Phys. Rev. E **64**, 025102 (2001)
26. Newman, M.E.J.: Networks: An Introduction, 1st edn. Oxford University Press, Oxford (2010)
27. Ponte, J.M., Croft, W.B.: A language modeling approach to information retrieval. In: Proceedings of the 21st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR 1998), pp. 275–281. ACM (1998)
28. Rennings, D., Moraes, F., Hauff, C.: An axiomatic approach to diagnosing neural IR models. In: Azzopardi, L., et al. (eds.) ECIR 2019. LNCS, vol. 11437, pp. 489–503. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-15712-8_32
29. Robertson, S.E., Zaragoza, H.: The probabilistic relevance framework: BM25 and beyond. Found. Trends Inf. Retrieval **3**(4), 333–389 (2009)

30. Rosset, C., Mitra, B., Xiong, C., Craswell, N., Song, X., Tiwary, S.: An axiomatic approach to regularizing neural ranking models. In: Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR 2019), pp. 981–984. ACM (2019)

31. Salton, G., Wong, A., Yang, C.: A vector space model for automatic indexing. Commun. ACM **18**(11), 613–620 (1975)

32. Sanz-Cruzado, J., Castells, P.: Contact recommendations in social networks. In: Berkovsky, S., Cantador, I., Tikk, D. (eds.) Collaborative Recommendations: Algorithms, Practical Challenges and Applications, pp. 519–569. World Scientific Publishing (2018)

33. Sanz-Cruzado, J., Castells, P.: Enhancing structural diversity in social networks by recommending weak ties. In: Proceedings of the 12th ACM Conference on Recommender Systems (RecSys 2018), pp. 233–241. ACM (2018)

34. Sanz-Cruzado, J., Castells, P.: Information retrieval models for contact recommendation in social networks. In: Azzopardi, L., et al. (eds.) ECIR 2019. LNCS, vol. 11437, pp. 148–163. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-15712-8_10

35. Shi, S., Wen, J.R., Yu, Q., Song, R., Ma, W.Y.: Gravitation-based model for information retrieval. In: Proceedings of the 28th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR 2005), pp. 488–495. ACM (2005)

36. Singhal, A., Choi, J., Hindle, D., Lewis, D.D., Pereira, F.C.N.: AT&T at TREC-7. In: Proceedings of the 7th Text REtrieval Conference (TREC 1998), pp. 186–198. NIST (1998)

37. Tang, J., Hu, X., Liu, H.: Social recommendation: a review. Soc. Netw. Anal. Min. **3**(4), 1113–1133 (2013). https://doi.org/10.1007/s13278-013-0141-9

38. Tao, T., Zhai, C.: An exploration of proximity measures in information retrieval. In: Proceedings of the 30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR 2007), pp. 295–302. ACM (2007)

39. Valcarce, D., Parapar, J., Barreiro, Á.: Axiomatic analysis of language modelling of recommender systems. Int. J. Uncertainty Fuzziness Knowl.-Based Syst. **25**(Suppl. 2), 113–127 (2017)

40. Valcarce, D., Parapar, J., Barreiro, Á.: Finding and analysing good neighbourhoods to improve collaborative filtering. Knowl.-Based Syst. **159**, 193–202 (2018)

41. White, S., Smyth, P.: Algorithms for estimating relative importance in networks. In: Proceedings of the 9th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD 2003), pp. 266–275. ACM (2003)

42. Zhai, C., Lafferty, J.: A study of smoothing methods for language models applied to information retrieval. ACM Trans. Inf. Syst. **22**(2), 179–214 (2004)

43. Zhou, T., Lü, L., Zhang, Y.C.: Predicting missing links via local information. Eur. Phys. J. B **71**(4), 623–630 (2009)