# Axiomatic Analysis of Co-occurrence Similarity Functions

**U Kang** **Mikhail Bilenko**[†] **Dengyong Zhou**[†]
**Christos Faloutsos**

February 2012
CMU-CS-12-102

School of Computer Science
Carnegie Mellon University
Pittsburgh, PA 15213

[†]Microsoft Research, Redmond, WA, USA

Research was sponsored by the Defense Threat Reduction Agency under contract No. HDTRA1-10-1-0120, and by the Army Research Laboratory under Cooperative Agreement Number W911NF-09-2-0053. This work is also partially supported by an IBM Faculty Award, a Google Focused Research Award, and a Yahoo Research Alliance Gift. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the Army Research Laboratory, the U.S. Government, or other funding parties. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation here on.

**Keywords:** co-occurrence similarity function, axiomatic analysis

**Abstract**

Finding similar items based on co-occurrence data is an important data mining task with applications ranging from recommender systems to keyword based advertising. A number of co-occurrence similarity functions have been proposed based on graph-theoretic, geometric, and statistical abstractions. Despite the variety of existing algorithms, however, there exists no formal methodology for analyzing their properties and comparing their benefits and limitations. At the same time, the wide range of applications and domains where co-occurrence-based similarity functions are deployed limits the conclusiveness of experimental evaluations beyond the narrow task typically considered by each method.

This paper proposes an axiomatic approach to analyzing co-occurrence similarity functions. The approach is based on formulating general, domain-independent constraints that well-behaved methods must satisfy to avoid producing degenerate results. Such constraints are derived based on the impact that continuous aggregation of the co-occurrence data is expected to have on absolute or relative similarity estimates. Proposed constraint-based analysis is applied to several representative, popular similarity functions and reveals that, surprisingly, none of them satisfy all constraints unconditionally. The analysis leads to the design of a theoretically well-justified similarity function called Random Walk with Sink (RWS). RWS is parameterized to satisfy the constraints unconditionally, with the parameterization having interesting probabilistic and graph-theoretic interpretations.

# 1 Introduction

Co-occurrence data is ubiquitous in modern data mining and machine learning applications as it provides a very rich signal source for inferring similarity between items, a common prediction task. The following are examples of problems where different types of co-occurrences are used to identify related items:

1. *Item recommendation*. Logs of consumption behavior (market-basket data) allow finding products for cross-promotion (e.g., in personalized recommendations) [17].
2. *Query suggestion*. Search engine logs associating queries with subsequently visited URLs allow identifying related queries based on URL co-visitation (e.g., for query suggestion or matching advertisements on relevant keywords) [5, 14].
3. *Related author search*. Bibliography databases containing co-authorship data or co-occurrences of publications in the same venues allow finding similar authors (e.g., for finding related work, collaborators or qualified reference letter writers) [18].

Because of the wide applicability of co-occurrence similarity functions, a number of them have been proposed in the context of different domains. Such methods are roughly grouped into the following groups based on underlying formalizations:

1. *Graph-theoretic* methods represent items as nodes in a bipartite graph, with occurrence contexts being opposite partition nodes connected to item nodes by edges representing occurrences. Similarity corresponds to node nearness measures (e.g., probability of reaching another node via a $k$-step random walk).
2. *Geometric* methods represent items as vectors in a metric space with occurrences corresponding to dimensions. Similarity corresponds to geometric measures of vector closeness (e.g., cosine similarity).
3. *Probabilistic* methods represent items as random events over which contexts define probability distributions, based on which similarity is then computed (e.g., Pointwise Mutual Information).

It is important to note that these groups are not disjoint, as a number of methods can be interpreted as belonging to more than one. However, the bipartite graph representation of co-occurrence data provides a common underlying formalism, and plays a central role in this paper.

## 1.1 Axiomatic Approach

Because similarity functions are typically a component in learning and mining applications, their experimental evaluation is tied to the specific task and domain at hand. Their performance then depends on the application suitability, making empirical evaluations highly domain-specific. Thus, a fundamental unanswered question remains: how do we comparatively analyze different co-occurrence-based similarity methods? This paper describes a general framework that provides the basis for such comparative analysis. The framework is based on an *axiomatic approach*: deriving fundamental properties (axioms) that capture basic intuitions that any reasonable co-occurrence-based similarity measure must obey. These properties are obtained by considering changes in similarity function output that are expected when new item co-occurrences are observed. Distinguishing between occurrences arriving in new or existing contexts, as well as between absolute versus relative similarity leads to several types of constraints described in the paper.

| Method \ Axiom | A-NCM | R-NCM | A-QCM | R-QCM | A-TCM | R-TCM | DR |
|---|---|---|---|---|---|---|---|
| Common Neighbors | Yes | Yes | Yes | Yes | Yes | Yes | No |
| Cosine | No† | No† | No† | Yes | No† | No† | No† |
| Jaccard | No† | Yes | No | No | No | No | No† |
| Pointwise Mutual Information | No† | No† | No | No | No | No | No† |
| Adamic-Adar | Yes | Yes | No | No | No | No | No |
| Forward Random Walk | No† | Yes | No† | Yes | Yes | Yes | No† |
| Backward Random Walk | No† | No† | Yes | Yes | No† | No† | No† |
| Mean Meeting Time | No† | No† | No† | No† | No† | No† | No† |
| **Random Walk with Sink** | **Yes*** | **Yes** | **Yes*** | **Yes** | **Yes** | **Yes** | **Yes*** |

Table 1: Summary of the axiomatic analysis. 'Yes': a method satisfies an axiom on any data. 'Yes*': a method satisfies an axiom on any data given appropriate parameter settings. 'No†': a method satisfies an axiom only on a specific set of data. 'No': a method does not satisfy an axiom on any data. Notice that only our proposed Random Walk with Sink (RWS) method, described in Section 5, satisfies *all* the axioms on *any* data given appropriate parameter settings.

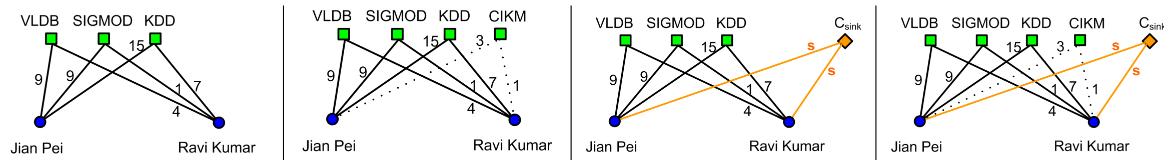
## 1.2 Our Contributions

Applying axiomatic analysis to a number of popular similarity functions yields surprising results: no single method satisfies all constraints unconditionally, as summarized in Table 1. For example, Figure 1 (a) and (b) show that the Forward Random Walk (FRW) similarity *decreases* after the addition a new context. For each method and each axiom, we either prove that the axiom is always satisfied, or identify the specific conditions that lead to axiom dissatisfaction. The ultimate utility of the proposed analysis framework is that it allows considering the shortcomings of current functions systematically, leading to derivation of their variants that overcome them, e.g., via data-dependent parameterization. This process is demonstrated by introducing a new variant of random walk-based similarity: *random walks with sink (RWS)*. The method has an intuitive interpretation that explains its flexibility related to smoothing, and it avoids axiom violations suffered by regular FRW; e.g., Figure 1 (c) and (d) show how RWS satisfies an axiom which was violated by FRW.

The axiomatic approach has been previously applied to clustering and information retrieval functions [10, 7, 1], leading to their better understanding, and our results indicate that it can be equally fruitful for analyzing co-occurrence similarity functions. The primary contributions of the paper are the following:

1. **Axiomatic framework.** We propose a principled methodology for analyzing co-occurrence similarity functions based on differential response to new observations.
2. **Analysis and proofs.** We analyze a number of commonly used similarity functions using our proposed abstraction called context-wise decomposition, and derive the conditions under which they satisfy the axioms. We prove that no single method satisfies all conditions unconditionally.
3. **Design of new similarity function.** We demonstrate how axiomatic analysis allows designing a new data-driven, theoretically well-justified co-occurrence similarity function without degenerate properties.

The rest of this paper is organized as follows. Section 2 surveys commonly used methods for computing similarities in bipartite graphs. Section 3 introduces the axiomatic framework, defining desirable properties expected of well-behaved similarity functions, followed by Section 4 where the framework is applied to analyze the popular similarity functions. Based on the analysis, we propose a new similarity function which




(a) Forward Random Walk: before adding 'CIKM'. Similarity: 0.2558

(b) Forward Random Walk: after adding 'CIKM'. Similarity: ↘ 0.2553.

(c) Random Walk with Sink: before adding 'CIKM'. Similarity: 0.06347

(d) Random Walk with Sink: after adding 'CIKM'. Similarity: ↗ 0.06758.

Figure 1: The comparison of the Forward Random Walk (FRW) and our proposed Random Walk with Sink (RWS) method. In (a) and (b), the new context 'CIKM' *decreases* the FRW similarity between Jian Pei and Ravi Kumar, which is counterintuitive and violating the Axiom 1 'New Context Monotonicity'. In (c) and (d), the same context increases the RWS similarity with the parameter $s = 100$, showing that RWS satisfies the Axiom 1.

satisfies all the proposed properties in Section 5. We summarize related work in Section 6 and conclude in Section 7.

## 2 Similarity Functions for Co-occurrence Data

In this section, we describe several highly popular co-occurrence similarity functions and briefly discuss their properties. All considered functions compute the similarity between a *query* item $q$ and a *target* item $u \in T \setminus \{q\}$, where $T$ is the set of all items. Items can be represented as nodes on one side of a bipartite graph, with the set of contexts, $C$, in which items are observed, represented by nodes on the other side of the bipartite graph. Graph edges encode occurrences, and may be unweighted or weighted, where weights would represent occurrence properties, e.g., occurrence count. For a dataset with $n$ items and $m$ contexts, the graph corresponds to an $n \times m$ adjacency matrix $W$ where its $(i, j)$th element $W_{ij}$ represents the occurrence weight of $i$-th item in $j$-th context. Table 2 summarizes the notation.

**Common Neighbors (CN)** method computes the similarity of two nodes as the total number of their common neighbors, and has been extensively used in social network graphs [15]. In a co-occurrence graph, this corresponds to computing the dot product of the vectors representing the neighborhoods of the two nodes. Let $\Gamma(q)$ and $\Gamma(u)$ be the sets of nodes connected to nodes $q$ and $u$, respectively. Then, the common neighbor similarity $CN_W(q, u)$ of a target node $u$ to the query $q$ based on weight matrix $W$ is:

$$CN_W(q, u) = \sum_{c \in \Gamma(q) \cap \Gamma(u)} W_{qc} W_{uc}.$$

**Cosine (COS)** similarity normalizes Common Neighbors by the total number of contexts in which the query and target items are observed, and is especially popular for computing textual similarity in information retrieval applications [13]. Formally, the similarity $COS_W(q, u)$ of a target node $u$ to the query $q$ based on weight matrix $W$ is:

$$COS_W(q, u) = \sum_{c \in \Gamma(q) \cap \Gamma(u)} \frac{W_{qc} W_{uc}}{||W_{q:}||_2 ||W_{u:}||_2},$$

where $W_{q:}$ and $W_{u:}$ are the $q$th and $u$th row of the $W$ matrix, respectively.

| Symbol | Definition |
|---|---|
| $q$ | Query item with respect to which similarities of other items are computed. |
| $T$ | Set of items similarity between which is computed. |
| $n$ | Number of items, $n = \|T\|$. |
| $C$ | Set of contexts in which items occurrences are observed. |
| $m$ | Number of observation contexts, $m = \|C\|$. |
| $W$ | $n \times m$ graph adjacency matrix. |
| $W_{ij}$ | $(i, j)$th element of $W$, meaning the occurrence weight for $i$-th item in $j$-th context. |
| $W_{i:}$ | Row vector containing the $i$th row of $W$, meaning the occurrence weights of all contexts for item $i$. |
| $Q$ | $n \times n$ diagonal degree matrix with $Q_{ii} = \sum_k W_{ik}$. |
| $D$ | $m \times m$ diagonal degree matrix with $D_{jj} = \sum_k W_{kj}$. |
| $f_W(q, u)$ | Similarity of item $u$ to query item $q$ computed via function $f$ based on weight matrix $W$. |
| $\Gamma(u)$ | Set of graph nodes connected to node $u$. |

Table 2: Table of symbols.

**Jaccard (JAC)** coefficient measures the similarity of two sets as the size of their intersection scaled by the size of the union, ignoring the occurrence weights in contrast to Common Neighbors and cosine similarity [13]. Jaccard similarity score $JAC_W(q, u)$ of a target node $u$ to the query node $q$ based on the weight matrix $W$ is:

$$JAC_W(q, u) = \frac{|\Gamma(q) \cap \Gamma(u)|}{|\Gamma(q) \cup \Gamma(u)|}.$$

**Pointwise Mutual Information (PMI)** has been used extensively in computational linguistics and information retrieval literature for similarity calculation between terms [19], modeling them as outcomes of random variables for different contexts. PMI of two events $q$ and $u$ is defines as $log\frac{p(q,u)}{p(q)p(u)}$, where $p(q)$, $p(u)$, and $p(q, u)$ are the probabilities that the events $q$, $u$, and $(q, u)$ are observed, respectively. In the co-occurrence data setting we are considering, probability $p(u)$ for an item $u \in T$ is defined as the probability that it has been observed in a randomly selected context $c \in C$, where $C$ is the set of all contexts: $p(u) = \frac{|\Gamma(u)|}{|C|}$, where $\Gamma(u)$ is the set of contexts in which $u$ occurred. Then, PMI similarity $PMI_W(q, u)$ of a target $u$ to the query $q$ based on weight matrix $W$ is:

$$PMI_W(q, u) = log\frac{|\Gamma(q) \cap \Gamma(u)|}{|\Gamma(q)||\Gamma(u)|}|C| \propto \frac{|\Gamma(q) \cap \Gamma(u)|}{|\Gamma(q)||\Gamma(u)|}. \quad (1)$$

We use the final term of Equation (1) for the definition of the PMI similarity score.

**Adamic-Adar (AA)** method measures the similarity of two nodes by aggregating the importance score of the common contexts between them [2]. The score of a common context is reciprocal to the number of item occurrences in it, on log scale. Formally, the Adamic-Adar score $AA_W(q, u)$ of a target node $u$ to the query node $q$ based on the weight matrix $W$ is:

$$AA_W(q, u) = \sum_{c \in \Gamma(q) \cap \Gamma(u)} \frac{1}{log|\Gamma(c)|}.$$

**Forward Random Walk (FRW)** method models the similarity as the probability of a random walk that started from the query node arriving at the target node after a specified number of steps [6]. Imagine a random walk on the graph starting from the node $q$. The probability score $FRW_W(q,u)$ of the walk arriving at node $u$ in 2 steps based on the weight matrix $W$ is:

$$FRW_W(q,u) = [Q^{-1}WD^{-1}W^\top]_{qu} = \sum_c \frac{W_{qc}}{Q_{qq}}\frac{W_{uc}}{D_{cc}},$$

where both $Q$ and $D$ are the diagonal degree matrices: $Q_{ii} = \sum_k W_{ik}$, $D_{jj} = \sum_k W_{kj}$.

**Backward Random Walk (BRW)** method computes similarity as the probability of a random walk starting from target node $u$ and arriving at the query node $q$, thus traversing in reverse direction compared to the forward walk [6]. The probability $BRW_W(q,u)$ that the walk arrives at item $q$ in 2 steps based on the weight matrix $W$ is:

$$BRW_W(q,u) = [WD^{-1}W^\top Q^{-1}]_{qu} = \sum_c \frac{W_{uc}}{Q_{uu}}\frac{W_{qc}}{D_{cc}}.$$

**Mean Meeting Time (MMT)** method computes similarity based on two independent random walks that start from the query and the target nodes, respectively [9]. MMT is defined as the one-step meeting probability that the two walks arrive in a shared context: $MMT_W(q,u)$ of a target node $u$ to the query $q$ based on the weight matrix $W$ is

$$MMT_W(q,u) = [Q^{-1}W(Q^{-1}W)^\top]_{qu} = \sum_c \frac{W_{qc}}{Q_{qq}}\frac{W_{uc}}{Q_{uu}}.$$

Mean meeting time can be more generally defined as the expectation of the minimum number of steps needed for the two random walks to meet in a common node, while forward and backward random walks can be more generally considered with a number of steps that is greater than two. In this paper, we focus on the basic cases of two-step forward and backward walks, and one-step mean meeting time, leaving the analysis of multi-step walks for future work.

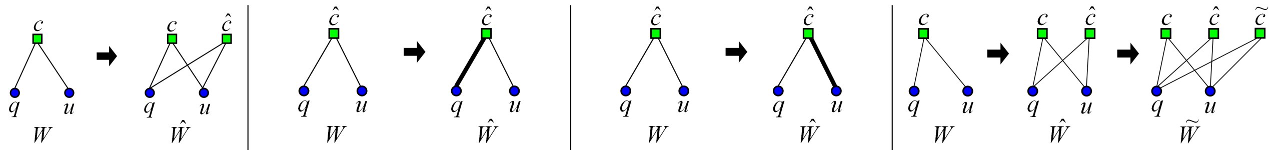
## 3 Similarity Axioms

What makes each of the method described in Section 2 suitable or unsuitable for a particular application? To answer this question, we need to define fundamental characteristics that are desirable of co-occurrence similarity functions. This section describes such characteristics and formalizes them as axioms based on the bipartite graph representation. The setting for these axioms captures the real-world phenomena underlying the continuous aggregation of new co-occurrence data: new papers being published, users issuing new queries or making purchases, articles being edited in Wikipedia. In this setting, each axiom corresponds to the effects that the arrival of new observations is expected to have on similarity function output.

There are two primary scenarios for the use of similarity functions in applications: selection of $k$ most similar items (nearest neighbors), and selection of all near neighbors whose similarity to the query item is greater than some threshold $R$ (neighborhood radius; e.g. see [3]). We will refer to these two scenarios as kNN and RNN selection[1]. The two scenarios give rise to different constraints because kNN selection is primarily concerned with the correctness of *ranking* target items, while RNN selection is primarily concerned with the accuracy of *similarity estimation* for the target items. The two corresponding types of axioms can then be formulated as *relative* constraints (concerning potential changes in ranking of a target item with respect to other target items), and *absolute* constraints (concerning potential changes in the actual similarity value for a target item).

In the following, we formally define a basic set of relative and absolute axioms which capture the intuitions of our expectation on the new observations of co-occurrence data.

[1] In practical applications, kNN and RNN selection are often hybridized.




(a) New Context Monotonicity (b) Query Co-occurrence Monotonicity (c) Target Co-occurrence Monotonicity (d) Diminishing Returns

Figure 2: (a) New Context Monotonicity (NCM): joint observations in a new shared context increase similarity. (b) Query Co-occurrence Monotonicity (QCM): new query observations in a shared context increase similarity. (c) Target Co-occurrence Monotonicity (TCM): new target observations in a shared context increase similarity. (d) Diminishing Returns (DR): incremental similarity gains due to subsequent observations decrease as items share more contexts.

### 3.1 Monotonicity

The first group of axioms encodes the intuition that new observations of query and target items in a *shared context* provide additional evidence of association between them. This additional evidence implies that the result of the observations must be an increase in similarity of the target item to the query (formulated via the absolute axioms), while the ranking of the target node cannot decrease (formulated via the relative axioms).

We first consider a scenario where a *new* context containing both the query and target items appears, corresponding to the arrival of a new context node in the bipartite graph, as shown in Figure 2 (a). Let $f_W(q, u)$ be the similarity of item $u$ to query item $q$ computed via function $f$ based on weight matrix $W$. The first axiom is defined as follows.

**Property 1** ***(New Context Monotonicity)*** *Let a new context $\hat{c}$ be observed with occurrences $\hat{W}_{q\hat{c}}$ of the query item $q$ and $\hat{W}_{u\hat{c}}$ of the target item $u$. Let $W$ be the original co-occurrence matrix, and $\hat{W}$ be the matrix after the addition of the new context. Then, a well-behaved co-occurrence similarity function must satisfy the following constraints:*

***Absolute New Context Monotonicity (A-NCM)****: $f_{\hat{W}}(q, u) > f_W(q, u)$.*

***Relative New Context Monotonicity (R-NCM)****: $f_{\hat{W}}(q, u) > f_{\hat{W}}(q, v)$, $\forall v$ s.t. $f_W(q, u) \geq f_W(q, v)$.*

The next two monotonicity axioms capture the expected response of the similarity function to the arrival of new observations in *existing* contexts where co-occurrences of target and query items were observed previously. In the bipartite graph representation, such observations correspond to an increase in the weight of an existing edge connecting items to the shared context, as shown in Figure 2 (b) and (c).

**Property 2** ***(Query Co-occurrence Monotonicity)*** *Let a new occurrence of query item $q$ with weight $\epsilon$ be observed in a context $\hat{c}$ where the target item has also been observed. Let $W$ be the original co-occurrence matrix, and $\hat{W}$ be the matrix after the new query item observation, differing from $W$ in a single element: $\hat{W}_{q\hat{c}} = W_{q\hat{c}} + \epsilon$. Then, a well-behaved co-occurrence similarity function must satisfy the following constraints:*

***Absolute Query Co-occurrence Monotonicity (A-QCM)****: $f_{\hat{W}}(q, u) > f_W(q, u)$.*

***Relative Query Co-occurrence Monotonicity (R-QCM)****: $f_{\hat{W}}(q, u) > f_{\hat{W}}(q, v)$, $\forall v$ s.t. $f_W(q, u) \geq f_W(q, v)$.*

**Property 3** ***(Target Co-occurrence Monotonicity)*** *Let a new occurrence of target item $u$ with weight $\epsilon$ be observed in a context $\hat{c}$ where the query item has also been observed. Let $W$ be the original co-occurrence matrix, and $\hat{W}$ be the matrix after the new target item observation, differing from $W$ in a single element: $\hat{W}_{u\hat{c}} = W_{u\hat{c}} + \epsilon$. Then, a well-behaved co-occurrence similarity function must satisfy the following constraints:*

***Absolute Target Co-occurrence Monotonicity (A-TCM)****: $f_{\hat{W}}(q,u) > f_W(q,u)$.*

***Relative Target Co-occurrence Monotonicity (R-TCM)****: $f_{\hat{W}}(q,u) > f_{\hat{W}}(q,v)$, $\forall v$ s.t. $f_W(q,u) \geq f_W(q,v)$.*

Overall, the monotonicity axioms guarantee that additional observations of the query and the target items in a shared context, either new or previously seen, imply a stronger degree of association between them and hence must result in a higher similarity estimate while not lowering the ranking of the target node.

### 3.2 Diminishing Returns

Next, we consider the rate of similarity increase, that is, the relative size of incremental similarity gains as more and more contexts are observed in which the query and the target items co-occur, as shown in Figure 2 (d). The diminishing returns axiom postulates that the relative impact of new co-occurrences must *decline* as more of them are observed. Intuitively, this property captures the process of continuously obtaining i.i.d. data occurrences from an underlying distribution that continues to generate new contexts (e.g., new user sessions in which queries co-occur, or new venues where researchers publish papers). Diminishing returns guarantees that the novelty of the new occurrence, conditioned on the previous occurrences of the same data, diminishes.

**Property 4** ***(Diminishing Returns (DR))*** *Let $W$ be the current weight matrix where query item $q$ and target item $u$ have been observed to co-occur. Let $\hat{W}$ be the weight matrix resulting from addition of a new context $\hat{c}$ in which $q$ and $u$ co-occur. Let $\tilde{W}$ be the weight matrix resulting from* ***subsequent*** *addition of a new context node $\tilde{c}$ in which $q$ and $u$ co-occur. Without loss of generality, assume all edges connecting $q$, $u$, $\hat{c}$ and $\tilde{c}$ have the equal weight $\theta$. Then, a well-behaved similarity function must satisfy the following constraint: $f_{\hat{W}}(q,u) - f_W(q,u) > f_{\tilde{W}}(q,u) - f_{\hat{W}}(q,u)$.*

## 4 Formal Analysis

In this section, we examine the compliance of each similarity function described in Section 2 with the axioms defined in Section 3. The analysis is simplified by introducing a *unifying* additive abstraction – *context-wise decomposition* – via which all of the considered similarity functions can be represented.

### 4.1 Unifying Framework for Similarity Functions

We observe that all the similarity functions in Section 2, although seemingly different, can be unified into our proposed abstraction called *context-wise decomposition*. Let us define the *evidence score* $e_W(c,q,u)$ to be the context $c$'s direct contribution in computing the similarity of $q$ and $u$ in a similarity function $f_W(q,u)$. Then, each of the function in Section 2 is represented by

$$f_W(q,u) = \sum_{c \in \Gamma(q) \cap \Gamma(u)} e_W(c,q,u), \tag{2}$$

| Method | Evidence $e_W(c,q,u)$ | Method | Evidence $e_W(c,q,u)$ |
|---|---|---|---|
| CN | $W_{qc}W_{uc}$ | AA | $\frac{1}{log\lvert\Gamma(c)\rvert}$ |
| COS | $\frac{W_{qc}W_{uc}}{\lvert\lvert W_{q:}\rvert\rvert_2\lvert\lvert W_{u:}\rvert\rvert_2}$ | FRW | $\frac{W_{qc}}{\sum_j W_{qj}}\frac{W_{uc}}{\sum_i W_{ic}}$ |
| JAC | $\frac{1}{\lvert\Gamma(q)\cup\Gamma(u)\rvert}$ | BRW | $\frac{W_{uc}}{\sum_j W_{uj}}\frac{W_{qc}}{\sum_i W_{ic}}$ |
| PMI | $\frac{1}{\lvert\Gamma(q)\rvert\lvert\Gamma(u)\rvert}$ | MMT | $\frac{W_{qc}}{\sum_j W_{qj}}\frac{W_{uc}}{\sum_i W_{ui}}$ |

Table 3: Evidence score functions for different similarity calculation methods. These functions are plugged into Equation (2) to define similarity methods.

meaning that the similarity score is the sum of contexts' evidence scores. Table 3 lists the evidence score functions for the similarity methods introduced in Section 2. As we will see later in this section, this unified abstraction eases the formal analysis of functions with regard to the axioms.

Table 1 summarizes the formal analysis of these methods based on the axioms, indicating which axioms are satisfied by which similarity functions unconditionally, and which do so only under certain constraints. The following subsections summarize and provide brief intuitions for each result, particularly focusing on cases whether the axioms are not satisfied under certain conditions. Full proofs are provided in the Appendix.

| Method | A-NCM | R-NCM |
|---|---|---|
| CN | Always Yes | Always Yes |
| COS | $\frac{\hat{W}_{q\hat{c}}\hat{W}_{u\hat{c}}+\sum_c W_{qc}W_{uc}}{\sqrt{\lvert\lvert W_{q:}\rvert\rvert_2^2+\hat{W}_{q\hat{c}}^2}\sqrt{\lvert\lvert W_{u:}\rvert\rvert_2^2+\hat{W}_{u\hat{c}}^2}} > f_W(q,u)$ | $\frac{\hat{W}_{q\hat{c}}\hat{W}_{u\hat{c}}+\sum_c W_{qc}W_{uc}}{\lvert\lvert W_{q:}\rvert\rvert_2^2\sqrt{\lvert\lvert W_{u:}\rvert\rvert_2^2+\hat{W}_{u\hat{c}}^2}} > f_W(q,v)$ |
| JAC | $1 > f_W(q,u)$ | Always Yes |
| PMI | $\frac{1}{\lvert\Gamma(q)\rvert+\lvert\Gamma(u)\rvert+1} > f_W(q,u)$ | $\frac{1}{\Gamma(q)} > f_W(q,u)$ |
| AA | Always Yes | Always Yes |
| FRW | $\frac{\hat{W}_{u\hat{c}}}{\hat{W}_{q\hat{c}}+\hat{W}_{u\hat{c}}} > f_W(q,u)$ | Always Yes |
| BRW | $\frac{\hat{W}_{q\hat{c}}}{\hat{W}_{q\hat{c}}+\hat{W}_{u\hat{c}}} > f_W(q,u)$ | $\frac{\hat{W}_{q\hat{c}}}{\hat{W}_{q\hat{c}}+\hat{W}_{u\hat{c}}} > f_W(q,u)$ |
| MMT | $\hat{W}_{q\hat{c}}\hat{W}_{u\hat{c}} > (\lambda-1)\sum_{c\neq\hat{c}} W_{qc}W_{uc}$ | $\hat{W}_{q\hat{c}}\hat{W}_{u\hat{c}} > \sum_{c\neq\hat{c}}(\gamma W_{qc}W_{vc} - W_{qc}W_{uc})$ |

Table 4: Summary of sufficient and necessary conditions for similarity functions to satisfy New Context Monotonicity. In MMT, $\lambda = \frac{(Q_{qq}+\hat{W}_{q\hat{c}})(Q_{uu}+\hat{W}_{u\hat{c}})}{Q_{qq}Q_{uu}}$, and $\gamma = \frac{Q_{uu}+\hat{W}_{u\hat{c}}}{Q_{vv}}$.

## 4.2 Analysis for New Context Monotonicity

In New Context Monotonicity axioms, a new context $\hat{c}$ is observed containing occurrences of target item $u$ and query item $q$. Let $f_W(q,u)$ and $f_{\hat{W}}(q,u)$ be the similarity of $u$ to $q$ before and after observing the new context, respectively, based on corresponding similarity matrices $W$ and $\hat{W}$, respectively. Analogously, let

$f_W(q,v)$ and $f_{\hat{W}}(q,v)$ be the similarities of another target item $v$ with respect to query $q$ before and after the new context observation, respectively. Then, based on Equation (2), the four scores can be written as:

$$f_W(q,u) = \sum_{c\in\Gamma(q)\cap\Gamma(u),c\neq\hat{c}} e_W(c,q,u), \tag{3}$$

$$f_{\hat{W}}(q,u) = e_{\hat{W}}(\hat{c},q,u) + \sum_{c\in\Gamma(q)\cap\Gamma(u),c\neq\hat{c}} e_{\hat{W}}(c,q,u), \tag{4}$$

$$f_W(q,v) = \sum_{c\in\Gamma(q)\cap\Gamma(v),c\neq\hat{c}} e_W(c,q,v), \tag{5}$$

$$f_{\hat{W}}(q,v) = \sum_{c\in\Gamma(q)\cap\Gamma(v),c\neq\hat{c}} e_{\hat{W}}(c,q,v). \tag{6}$$

#### 4.2.1 Absolute New Context Monotonicity (A-NCM)

We first provide a sufficient condition for A-NCM to hold.

**Lemma 1** *Similarity function $f$ satisfies A-NCM if evidence for all contexts observed before $\hat{c}$ is not changed by observation of $\hat{c}$: $e_W(c,q,u) = e_{\hat{W}}(c,q,u), \forall c \neq \hat{c}$.*

**Proof 1** *If $e_W(c,q,u) = e_{\hat{W}}(c,q,u), \forall c \neq \hat{c}$, and $e_W(\hat{c},q,u) > 0, \forall q, \forall u, \forall W$, then $f_{\hat{W}}(q,u) = e_{\hat{W}}(\hat{c},q,u) + \sum_{c\in\Gamma(q)\cap\Gamma(u),c\neq\hat{c}} e_{\hat{W}}(c,q,u) = e_{\hat{W}}(\hat{c},q,u) + f_W(q,u) > f_W(q,u)$.*

| Method | A-QCM | R-QCM |
|---|---|---|
| CN | Always Yes | Always Yes |
| COS | $\frac{\epsilon W_{u\hat{c}}+\sum_c W_{qc}W_{uc}}{\sqrt{\|\|W_{q:}\|\|_2^2+\epsilon(\epsilon+2W_{q\hat{c}})}\|\|W_{u:}\|\|_2} > f_W(q,u)$ | Always Yes |
| JAC | Always No | Always No |
| PMI | Always No | Always No |
| AA | Always No | Always No |
| FRW | $\epsilon < \frac{1}{f_W(q,u)}(W_{u\hat{c}} - \frac{W_{q\hat{c}}W_{u\hat{c}}}{D_{\hat{c}\hat{c}}}) - D_{\hat{c}\hat{c}}$ | Always Yes |
| BRW | Always Yes | Always Yes |
| MMT | $\left[\frac{W_{q\hat{c}}+\epsilon}{Q_{qq}+\epsilon} - \frac{W_{q\hat{c}}}{Q_{qq}}\right]\frac{W_{u\hat{c}}}{Q_{uu}} > \frac{\epsilon}{Q_{qq}+\epsilon}\sum_{c\neq\hat{c}}\frac{W_{qc}}{Q_{qq}}\frac{W_{uc}}{Q_{uu}}$ | $\epsilon > -\hat{W}_{q\hat{c}} + \frac{Q_{uu}}{\hat{W}_{u\hat{c}}}(\sum_{c\neq\hat{c}} W_{qc}(\frac{W_{vc}}{Q_{vv}} - \frac{W_{uc}}{Q_{uu}}))$ |

Table 5: Summary of sufficient and necessary conditions for similarity functions to satisfy Query Co-occurrence Monotonicity. $\epsilon$ is the additional edge weight as described in Axiom 2 at Section 3.

The lemma effectively states that if the addition of a new context $c$ does not affect the evidence scores for existing contexts, similarity is guaranteed to increase (as long as evidence from the new context is positive). Of the similarity functions above, only Common Neighbors and Adamic-Adar satisfy the condition in Lemma 1, and hence unconditionally satisfy the A-NCM axioms. Other methods only satisfy A-NCM conditionally, with Table 4 summarizing the conditions for each one. We note that all the conditions from Table 4 to 7 are both sufficient and necessary.

### 4.2.2 Relative New Context Monotonicity (R-NCM)

Next, we provide a sufficient condition for R-NCM to hold, in Lemma 2. It effectively states that R-NCM holds for all similarity functions which maintain the ranking of nodes after the new context addition without accounting for the additional evidence score yielded by the new context.

**Lemma 2** *If a similarity function $f$ satisfies $f_{\hat{W}}(q,u) - e_{\hat{W}}(\hat{c},q,u) \geq f_{\hat{W}}(q,v)$, and $e_{\hat{W}}(\hat{c},q,u) > 0$, then $f$ satisfies R-NCM.*

**Proof 2** $f_{\hat{W}}(q,u) = (f_{\hat{W}}(q,u) - e_{\hat{W}}(\hat{c},q,u)) + e_{\hat{W}}(\hat{c},q,u) \geq f_{\hat{W}}(q,v) + e_{\hat{W}}(\hat{c},q,u) > f_{\hat{W}}(q,v)$. *Thus,* $f_{\hat{W}}(q,u) > f_{\hat{W}}(q,v)$.

It can be checked that Common Neighbors, Adamic-Adar, and Forward Random Walk satisfy the condition in Lemma 2. Jaccard similarity does not satisfy the condition, but it can nonetheless be shown to always satisfy R-NCM. Other methods only satisfy R-NCM conditionally. Table 4 provides a summary of the conditions for the different similarity functions to satisfy NCM.

## 4.3 Analysis for Query Co-occurrence Monotonicity

Let $f_W(q,u)$ and $f_{\hat{W}}(q,u)$ be the similarity scores of target item $u$ to query item $q$ before and after the new query observation, respectively. Using the evidence score decomposition in Equation (2), the two scores can be written as:

$$f_W(q,u) = e_W(\hat{c},q,u) + \sum_{c \in \Gamma(q) \cap \Gamma(u), c \neq \hat{c}} e_W(c,q,u), \tag{7}$$

$$f_{\hat{W}}(q,u) = e_{\hat{W}}(\hat{c},q,u) + \sum_{c \in \Gamma(q) \cap \Gamma(u), c \neq \hat{c}} e_{\hat{W}}(c,q,u). \tag{8}$$

For the similarities of target $v$ to query $q$, $f_W(q,v)$ and $f_{\hat{W}}(q,v)$, we utilize the expressions in Equations (5) and (6).

### 4.3.1 Absolute Query Co-occurrence Monotonicity (A-QCM)

A sufficient condition to satisfy A-QCM is described in Lemma 3. It effectively states that when more query item occurrences are observed in a context already shared with a target item, similarity of the target item will increase as long as the corresponding evidence score increases, and the evidence scores for other shared contexts between the query and the target item don't decrease.

**Lemma 3** *If the evidence score function $e_W$ of a similarity function $f$ satisfies $e_W(\hat{c},q,u) < e_{\hat{W}}(\hat{c},q,u)$, and $\sum_{c \in C^*} e_W(c,q,u) \leq \sum_{cC^*} e_{\hat{W}}(c,q,u)$, where $C^* = \{c | c \in \Gamma(q) \cap \Gamma(u)$, and $c \neq \hat{c}\}$, then $f$ satisfies A-QCM.*

**Proof 3** *Directly follows from applying Equations* (5)-(8) *to the condition.*

It can be checked that Common Neighbors and Backward Random Walk meet the condition in Lemma 3, and thus satisfy A-QCM unconditionally. Jaccard similarity, PMI, and Adamic-Adar don't satisfy A-QCM, as they don't take into consideration the edge weights in their computation. Cosine similarity, Forward Random Walk and Mean Meeting Time satisfy A-QCM conditionally, with the conditions for each one summarized in Table 5.

| Method | A-TCM | R-TCM |
| --- | --- | --- |
| CN | Always Yes | Always Yes |
| COS | $\frac{\epsilon W_{q\hat{c}}+\sum_c W_{qc}W_{uc}}{\lVert W_{q:}\rVert_2\sqrt{\lVert W_{u:}\rVert_2^2+\epsilon(\epsilon+2W_{u\hat{c}})}} > f_W(q,u)$ | $\frac{\epsilon W_{q\hat{c}}+\sum_c W_{qc}W_{uc}}{\lVert W_{q:}\rVert_2\sqrt{\lVert W_{u:}\rVert_2^2+\epsilon(\epsilon+2W_{u\hat{c}})}} > f_W(q,u)$ |
| JAC | Always No | Always No |
| PMI | Always No | Always No |
| AA | Always No | Always No |
| FRW | Always Yes | Always Yes |
| BRW | $\epsilon < \frac{1}{f_W(q,u)}(W_{q\hat{c}} - \frac{W_{u\hat{c}}W_{q\hat{c}}}{D_{\hat{c}\hat{c}}}) - D_{\hat{c}\hat{c}}$ | $\epsilon < \frac{1}{f_W(q,u)}(W_{q\hat{c}} - \frac{W_{u\hat{c}}W_{q\hat{c}}}{D_{\hat{c}\hat{c}}}) - D_{\hat{c}\hat{c}}$ |
| MMT | $\left[\frac{W_{u\hat{c}}+\epsilon}{Q_{uu}+\epsilon} - \frac{W_{u\hat{c}}}{Q_{uu}}\right]\frac{W_{q\hat{c}}}{Q_{qq}} > \frac{\epsilon}{Q_{uu}+\epsilon}\sum_{c\neq\hat{c}}\frac{W_{uc}}{Q_{uu}}\frac{W_{qc}}{Q_{qq}}$ | $\epsilon(W_{q\hat{c}} - \sum_{c\neq\hat{c}} W_{qc}\frac{W_{vc}}{Q_{vv}}) >$ $\sum_{c\neq\hat{c}} W_{qc}(\frac{W_{vc}Q_{uu}}{Q_{vv}} - W_{uc})$ |

Table 6: Summary of sufficient and necessary conditions for similarity functions to satisfy Target Co-occurrence Monotonicity. $\epsilon$ is the additional edge weight as described in Axiom 3 at Section 3.

#### 4.3.2 Relative Query Co-occurrence Monotonicity (R-QCM)

Next, we describe a sufficient condition for the axiom R-QCM to hold, in Lemma 4.

**Lemma 4** *Let $q$ be a query node, and $u$ and $v$ are target nodes with $f_W(q,u) \geq f_W(q,v)$. If a similarity function $f$ satisfies the following condition under new appearances of $q$ in context $\hat{c}$ shared with $u$ but not with $v$, it preserves their ranking and hence satisfies axiom R-QCM:*
$\frac{f_{\hat{W}}(q,v)}{f_W(q,v)} = \frac{\sum_{c\in\Gamma(q)\cap\Gamma(u),c\neq\hat{c}} e_{\hat{W}}(c,q,u)}{\sum_{c\in\Gamma(q)\cap\Gamma(u),c\neq\hat{c}} e_W(c,q,u)} < \frac{e_{\hat{W}}(\hat{c},q,u)}{e_W(\hat{c},q,u)}$.

**Proof 4** *Let $C^* = \{c | c \in \Gamma(q)\cap\Gamma(u)$, and $c \neq \hat{c}\}$. Then,*
$f_{\hat{W}}(q,u) = e_{\hat{W}}(\hat{c},q,u) + \sum_{c\in C^*} e_{\hat{W}}(c,q,u)$
$> \frac{\sum_{c\in C^*} e_{\hat{W}}(c,q,u)}{\sum_{c\in C^*} e_W(c,q,u)} e_W(\hat{c},q,u) + \sum_{c\in C^*} e_{\hat{W}}(c,q,u)$
$= \frac{\sum_{c\in C^*} e_{\hat{W}}(c,q,u)}{\sum_{c\in C^*} e_W(c,q,u)}(e_W(\hat{c},q,u) + \sum_{c\in C^*} e_W(c,q,u))$
$\geq \frac{\sum_{c\in C^*} e_{\hat{W}}(c,q,u)}{\sum_{c\in C^*} e_W(c,q,u)} f_W(q,v) = f_{\hat{W}}(q,v)$.
*Thus, $f_{\hat{W}}(q,u) > f_{\hat{W}}(q,v)$.*

It can be checked that Common Neighbors, Cosine similarity, Forward and Backward Random Walk satisfy the condition in Lemma 4, and thus satisfy axiom R-QCM. Jaccard Similarity, PMI, and Adamic-Adar do not satisfy the axiom due to their ignorance of edge weights and hence of any new item observations in contexts where they have been seen previously. Mean Meeting Time satisfies R-QCM conditionally as summarized in Table 5.

### 4.4 Analysis of Target Co-occurrence Monotonicity

Analogously to the previous section, the evidence score decomposition in Equations (5)-(8) allows us to examine the compliance of similarity functions with the absolute and relative constraints when new observations of a target item $u$ are seen in a context $\hat{c}$ shared with the query item $q$.

### 4.4.1 Absolute Target Co-occurrence Monotonicity (A-TCM)

The sufficient condition for the axiom A-TCM can be stated analogously to that of A-QCM: when more target occurrences are observed in a context already shared with the query item, similarity of the target item will increase as long as the evidence score for $\hat{c}$ increases, and the evidence scores for other shared contexts between the query and the target item don't decrease.

**Lemma 5** *If the evidence score function $e_W$ of a similarity function $f$ satisfies $e_W(\hat{c}, q, u) < e_{\hat{W}}(\hat{c}, q, u)$, and $\sum_{c \in C^*} e_W(c, q, u) \leq \sum_{c \in C^*} e_{\hat{W}}(c, q, u)$, where $C^* = \{c | c \in \Gamma(q) \cap \Gamma(u), \text{and } c \neq \hat{c}\}$, then $f$ satisfies A-TCM.*

**Proof 5** *The proof is analogous to that of Lemma 3, and we omit it for brevity.*

It can be checked that Common Neighbors and Forward Random Walk fulfill the condition in Lemma 5, and thus satisfy axiom A-TCM. Jaccard similarity, PMI, and Adamic-Adar don't satisfy A-TCM as they ignore the edge weights and hence any additional occurrences in previous contexts. Cosine similarity, Backward Random Walk and Mean Meeting Time satisfy A-TCM conditionally as summarized in Table 6.

### 4.4.2 Relative Target Co-occurrence Monotonicity (R-TCM)

Next, we provide a sufficient condition for R-TCM, which is analogous to that of R-QCM.

**Lemma 6** *Let $q$ be a query node, and $u$ and $v$ are target nodes with $f_W(q, u) \geq f_W(q, v)$. If a similarity function $f$ satisfies the following condition under new appearances of $q$ in context $\hat{c}$ shared with $u$ but not with $v$, it preserves their ranking and hence satisfies axiom R-TCM:*
$\frac{f_{\hat{W}}(q,v)}{f_W(q,v)} = \frac{\sum_{c \in \Gamma(q) \cap \Gamma(u), c \neq \hat{c}} e_{\hat{W}}(c,q,u)}{\sum_{c \in \Gamma(q) \cap \Gamma(u), c \neq \hat{c}} e_W(c,q,u)} < \frac{e_{\hat{W}}(\hat{c},q,u)}{e_W(\hat{c},q,u)}$.

**Proof 6** *The proof is analogous to that of Lemma 4, and we omit it for brevity.*

It can be checked that Common Neighbors and Forward Random Walk satisfy the condition in Lemma 6 and thus satisfy R-TCM unconditionally. Jaccard similarity, PMI, and Adamic-Adar do not satisfy R-TCM as they ignore edge weights and hence don't change their output when additional item occurrences are observed in existing contexts. Cosine similarity, Backward Random Walk, and Mean Meeting Time satisfy R-TCM conditionally as summarized in Table 6.

## 4.5 Analysis for Diminishing Returns

Because the increase in similarity due to every subsequently observed co-occurrence context does not depend on the total number of shared contexts for Common Neighbors and Adamic-Adar similarities, they never satisfy the diminishing returns axiom, linearly increasing similarity with every subsequent shared occurrence. Other methods satisfy Diminishing Returns conditionally, as summarized in Table 7. This indicates that in data streaming domains where new contexts are continually observed (e.g., new search query sessions), Common Neighbors and Adamic-Adar are not appropriate, while other methods should be monitored to ensure that similarity values grow sublinearly and converge as data is continuously aggregated.

# 5 Random Walk with Sink

The previous section demonstrates that no similarity function under consideration satisfies all axioms unconditionally, implying that they may exhibit unintuitive, degenerate behavior. A natural question is, can we design a similarity function that satisfies all axioms? This section demonstrates how this can be achieved by a regularized variant of random-walk based similarity, which shows the benefit of our axiom based analysis.

| **Method** | **DR** |
|---|---|
| CN | Always No |
| COS | $\frac{2\theta^2+2\sum_{c\neq\hat{c},\tilde{c}} W_{qc}W_{uc}}{\sqrt{\|\|W_{q:}\|\|_2^2+\theta^2}\sqrt{\|\|W_{u:}\|\|_2^2+\theta^2}} - \frac{2\theta^2+\sum_{c\neq\hat{c},\tilde{c}} W_{qc}W_{uc}}{\sqrt{\|\|W_{q:}\|\|_2^2+2\theta^2}\sqrt{\|\|W_{u:}\|\|_2^2+2\theta^2}} > f_W(q,u)$ |
| JAC | $1 > f_W(q,u)$ |
| PMI | $\frac{1}{\|\Gamma(q)\|+\|\Gamma(u)\|+1} > f_W(q,u)$ |
| AA | Always No |
| FRW | $\theta > \frac{2D_{\tilde{c}\tilde{c}}D_{\hat{c}\hat{c}}}{3D_{\tilde{c}\tilde{c}}-D_{\hat{c}\hat{c}}}(f_W(q,u) - \frac{Q_{qq}(D_{\tilde{c}\tilde{c}}-D_{\hat{c}\hat{c}})}{2D_{\tilde{c}\tilde{c}}D_{\hat{c}\hat{c}}})$ |
| BRW | $\theta > \frac{2D_{\tilde{c}\tilde{c}}D_{\hat{c}\hat{c}}}{3D_{\tilde{c}\tilde{c}}-D_{\hat{c}\hat{c}}}(f_W(q,u) - \frac{Q_{uu}(D_{\tilde{c}\tilde{c}}-D_{\hat{c}\hat{c}})}{2D_{\tilde{c}\tilde{c}}D_{\hat{c}\hat{c}}})$ |
| MMT | $3\theta^2 + (Q_{qq}+Q_{uu})\theta > (2\theta^2 + 3(Q_{qq}+Q_{uu})\theta + Q_{qq}^2 + Q_{uu}^2 + Q_{qq}Q_{uu})f_W(q,u)$ |

Table 7: Summary of sufficient and necessary conditions for similarity functions to satisfy Diminishing Returns. $\theta$ is the weight of the new edges as described in Axiom 4.

## 5.1 Main Idea

We begin by observing from Table 1 that Absolute New Context Monotonicity (A-NCM) is surprisingly not satisfied by any random-walk based methods, while intuition suggests that observing new, exclusive co-occurrences between two items will always increase the visitation probability for walks between them. While the appearance of a new shared context adds a new visitation path with associated probability mass (evidence score), it also triggers re-normalization, which may lead to decline in total probability of reaching the destination via other co-occurrence contexts.

We propose to remedy this issue by attaching an absorbing 'sink' context to all item nodes, effectively smoothing the associated outgoing context probabilities. The weights of the edges between item nodes and the sink node are constant, regardless of the degree of the item nodes. The sink context does not contribute any evidence score to the overall similarity, but by re-distributing the probability mass (evidence) among the co-occurrences, it ensures that addition of a new co-occurrence does not remove more visitation probability than it contributes.

Formally, our proposed Random Walk with Sink (RWS) similarity is defined as the visitation probability of a forward random walk originating from the query node to visit the target node, with all item nodes being observed with the weight $s$ in an absorbing sink context $c_{sink}$. That is, $RWS_W(q,u)$ from query node $q$ to target node $u$ based on the weight matrix $W$ is:

$$RWS_W(q,u) = [Q'^{-1}WD^{-1}W^\top]_{qu} = \sum_c \frac{W_{qc}}{s+Q_{qq}}\frac{W_{uc}}{D_{cc}},$$

where $Q'$ is an adjusted diagonal degree matrix with $Q'_{ii} = s + \sum_k W_{ik}$.

Let us see how the addition of the sink node with the fixed edge weight solves the problem of FRW's conditional satisfiability of A-NCM. In RWS, the similarity increase by the addition of the new context in A-NCM is given by $\frac{\hat{W}_{q\hat{c}}}{s+Q_{qq}+\hat{W}_{q\hat{c}}}\frac{\hat{W}_{u\hat{c}}}{\hat{W}_{q\hat{c}}+\hat{W}_{u\hat{c}}}$, and the similarity decrease of existing contexts by the re-normalization is given by $\sum_{c\neq\hat{c}}\frac{W_{qc}}{s+Q_{qq}}\frac{W_{uc}}{D_{cc}} - \sum_{c\neq\hat{c}}\frac{W_{qc}}{s+Q_{qq}+\hat{W}_{q\hat{c}}}\frac{W_{uc}}{D_{cc}}$. Satisfying A-NCM requires the increase is larger than the decrease, which leads to the condition $\frac{\hat{W}_{u\hat{c}}}{\hat{W}_{q\hat{c}}+\hat{W}_{u\hat{c}}} > \sum_{c\neq\hat{c}}\frac{W_{qc}W_{uc}}{(s+Q_{qq})D_{cc}}$. Notice that by increasing $s$ enough, and thereby decreasing the value of the right hand term, the condition can be satisfied. Thus, RWS can be parameterized to satisfy A-NCM by imposing enough smoothing on the random walk similarity.

We show a working example of RWS in Figure 1 which contrasts the performance of Forward Random Walk and RWS on the DBLP bibliography dataset [2]. We compute the similarity between authors by treating venues in which they publish as co-occurrence contexts. Figure 1 (a) and (b) illustrate the effect on Forward Random Walk similarity between authors 'Jian Pei' and 'Ravi Kumar' from the addition of a new co-occurrence context 'CIKM'. The increase in visitation probability through this context is insufficient to make up for the decrease in probability mass going through the other co-occurrence contexts ('VLDB', 'SIGMOD', and 'KDD'). Figure 1 (c) and (d) show that introducing the sink node $c_{sink}$ with a sufficient smoothing level results in similarity not decreasing when a new shared co-occurrence is observed.

## 5.2 Analysis

Analysis of axiom satisfiability by RWS is performed analogously to that of Forward Random Walk, demonstrating that RWS also satisfies R-NCM, R-QCM, A-TCM, and R-TCM axioms. For axioms A-NCM, A-QCM, and DR, setting the parameter $s$ appropriately allows RWS to satisfy them, as summarized by the following results.

**Lemma 7** *RWS satisfies A-NCM if and only if* $s > -Q_{qq} + \frac{\hat{W}_{q\hat{c}}+\hat{W}_{u\hat{c}}}{\hat{W}_{u\hat{c}}} \sum_{c\neq\hat{c}} \frac{W_{qc}W_{uc}}{D_{cc}}$.

**Proof 7** *The full proof is provided in the Appendix.*

**Lemma 8** *RWS satisfies A-QCM if and only if*
$s > \frac{1}{D_{\hat{c}\hat{c}}-W_{q\hat{c}}}(W_{q\hat{c}}(D_{\hat{c}\hat{c}}+\epsilon) - Q_{qq}(D_{\hat{c}\hat{c}}-W_{q\hat{c}}) + \frac{(D_{\hat{c}\hat{c}}+\epsilon)D_{\hat{c}\hat{c}}}{W_{u\hat{c}}} \sum_{c\neq\hat{c}} \frac{W_{qc}W_{uc}}{D_{cc}})$.

**Proof 8** *The full proof is provided in the Appendix.*

**Lemma 9** *RWS satisfies DR if and only if* $s > 2\alpha - Q_{qq}$ *where* $\alpha = \sum_c W_{qc}\frac{W_{uc}}{D_{cc}}$.

**Proof 9** *The full proof is provided in the Appendix.*

Thus, the addition of the sink node enables RWS to satisfy all axioms, via a setting of the parameter $s$, as shown in Lemmas 7∼9.

We note that a variant of Backward Random Walk analogous to RWS can be designed by adding a sink with an axiom-driven parameterization. We also point out that RWS' reliance on a *constant* smoothing parameter, $s$, is a key distinction from PageRank-style smoothing of random walks, where the edge weight changes in proportion to the current degrees of the node to make the random jump probability constant. It can be shown that PageRank-style smoothing does not lead to unconditional axiom satisfaction, and hence is not an appropriate strategy for designing well-behaved similarity functions.

[2] http://www.informatik.uni-trier.de/~ley/db/

## 6 Related Work

Similarity, distance and distortion measures have been an active research topic for several decades across many areas of computer science and mathematics, and this paper focuses on their narrow subset that has high practical significance: co-occurrence-based similarity. Beyond the popular similarity functions introduced in Section 2, number of other measures were studied in the context of link prediction in social networks [11]. In recent work, Sarkar et al. performed learning-theoretic analysis of several link prediction heuristics under the assumption that they approximate a distance metric in a latent space [16]. Our approach avoids relying on metric assumptions, as it has been demonstrated in cognitive psychology literature that their key properties (minimality, symmetry, triangle inequality) are routinely violated in application domains [20, 12].

This paper's core contribution lies in developing an axiomatic approach for analyzing the capacity of various similarity methods to satisfy properties desired of them during continuous co-occurrence data aggregation. The axiomatic approach has previously been proven particularly fruitful in clustering function analysis, where its introduction by Kleinberg [10] was been followed by a number of results that study a variety of axiomatic properties for different clustering methods [4, 1]. The axiomatic approach has also been studied in the context of information retrieval where it was employed to analyze retrieval models [7, 8].

## 7 Conclusion

In this paper, we propose an axiomatic approach to analyzing co-occurrence similarity functions. The main contributions are the followings.

1. **Axiomatic framework.** We propose a principled methodology for analyzing co-occurrence similarity functions based on differential response to new observations.
2. **Analysis and proofs.** We perform extensive analysis on a number of common similarity functions using our proposed unifying abstraction, and prove that there exists no single method which satisfies all conditions unconditionally.
3. **Design of new similarity function.** We demonstrate how axiomatic analysis allows designing a new data-driven, theoretically well-justified co-occurrence similarity function without degenerate properties.

Future research directions include extending the axioms that capture important properties of other similarity functions in more general contexts.

## References


[1] Margareta Ackerman, Shai Ben-David, and David Loker. Towards property-based classification of clustering paradigms. In *NIPS*, 2010.

[2] Lada A. Adamic and E. Adar. Friends and neighbors on the web. *Social Networks*, 25(3):211–230, 2003.

[3] A. Andoni and P. Indyk. Near-optimal hashing algorithms for approximate nearest neighbor in high dimensions. *Communications of ACM*, 51(1):117–122, 2008.

[4] Shai Ben-David and Margareta Ackerman. Measures of clustering quality: A working set of axioms for clustering. In *NIPS*, 2008.

[5] Huanhuan Cao, Daxin Jiang, Jian Pei, Qi He, Zhen Liao, Enhong Chen, and Hang Li. Context-aware query suggestion by mining click-through and session data. In *KDD*, 2008.

[6] Nick Craswell and Martin Szummer. Random walks on the click graph. In *SIGIR*, 2007.

[7] Hui Fang, Tao Tao, and ChengXiang Zhai. A formal study of information retrieval heuristics. In *SIGIR*, 2004.

[8] Hui Fang and ChengXiang Zhai. An exploration of axiomatic approaches to information retrieval. In *SIGIR*, 2005.

[9] G. Jeh and J. Widom. SimRank: a measure of structural-context similarity. In *KDD*, 2002.

[10] Jon Kleinberg. An impossibility theorem for clustering. *NIPS*, 2002.

[11] David Liben-Nowell and Jon M. Kleinberg. The link-prediction problem for social networks. *JASIST*, 58(7):1019–1031, 2007.

[12] Bradley C. Love. A computational level theory of similarity. In *CogSci*, 2000.

[13] C.D. Manning, P. Raghavan, and H. Schütze. *Introduction to information retrieval*. Cambridge University Press, 2008.

[14] Q. Mei, D. Zhou, and K. Church. Query suggestion using hitting time. In *CIKM*, 2008.

[15] M.E.J. Newman. Clustering and preferential attachment in growing networks. *Physical Review E*, 64(2):025102, 2001.

[16] P. Sarkar, D. Chakrabarti, and A. W. Moore. Theoretical justification of popular link prediction heuristics. In *COLT*, 2010.

[17] Badrul Sarwar, George Karypis, Joseph Konstan, and John Reidl. Item-based collaborative filtering recommendation algorithms. In *WWW*, 2001.

[18] Jimeng Sun, Huiming Qu, Deepayan Chakrabarti, and Christos Faloutsos. Neighborhood formation and anomaly detection in bipartite graphs. In *ICDM*, 2005.

[19] P. Turney. Mining the Web for Synonyms: PMI-IR versus LSA on TOEFL. In *ECML*, 2001.

[20] Amos Tversky. Features of similarity. *Psychological Review*, 84:327–352, 1977.

# A Proofs

In this section, we give proofs of the analysis in Table 1.

## A.1 Proofs for New Context Monotonicity (NCM)

It is obvious that Common Neighbors satisfies A-NCM and R-NCM from its definition, so we omit the proofs. We prove conditions for the Cosine similarity.

**Lemma 10** *Cosine similarity satisfies A-NCM if and only if* $\frac{\hat{W}_{q\hat{c}}\hat{W}_{u\hat{c}}+\sum_c W_{qc}W_{uc}}{\sqrt{||W_{q:}||_2^2+\hat{W}_{q\hat{c}}^2}\sqrt{||W_{u:}||_2^2+\hat{W}_{u\hat{c}}^2}} > f_W(q,u)$.

**Proof 10**

$$f_{\hat{W}}(q,u) > f_W(q,u)$$
$$\frac{\hat{W}_{q\hat{c}}\hat{W}_{u\hat{c}}+\sum_c W_{qc}W_{uc}}{\sqrt{||W_{q:}||_2^2+\hat{W}_{q\hat{c}}^2}\sqrt{||W_{u:}||_2^2+\hat{W}_{u\hat{c}}^2}} > f_W(q,u)$$

□

**Lemma 11** *Cosine similarity satisfies R-NCM if and only if* $\frac{\hat{W}_{q\hat{c}}\hat{W}_{u\hat{c}}+\sum_c W_{qc}W_{uc}}{||W_{q:}||_2^2\sqrt{||W_{u:}||_2^2+\hat{W}_{u\hat{c}}^2}} > f_W(q,u)$.

**Proof 11**

$$f_{\hat{W}}(q,u) > f_{\hat{W}}(q,v)$$
$$\frac{\hat{W}_{q\hat{c}}\hat{W}_{u\hat{c}}+\sum_c W_{qc}W_{uc}}{\sqrt{||W_{q:}||_2^2+\hat{W}_{q\hat{c}}^2}\sqrt{||W_{u:}||_2^2+\hat{W}_{u\hat{c}}^2}} > \frac{\sum_c W_{qc}W_{vc}}{\sqrt{||W_{q:}||_2^2+\hat{W}_{q\hat{c}}^2}||W_{v:}||_2}$$
$$\frac{\hat{W}_{q\hat{c}}\hat{W}_{u\hat{c}}+\sum_c W_{qc}W_{uc}}{\sqrt{||W_{u:}||_2^2+\hat{W}_{u\hat{c}}^2}} > \frac{\sum_c W_{qc}W_{vc}}{||W_{v:}||_2}$$

*An equivalent condition to the last line is*

$$\frac{\hat{W}_{q\hat{c}}\hat{W}_{u\hat{c}}+\sum_c W_{qc}W_{uc}}{||W_{q:}||_2^2\sqrt{||W_{u:}||_2^2+\hat{W}_{u\hat{c}}^2}} > \frac{\sum_c W_{qc}W_{vc}}{||W_{q:}||_2||W_{v:}||_2}$$
$$\frac{\hat{W}_{q\hat{c}}\hat{W}_{u\hat{c}}+\sum_c W_{qc}W_{uc}}{||W_{q:}||_2^2\sqrt{||W_{u:}||_2^2+\hat{W}_{u\hat{c}}^2}} > f_W(q,v)$$

□

Next, we prove conditions for the Jaccard similarity.

**Lemma 12** *Jaccard similarity satisfies A-NCM if and only if* $1 > f_W(q,u)$.

**Proof 12**

$$
\begin{aligned}
f_{\hat{W}}(q,u) &> f_W(q,u) \\
\frac{|\Gamma(q)\cap\Gamma(u)|+1}{|\Gamma(q)\cup\Gamma(u)|+1} &> \frac{|\Gamma(q)\cap\Gamma(u)|}{|\Gamma(q)\cup\Gamma(u)|} \\
|\Gamma(q)\cup\Gamma(u)| &> |\Gamma(q)\cap\Gamma(u)| \\
1 &> f_W(q,u)
\end{aligned}
$$

□

**Lemma 13** *Jaccard similarity always satisfies R-NCM.*

**Proof 13** *For R-NCM to hold, the following must be true for* $\forall v$ *such that* $\frac{|\Gamma(q)\cap\Gamma(u)|}{|\Gamma(q)\cup\Gamma(u)|} \geq \frac{|\Gamma(q)\cap\Gamma(v)|}{|\Gamma(q)\cup\Gamma(v)|}$

$$
\begin{aligned}
f_{\hat{W}}(q,u) &> f_{\hat{W}}(q,v) \\
\frac{|\Gamma(q)\cap\Gamma(u)|+1}{|\Gamma(q)\cup\Gamma(u)|+1} &> \frac{|\Gamma(q)\cap\Gamma(v)|}{|\Gamma(q)\cup\Gamma(v)|+1}
\end{aligned}
$$

*which is always true since*

$$
\frac{|\Gamma(q)\cap\Gamma(u)|+1}{|\Gamma(q)\cup\Gamma(u)|+1} \geq \frac{|\Gamma(q)\cap\Gamma(u)|}{|\Gamma(q)\cup\Gamma(u)|} \geq \frac{|\Gamma(q)\cap\Gamma(v)|}{|\Gamma(q)\cup\Gamma(v)|} > \frac{|\Gamma(q)\cap\Gamma(v)|}{|\Gamma(q)\cup\Gamma(v)|+1}.
$$

□

Next, we prove conditions for the Pointwise Mutual Information.

**Lemma 14** *Pointwise Mutual Information satisfies A-NCM if and only if* $\frac{1}{|\Gamma(q)|+|\Gamma(u)|+1} > f_W(q,u)$.

**Proof 14**

$$
\begin{aligned}
f_{\hat{W}}(q,u) &> f_W(q,u) \\
\frac{|\Gamma(q)\cap\Gamma(u)|+1}{(|\Gamma(q)|+1)(|\Gamma(u)|+1)} &> \frac{|\Gamma(q)\cap\Gamma(u)|}{|\Gamma(q)||\Gamma(u)|} \\
\frac{|\Gamma(q)||\Gamma(u)|}{|\Gamma(q)|+|\Gamma(u)|+1} &> |\Gamma(q)\cap\Gamma(u)| \\
\frac{1}{|\Gamma(q)|+|\Gamma(u)|+1} &> f_W(q,u)
\end{aligned}
$$

□

**Lemma 15** *Pointwise Mutual Information satisfies R-NCM if and only if* $\frac{1}{\Gamma(q)} > f_W(q,u)$.

**Proof 15** *For R-NCM to hold, the following must be true for* $\forall v$ *such that* $f_W(q,u) \geq f_W(q,v)$

$$
\begin{aligned}
f_{\hat{W}}(q,u) &> f_{\hat{W}}(q,v) \\
\frac{|\Gamma(q) \cap \Gamma(u)| + 1}{(|\Gamma(q)| + 1)(|\Gamma(u)| + 1)} &> \frac{|\Gamma(q) \cap \Gamma(v)|}{|\Gamma(q) + 1||\Gamma(v)|} \\
|\Gamma(u)| &> |\Gamma(q) \cap \Gamma(u)| \\
\frac{1}{\Gamma(q)} &> f_W(q,u)
\end{aligned}
$$

□

Next, we prove conditions for the Forward Random Walk.

**Lemma 16** *Forward Random Walk satisfies A-NCM if and only if* $\frac{\hat{W}_{u\hat{c}}}{\hat{W}_{q\hat{c}}+\hat{W}_{u\hat{c}}} > f_W(q,u)$.

**Proof 16** *For A-NCM to hold, the following must be true:*

$$
\begin{aligned}
f_{\hat{W}}(q,u) &> f_W(q,u) \\
\frac{\hat{W}_{q\hat{c}}}{Q_{qq} + \hat{W}_{q\hat{c}}} \frac{\hat{W}_{u\hat{c}}}{\hat{W}_{q\hat{c}} + \hat{W}_{u\hat{c}}} + \sum_{c \neq \hat{c}} \frac{W_{qc}}{Q_{qq} + \hat{W}_{q\hat{c}}} \frac{W_{uc}}{D_{cc}} &> \sum_{c \neq \hat{c}} \frac{W_{qc}}{Q_{qq}} \frac{W_{uc}}{D_{cc}} \\
\frac{\hat{W}_{q\hat{c}}}{Q_{qq} + \hat{W}_{q\hat{c}}} \frac{\hat{W}_{u\hat{c}}}{\hat{W}_{q\hat{c}} + \hat{W}_{u\hat{c}}} + \sum_{c \neq \hat{c}} \frac{W_{qc} W_{uc}}{Q_{qq} D_{cc} (Q_{qq} + \hat{W}_{q\hat{c}})} (Q_{qq} - Q_{qq} - \hat{W}_{q\hat{c}}) &> 0 \\
\frac{\hat{W}_{q\hat{c}}}{Q_{qq} + \hat{W}_{q\hat{c}}} \left( \frac{\hat{W}_{u\hat{c}}}{\hat{W}_{q\hat{c}} + \hat{W}_{u\hat{c}}} - \sum_{c \neq \hat{c}} \frac{W_{qc} W_{uc}}{Q_{qq} D_{cc}} \right) &> 0 \\
\frac{\hat{W}_{u\hat{c}}}{\hat{W}_{q\hat{c}} + \hat{W}_{u\hat{c}}} &> \sum_{c \neq \hat{c}} \frac{W_{qc} W_{uc}}{Q_{qq} D_{cc}} \\
\frac{\hat{W}_{u\hat{c}}}{\hat{W}_{q\hat{c}} + \hat{W}_{u\hat{c}}} &> f_W(q,u)
\end{aligned}
$$

*which holds only when the new context occurrence weights* $\hat{W}_{q\hat{c}}$ *and* $\hat{W}_{u\hat{c}}$ *satisfy the last inequality. Thus, for the A-NCM axiom to hold for forward random walks, the target node's proportion of occurrences in the new context must be no less than the current value of similarity.* □

**Lemma 17** *Forward Random Walk satisfies R-NCM always.*

**Proof 17** *For R-NCM to hold, the following must be true for* $\forall v$ *such that* $f_W(q,u) \geq f_W(q,v)$*,*

$$
\begin{aligned}
f_{\hat{W}}(q,u) &> f_{\hat{W}}(q,v) \\
\frac{\hat{W}_{q\hat{c}}}{Q_{qq} + \hat{W}_{q\hat{c}}} \frac{\hat{W}_{u\hat{c}}}{\hat{W}_{q\hat{c}} + \hat{W}_{u\hat{c}}} + \sum_{c \neq \hat{c}} \frac{W_{qc}}{Q_{qq} + \hat{W}_{q\hat{c}}} \frac{W_{uc}}{D_{cc}} &> \sum_{c \neq \hat{c}} \frac{W_{qc}}{Q_{qq} + \hat{W}_{q\hat{c}}} \frac{W_{vc}}{D_{cc}} \\
\frac{\hat{W}_{q\hat{c}}}{Q_{qq} + \hat{W}_{q\hat{c}}} \frac{\hat{W}_{u\hat{c}}}{\hat{W}_{q\hat{c}} + \hat{W}_{u\hat{c}}} + \sum_{c \neq \hat{c}} \frac{Q_{qq}}{Q_{qq} + \hat{W}_{q\hat{c}}} \frac{W_{qc}}{Q_{qq}} \frac{W_{uc} - W_{vc}}{D_{cc}} &> 0 \\
\frac{\hat{W}_{q\hat{c}}}{Q_{qq} + \hat{W}_{q\hat{c}}} \frac{\hat{W}_{u\hat{c}}}{\hat{W}_{q\hat{c}} + \hat{W}_{u\hat{c}}} + \frac{Q_{qq}}{Q_{qq} + \hat{W}_{q\hat{c}}} (f_W(q,u) - f_W(q,v)) &> 0
\end{aligned}
$$

*which is always true as the first summand is positive and the second summand is nonnegative, hence the R-NCM axiom is always satisfied for Forward Random Walk similarity.* □

For the Backward Random Walk (BRW), it can be shown that BRW satisfies A-NCM if and only if $\frac{\hat{W}_{q\hat{c}}}{\hat{W}_{q\hat{c}}+\hat{W}_{u\hat{c}}} > f_W(q,u)$, using a derivation very similar to the A-NCM for FRW. Since the addition of a new context doesn't affect other target nodes in NCM for BRW, the condition that BRW satisfies R-NCM is exactly the same as those for A-NCM.

Next, we prove conditions for the Mean Meeting Time.

**Lemma 18** *Mean Meeting Time satisfies A-NCM if and only if* $\hat{W}_{q\hat{c}}\hat{W}_{u\hat{c}} > (\lambda - 1)\sum_{c\neq\hat{c}} W_{qc}W_{uc}$*, where* $\lambda = \frac{(Q_{qq}+\hat{W}_{q\hat{c}})(Q_{uu}+\hat{W}_{u\hat{c}})}{Q_{qq}Q_{uu}}$.

**Proof 18** *For A-NCM to hold, the following must be true:*

$$f_{\hat{W}}(q,u) > f_W(q,u)$$

$$\frac{\hat{W}_{q\hat{c}}}{Q_{qq}+\hat{W}_{q\hat{c}}}\frac{\hat{W}_{u\hat{c}}}{Q_{uu}+\hat{W}_{u\hat{c}}} + \sum_{c\neq\hat{c}}\frac{W_{qc}}{Q_{qq}+\hat{W}_{q\hat{c}}}\frac{W_{uc}}{Q_{uu}+\hat{W}_{u\hat{c}}} > \sum_{c\neq\hat{c}}\frac{W_{qc}}{Q_{qq}}\frac{W_{uc}}{Q_{uu}}$$

*Thus,*

$$\hat{W}_{q\hat{c}}\hat{W}_{u\hat{c}} > (\lambda - 1)\sum_{c\neq\hat{c}} W_{qc}W_{uc}.$$

□

**Lemma 19** *Mean Meeting Time satisfies R-NCM if and only if* $\hat{W}_{q\hat{c}}\hat{W}_{u\hat{c}} > \sum_{c\neq\hat{c}}(\gamma W_{qc}W_{vc} - W_{qc}W_{uc})$*, where* $\gamma = \frac{Q_{uu}+\hat{W}_{u\hat{c}}}{Q_{vv}}$.

**Proof 19** *For R-NCM to hold, the following must be true :*

$$f_{\hat{W}}(q,u) > f_{\hat{W}}(q,v)$$

$$\frac{\hat{W}_{q\hat{c}}}{Q_{qq}+\hat{W}_{q\hat{c}}}\frac{\hat{W}_{u\hat{c}}}{Q_{uu}+\hat{W}_{u\hat{c}}} + \sum_{c\neq\hat{c}}\frac{W_{qc}}{Q_{qq}+\hat{W}_{q\hat{c}}}\frac{W_{uc}}{Q_{uu}+\hat{W}_{u\hat{c}}} > \sum_{c\neq\hat{c}}\frac{W_{qc}}{Q_{qq}+\hat{W}_{q\hat{c}}}\frac{W_{vc}}{Q_{vv}}$$

$$\frac{\hat{W}_{q\hat{c}}\hat{W}_{u\hat{c}}}{Q_{uu}+\hat{W}_{u\hat{c}}} + \sum_{c\neq\hat{c}}\frac{W_{qc}W_{uc}}{Q_{uu}+\hat{W}_{u\hat{c}}} > \sum_{c\neq\hat{c}}\frac{W_{qc}W_{vc}}{Q_{vv}}$$

*Thus,*

$$\hat{W}_{q\hat{c}}\hat{W}_{u\hat{c}} > \sum_{c\neq\hat{c}}(\gamma W_{qc}W_{vc} - W_{qc}W_{uc})$$

. □

Next, we prove conditions for the Random Walk with Sink.

**Lemma 20** *RWS satisfies A-NCM if and only if* $s > -Q_{qq} + \frac{\hat{W}_{q\hat{c}}+\hat{W}_{u\hat{c}}}{\hat{W}_{u\hat{c}}}\sum_{c\neq\hat{c}}\frac{W_{qc}W_{uc}}{D_{cc}}$.

**Proof 20** $f_{\hat{W}}(q,u) > f_W(q,u)$
$\Leftrightarrow \frac{\hat{W}_{q\hat{c}}}{s+Q_{qq}+\hat{W}_{q\hat{c}}} \frac{\hat{W}_{u\hat{c}}}{\hat{W}_{q\hat{c}}+\hat{W}_{u\hat{c}}} + \sum_{c\neq\hat{c}} \frac{W_{qc}}{s+Q_{qq}+\hat{W}_{q\hat{c}}} \frac{W_{uc}}{D_{cc}} > \sum_{c\neq\hat{c}} \frac{W_{qc}}{s+Q_{qq}} \frac{W_{uc}}{D_{cc}}$
$\Leftrightarrow \sum_{c\neq\hat{c}} \frac{W_{qc}W_{uc}}{(s+Q_{qq})D_{cc}(s+Q_{qq}+\hat{W}_{q\hat{c}})}(s+Q_{qq}-s-Q_{qq}-\hat{W}_{q\hat{c}}) > -\frac{\hat{W}_{q\hat{c}}}{s+Q_{qq}+\hat{W}_{q\hat{c}}} \frac{\hat{W}_{u\hat{c}}}{\hat{W}_{q\hat{c}}+\hat{W}_{u\hat{c}}}$
$\Leftrightarrow \frac{\hat{W}_{q\hat{c}}}{s+Q_{qq}+\hat{W}_{q\hat{c}}}(\frac{\hat{W}_{u\hat{c}}}{\hat{W}_{q\hat{c}}+\hat{W}_{u\hat{c}}} - \sum_{c\neq\hat{c}} \frac{W_{qc}W_{uc}}{(s+Q_{qq})D_{cc}}) > 0$
$\Leftrightarrow \frac{\hat{W}_{u\hat{c}}}{\hat{W}_{q\hat{c}}+\hat{W}_{u\hat{c}}} > \sum_{c\neq\hat{c}} \frac{W_{qc}W_{uc}}{(s+Q_{qq})D_{cc}}$.

*The lemma is proved by rearranging terms in the last equation.*

## A.2 Proofs for Query Co-occurrence Monotonicity (QCM)

It is obvious that Common Neighbors satisfies A-QCM and R-QCM, so we omit the proofs. We prove conditions for the Cosine similarity.

**Lemma 21** *Cosine similarity satisfies A-QCM if and only if* $\frac{\epsilon W_{u\hat{c}}+\sum_c W_{qc}W_{uc}}{\sqrt{||W_{q:}||_2^2+\epsilon(\epsilon+2W_{q\hat{c}})}||W_{u:}||_2} > f_W(q,u)$.

**Proof 21** *For A-QCM to hold for the Cosine similarity, the following must be true:*

$$
\begin{aligned}
f_{\hat{W}}(q,u) &> f_W(q,u) \\
\frac{(W_{q\hat{c}}+\epsilon)W_{u\hat{c}} + \sum_{c\neq\hat{c}} W_{qc}W_{uc}}{\sqrt{||W_{q:}||_2^2+\epsilon(\epsilon+2W_{q\hat{c}})}||W_{u:}||_2} &> f_W(q,u) \\
\frac{\epsilon W_{u\hat{c}} + \sum_c W_{qc}W_{uc}}{\sqrt{||W_{q:}||_2^2+\epsilon(\epsilon+2W_{q\hat{c}})}||W_{u:}||_2} &> f_W(q,u)
\end{aligned}
$$

□

The Jaccard similarity, the Pointwise Mutual Information, and the Adamic-Adar similarity do not satisfy A-QCM nor R-QCM since they do not consider the edge weights.

Next, we prove conditions for the Forward Random Walk.

**Lemma 22** *Forward Random Walk satisfies A-QCM if and only if* $\epsilon < \frac{1}{f_W(q,u)}(W_{u\hat{c}} - \frac{W_{q\hat{c}}W_{u\hat{c}}}{D_{\hat{c}\hat{c}}}) - D_{\hat{c}\hat{c}}$.

**Proof 22** *For A-QCM to hold for the Forward Random Walk, the following must be true:*

$$f_{\hat{W}}(q,u) > f_W(q,u)$$
$$\frac{(W_{q\hat{c}}+\epsilon)}{Q_{qq}+\epsilon}\frac{W_{u\hat{c}}}{D_{\hat{c}\hat{c}}+\epsilon} + \sum_{c\neq\hat{c}} \frac{W_{qc}}{Q_{qq}+\epsilon}\frac{W_{uc}}{D_{cc}} > \frac{W_{q\hat{c}}}{Q_{qq}}\frac{W_{u\hat{c}}}{D_{\hat{c}\hat{c}}} + \sum_{c\neq\hat{c}} \frac{W_{qc}}{Q_{qq}}\frac{W_{uc}}{D_{cc}}$$
$$\frac{W_{u\hat{c}}(Q_{qq}D_{\hat{c}\hat{c}}W_{q\hat{c}} + Q_{qq}D_{\hat{c}\hat{c}}\epsilon - W_{q\hat{c}}(Q_{qq}D_{\hat{c}\hat{c}} + \epsilon(Q_{qq}+D_{\hat{c}\hat{c}}+\epsilon)))}{Q_{qq}D_{\hat{c}\hat{c}}(Q_{qq}+\epsilon)(D_{\hat{c}\hat{c}}+\epsilon)} + \sum_{c\neq\hat{c}} \frac{W_{uc}W_{qc}(Q_{qq}-Q_{qq}-\epsilon)}{Q_{qq}D_{cc}(Q_{qq}+\epsilon)} > 0$$
$$\frac{\epsilon}{Q_{qq}+\epsilon}\frac{W_{u\hat{c}}Q_{qq}(D_{\hat{c}\hat{c}}-W_{q\hat{c}})}{Q_{qq}D_{\hat{c}\hat{c}}(D_{\hat{c}\hat{c}}+\epsilon)} - \frac{\epsilon}{Q_{qq}+\epsilon}\frac{W_{u\hat{c}}W_{q\hat{c}}}{Q_{qq}D_{\hat{c}\hat{c}}} - \frac{\epsilon}{Q_{qq}+\epsilon}\sum_{c\neq\hat{c}}\frac{W_{qc}}{Q_{qq}}\frac{W_{uc}}{D_{cc}} > 0$$
$$\frac{W_{u\hat{c}}}{D_{\hat{c}\hat{c}}}\frac{D_{\hat{c}\hat{c}}-W_{q\hat{c}}}{D_{\hat{c}\hat{c}}+\epsilon} - f_W(q,u) > 0$$
$$\epsilon < \frac{1}{f_W(q,u)}(W_{u\hat{c}} - \frac{W_{q\hat{c}}W_{u\hat{c}}}{D_{\hat{c}\hat{c}}}) - D_{\hat{c}\hat{c}}$$

*We note that for the case where $q$ and $u$ are the only nodes occurring in context $\hat{c}$, and therefore $D_{\hat{c}\hat{c}} = W_{q\hat{c}} + W_{u\hat{c}}$, the above constraint is equivalent to:*

$$\epsilon < \frac{1}{f_W(q,u)}\frac{W_{u\hat{c}}^2}{W_{q\hat{c}}+W_{u\hat{c}}} - W_{u\hat{c}} - W_{q\hat{c}}$$

□

Next, we prove conditions for the Mean Meeting Time.

**Lemma 23** *Mean Meeting Time satisfies A-QCM if and only if* $\left[\frac{W_{q\hat{c}}+\epsilon}{Q_{qq}+\epsilon} - \frac{W_{q\hat{c}}}{Q_{qq}}\right]\frac{W_{u\hat{c}}}{Q_{uu}} > \frac{\epsilon}{Q_{qq}+\epsilon}\sum_{c\neq\hat{c}}\frac{W_{qc}}{Q_{qq}}\frac{W_{uc}}{Q_{uu}}$.

**Proof 23** *For A-QCM to hold, the following must be true:*

$$f_{\hat{W}}(q,u) > f_W(q,u)$$
$$\frac{W_{q\hat{c}}+\epsilon}{Q_{qq}+\epsilon}\frac{W_{u\hat{c}}}{Q_{uu}} + \sum_{c\neq\hat{c}}\frac{W_{qc}}{Q_{qq}+\epsilon}\frac{W_{uc}}{Q_{uu}} > \frac{W_{q\hat{c}}}{Q_{qq}}\frac{W_{u\hat{c}}}{Q_{uu}} + \sum_{c\neq\hat{c}}\frac{W_{qc}}{Q_{qq}}\frac{W_{uc}}{Q_{uu}}$$
$$\left[\frac{W_{q\hat{c}}+\epsilon}{Q_{qq}+\epsilon} - \frac{W_{q\hat{c}}}{Q_{qq}}\right]\frac{W_{u\hat{c}}}{Q_{uu}} > \frac{\epsilon}{Q_{qq}+\epsilon}\sum_{c\neq\hat{c}}\frac{W_{qc}}{Q_{qq}}\frac{W_{uc}}{Q_{uu}}$$

□

**Lemma 24** *Mean Meeting Time satisfies R-QCM if and only if* $\epsilon > -\hat{W}_{q\hat{c}} + \frac{Q_{uu}}{\hat{W}_{u\hat{c}}}(\sum_{c\neq\hat{c}} W_{qc}(\frac{W_{vc}}{Q_{vv}} - \frac{W_{uc}}{Q_{uu}}))$.

**Proof 24** *For A-QCM to hold, the following must be true:*

$$f_{\hat{W}}(q,u) > f_{\hat{W}}(q,v)$$
$$\frac{W_{q\hat{c}}+\epsilon}{Q_{qq}+\epsilon}\frac{W_{u\hat{c}}}{Q_{uu}} + \sum_{c\neq\hat{c}}\frac{W_{qc}}{Q_{qq}+\epsilon}\frac{W_{uc}}{Q_{uu}} > \sum_{c\neq\hat{c}}\frac{W_{qc}}{Q_{qq}+\epsilon}\frac{W_{vc}}{Q_{vv}}$$
$$(W_{q\hat{c}}+\epsilon)\frac{W_{u\hat{c}}}{Q_{uu}} > \sum_{c\neq\hat{c}} W_{qc}(\frac{W_{vc}}{Q_{vv}} - \frac{W_{uc}}{Q_{uu}})$$
$$\epsilon > -W_{q\hat{c}} + \frac{Q_{uu}}{W_{u\hat{c}}}(\sum_{c\neq\hat{c}} W_{qc}(\frac{W_{vc}}{Q_{vv}} - \frac{W_{uc}}{Q_{uu}}))$$

□

Next, we show that RWS satisfies A-QCM with an appropriate setting of the $s$ parameter.

**Lemma 25** *RWS satisfies A-QCM if and only if* $s > \frac{1}{D_{\hat{c}\hat{c}} - W_{q\hat{c}}}(W_{q\hat{c}}(D_{\hat{c}\hat{c}} + \epsilon) - Q_{qq}(D_{\hat{c}\hat{c}} - W_{q\hat{c}}) + \frac{(D_{\hat{c}\hat{c}}+\epsilon)D_{\hat{c}\hat{c}}}{W_{u\hat{c}}}\sum_{c\neq\hat{c}}\frac{W_{qc}W_{uc}}{D_{cc}})$ .

**Proof 25**

$$
\begin{aligned}
f_{\hat{W}}(q,u) &> f_W(q,u) \\
\frac{W_{q\hat{c}}+\epsilon}{s+Q_{qq}+\epsilon}\frac{W_{u\hat{c}}}{D_{\hat{c}\hat{c}}+\epsilon} + \sum_{c\neq\hat{c}}\frac{W_{qc}}{s+Q_{qq}+\epsilon}\frac{W_{uc}}{D_{cc}} &> \frac{W_{q\hat{c}}}{s+Q_{qq}}\frac{W_{u\hat{c}}}{D_{\hat{c}\hat{c}}} + \sum_{c\neq\hat{c}}\frac{W_{qc}}{s+Q_{qq}}\frac{W_{uc}}{D_{cc}}
\end{aligned}
$$

*The lemma is proved by rearranging terms in the last equation.* □

### A.3 Proofs for Target Co-occurrence Monotonicity (TCM)

It is obvious that Common Neighbors satisfies A-TCM and R-TCM, so we omit the proofs. We prove conditions for the Cosine similarity.

**Lemma 26** *Cosine similarity satisfies A-TCM if and only if* $\frac{\epsilon W_{q\hat{c}} + \sum_c W_{qc}W_{uc}}{||W_{q:}||_2\sqrt{||W_{u:}||_2^2 + \epsilon(\epsilon+2W_{u\hat{c}})}} > f_W(q,u)$.

**Proof 26** *For A-TCM to hold for cosine, the following must be true:*

$$
\begin{aligned}
f_{\hat{W}}(q,u) &> f_W(q,u) \\
\frac{(W_{u\hat{c}}+\epsilon)W_{q\hat{c}} + \sum_{c\neq\hat{c}} W_{qc}W_{uc}}{||W_{q:}||_2\sqrt{||W_{u:}||_2^2 + \epsilon(\epsilon+2W_{u\hat{c}})}} &> f_W(q,u) \\
\frac{\epsilon W_{q\hat{c}} + \sum_c W_{qc}W_{uc}}{||W_{q:}||_2\sqrt{||W_{u:}||_2^2 + \epsilon(\epsilon+2W_{u\hat{c}})}} &> f_W(q,u)
\end{aligned}
$$

□

Since the increase of the weight of the target edge doesn't affect other target nodes' Cosine similarities, the condition that the Cosine similarity satisfies R-TCM is exactly the same as those for A-TCM.

The Jaccard similarity, the Pointwise Mutual Information, and the Adamic-Adar similarity do not satisfy A-TCM nor R-TCM since they do not consider the edge weights.

Next, we prove conditions for the Backward Random Walk.

**Lemma 27** *Backward Random Walk satisfies A-TCM if and only if* $\epsilon < \frac{1}{f_W(q,u)}(W_{q\hat{c}} - \frac{W_{u\hat{c}}W_{q\hat{c}}}{D_{\hat{c}\hat{c}}}) - D_{\hat{c}\hat{c}}$.

**Proof 27** *For A-TCM to hold for the Backward Random Walk, the following must be true:*

$$
\begin{gathered}
f_{\hat{W}}(q,u) > f_W(q,u) \\
\frac{(W_{u\hat{c}}+\epsilon)}{Q_{uu}+\epsilon}\frac{W_{q\hat{c}}}{D_{\hat{c}\hat{c}}+\epsilon} + \sum_{c\neq\hat{c}} \frac{W_{qc}}{Q_{uu}+\epsilon}\frac{W_{uc}}{D_{cc}} > \frac{W_{u\hat{c}}}{Q_{uu}}\frac{W_{q\hat{c}}}{D_{\hat{c}\hat{c}}} + \sum_{c\neq\hat{c}} \frac{W_{qc}}{Q_{uu}}\frac{W_{uc}}{D_{cc}} \\
\frac{W_{q\hat{c}}(Q_{uu}D_{\hat{c}\hat{c}}W_{u\hat{c}} + Q_{uu}D_{\hat{c}\hat{c}}\epsilon - W_{u\hat{c}}(Q_{uu}D_{\hat{c}\hat{c}} + \epsilon(Q_{uu}+D_{\hat{c}\hat{c}}+\epsilon)))}{Q_{uu}D_{\hat{c}\hat{c}}(Q_{uu}+\epsilon)(D_{\hat{c}\hat{c}}+\epsilon)} + \sum_{c\neq\hat{c}} \frac{W_{uc}W_{qc}(Q_{uu}-Q_{uu}-\epsilon)}{Q_{uu}D_{cc}(Q_{uu}+\epsilon)} > 0 \\
\frac{\epsilon}{Q_{uu}+\epsilon}\frac{W_{q\hat{c}}Q_{uu}(D_{\hat{c}\hat{c}}-W_{u\hat{c}})}{Q_{uu}D_{\hat{c}\hat{c}}(D_{\hat{c}\hat{c}}+\epsilon)} - \frac{\epsilon}{Q_{uu}+\epsilon}\frac{W_{q\hat{c}}W_{u\hat{c}}}{Q_{uu}D_{\hat{c}\hat{c}}} - \frac{\epsilon}{Q_{uu}+\epsilon}\sum_{c\neq\hat{c}}\frac{W_{qc}}{Q_{uu}}\frac{W_{uc}}{D_{cc}} > 0 \\
\frac{W_{q\hat{c}}}{D_{\hat{c}\hat{c}}}\frac{D_{\hat{c}\hat{c}}-W_{u\hat{c}}}{D_{\hat{c}\hat{c}}+\epsilon} - f_W(q,u) > 0 \\
\epsilon < \frac{1}{f_W(q,u)}(W_{q\hat{c}} - \frac{W_{u\hat{c}}W_{q\hat{c}}}{D_{\hat{c}\hat{c}}}) - D_{\hat{c}\hat{c}}
\end{gathered}
$$

*For the case where $q$ and $u$ are the only nodes occurring in context $\hat{c}$, and therefore $D_{\hat{c}\hat{c}} = W_{u\hat{c}} + W_{q\hat{c}}$, the above constraint is equivalent to:*

$$
\epsilon < \frac{1}{f_W(q,u)}\frac{W_{q\hat{c}}^2}{W_{u\hat{c}}+W_{q\hat{c}}} - W_{q\hat{c}} - W_{u\hat{c}}
$$

□

Since the addition of a weight to the target context doesn't affect other target nodes in TCM for BRW, the condition that BRW satisfies R-TCM is exactly the same of those for A-TCM.

Next, we prove conditions for the Mean Meeting Time.

**Lemma 28** *Mean Meeting Time satisfies A-TCM if and only if* $\left[\frac{W_{u\hat{c}}+\epsilon}{Q_{uu}+\epsilon} - \frac{W_{u\hat{c}}}{Q_{uu}}\right]\frac{W_{q\hat{c}}}{Q_{qq}} > \frac{\epsilon}{Q_{uu}+\epsilon}\sum_{c\neq\hat{c}}\frac{W_{uc}}{Q_{uu}}\frac{W_{qc}}{Q_{qq}}$.

**Proof 28** *For A-TCM to hold, the following must be true:*

$$
\begin{gathered}
f_{\hat{W}}(q,u) > f_W(q,u) \\
\frac{W_{u\hat{c}}+\epsilon}{Q_{uu}+\epsilon}\frac{W_{q\hat{c}}}{Q_{qq}} + \sum_{c\neq\hat{c}}\frac{W_{uc}}{Q_{uu}+\epsilon}\frac{W_{qc}}{Q_{qq}} > \frac{W_{u\hat{c}}}{Q_{uu}}\frac{W_{q\hat{c}}}{Q_{qq}} + \sum_{c\neq\hat{c}}\frac{W_{uc}}{Q_{uu}}\frac{W_{qc}}{Q_{qq}} \\
\left[\frac{W_{u\hat{c}}+\epsilon}{Q_{uu}+\epsilon} - \frac{W_{u\hat{c}}}{Q_{uu}}\right]\frac{W_{q\hat{c}}}{Q_{qq}} > \frac{\epsilon}{Q_{uu}+\epsilon}\sum_{c\neq\hat{c}}\frac{W_{uc}}{Q_{uu}}\frac{W_{qc}}{Q_{qq}}
\end{gathered}
$$

□

**Lemma 29** *Mean Meeting Time satisfies R-TCM if and only if* $\epsilon(W_{q\hat{c}} - \sum_{c\neq\hat{c}} W_{qc}\frac{W_{vc}}{Q_{vv}}) > \sum_{c\neq\hat{c}} W_{qc}(\frac{W_{vc}Q_{uu}}{Q_{vv}} - W_{uc})$.

**Proof 29** *For R-TCM to hold, the following must be true:*

$$
\begin{aligned}
f_{\hat{W}}(q,u) &> f_{\hat{W}}(q,v) \\
\frac{W_{q\hat{c}}}{Q_{qq}}\frac{W_{u\hat{c}}+\epsilon}{Q_{uu}+\epsilon} + \sum_{c\neq\hat{c}} \frac{W_{qc}}{Q_{qq}}\frac{W_{uc}}{Q_{uu}+\epsilon} &> \sum_{c\neq\hat{c}} \frac{W_{qc}}{Q_{qq}}\frac{W_{vc}}{Q_{vv}} \\
W_{q\hat{c}}(W_{u\hat{c}}+\epsilon) - (Q_{uu}+\epsilon)\sum_{c\neq\hat{c}} W_{qc}\frac{W_{vc}}{Q_{vv}} &> -\sum_{c\neq\hat{c}} W_{qc}W_{uc} \\
\epsilon(W_{q\hat{c}} - \sum_{c\neq\hat{c}} W_{qc}\frac{W_{vc}}{Q_{vv}}) &> \sum_{c\neq\hat{c}} W_{qc}(\frac{W_{vc}Q_{uu}}{Q_{vv}} - W_{uc})
\end{aligned}
$$

□

## A.4 Proofs for Diminishing Returns (DR)

It is obvious that Common Neighbors does not satisfy the Diminishing Returns (DR). We prove conditions for the Cosine similarity.

**Lemma 30** *Cosine Similarity satisfies DR if and only if* $\frac{2\theta^2+2\sum_{c\neq\hat{c},\tilde{c}} W_{qc}W_{uc}}{\sqrt{||W_{q:}||_2^2+\theta^2}\sqrt{||W_{u:}||_2^2+\theta^2}} - \frac{2\theta^2+\sum_{c\neq\hat{c},\tilde{c}} W_{qc}W_{uc}}{\sqrt{||W_{q:}||_2^2+2\theta^2}\sqrt{||W_{u:}||_2^2+2\theta^2}} > f_W(q,u)$.

**Proof 30** *For DR to hold, the following must be true:*

$$
\begin{aligned}
f_{\hat{W}}(q,u) - f_W(q,u) &> f_{\tilde{W}}(q,u) - f_{\hat{W}}(q,u) \\
2f_{\hat{W}}(q,u) &> f_W(q,u) + f_{\tilde{W}}(q,u) \\
\frac{2\theta^2+2\sum_{c\neq\hat{c},\tilde{c}} W_{qc}W_{uc}}{\sqrt{||W_{q:}||_2^2+\theta^2}\sqrt{||W_{u:}||_2^2+\theta^2}} &> \frac{\sum_{c\neq\hat{c},\tilde{c}} W_{qc}W_{uc}}{||W_{q:}||_2||W_{u:}||_2} + \frac{2\theta^2+\sum_{c\neq\hat{c},\tilde{c}} W_{qc}W_{uc}}{\sqrt{||W_{q:}||_2^2+2\theta^2}\sqrt{||W_{u:}||_2^2+2\theta^2}} \\
\frac{2\theta^2+2\sum_{c\neq\hat{c},\tilde{c}} W_{qc}W_{uc}}{\sqrt{||W_{q:}||_2^2+\theta^2}\sqrt{||W_{u:}||_2^2+\theta^2}} - \frac{2\theta^2+\sum_{c\neq\hat{c},\tilde{c}} W_{qc}W_{uc}}{\sqrt{||W_{q:}||_2^2+2\theta^2}\sqrt{||W_{u:}||_2^2+2\theta^2}} &> \frac{\sum_{c\neq\hat{c},\tilde{c}} W_{qc}W_{uc}}{||W_{q:}||_2||W_{u:}||_2} \\
\frac{2\theta^2+2\sum_{c\neq\hat{c},\tilde{c}} W_{qc}W_{uc}}{\sqrt{||W_{q:}||_2^2+\theta^2}\sqrt{||W_{u:}||_2^2+\theta^2}} - \frac{2\theta^2+\sum_{c\neq\hat{c},\tilde{c}} W_{qc}W_{uc}}{\sqrt{||W_{q:}||_2^2+2\theta^2}\sqrt{||W_{u:}||_2^2+2\theta^2}} &> f_W(q,u)
\end{aligned}
$$

□

Next, we prove conditions for the Jaccard Similarity.

**Lemma 31** *Jaccard Similarity satisfies DR if and only if* $1 > f_W(q,u)$.

**Proof 31** *For DR to hold, the following must be true:*

$$
\begin{aligned}
\frac{|\Gamma(q)\cap\Gamma(u)|+1}{|\Gamma(q)\cup\Gamma(u)|+1} - \frac{|\Gamma(q)\cap\Gamma(u)|}{|\Gamma(q)\cup\Gamma(u)|} &> \frac{|\Gamma(q)\cap\Gamma(u)|+2}{|\Gamma(q)\cup\Gamma(u)|+2} - \frac{|\Gamma(q)\cap\Gamma(u)|+1}{|\Gamma(q)\cup\Gamma(u)|+1} \\
\frac{|\Gamma(q)\cup\Gamma(u)| - |\Gamma(q)\cap\Gamma(u)|}{|\Gamma(q)\cup\Gamma(u)|(|\Gamma(q)\cup\Gamma(u)|+1)} &> \frac{|\Gamma(q)\cup\Gamma(u)| - |\Gamma(q)\cap\Gamma(u)|}{(|\Gamma(q)\cup\Gamma(u)|+2)(|\Gamma(q)\cup\Gamma(u)|+1)}
\end{aligned}
$$

*which is true, since the denominator of the left side is smaller than the right side.* □

Next, we prove conditions for the Pointwise Mutual Information.

**Lemma 32** *Pointwise Mutual Information satisfies DR if and only if* $\frac{1}{|\Gamma(q)|+|\Gamma(u)|+1} > f_W(q,u)$.

**Proof 32** *For DR to hold, the following must be true:*

$$\frac{|\Gamma(q)\cap\Gamma(u)|+1}{(|\Gamma(q)|+1)(|\Gamma(u)|+1)} - \frac{|\Gamma(q)\cap\Gamma(u)|}{|\Gamma(q)||\Gamma(u)|} > \frac{|\Gamma(q)\cap\Gamma(u)|+2}{(|\Gamma(q)|+2)(|\Gamma(u)|+2)} - \frac{|\Gamma(q)\cap\Gamma(u)|+1}{(|\Gamma(q)|+1)(|\Gamma(u)|+1)},$$
$$\frac{|\Gamma(q)||\Gamma(u)| - |\Gamma(q)\cap\Gamma(u)|(|\Gamma(q)|+|\Gamma(u)|+1)}{(|\Gamma(q)|+1)(|\Gamma(u)|+1)|\Gamma(q)||\Gamma(u)|} >$$
$$\frac{|\Gamma(q)||\Gamma(u)| - |\Gamma(q)\cap\Gamma(u)|(|\Gamma(q)|+|\Gamma(u)|+1) - 2|\Gamma(q)\cap\Gamma(u)| - 2}{(|\Gamma(q)|+2)(|\Gamma(u)|+2)(|\Gamma(q)|+1)(|\Gamma(u)|+1)},$$
$$\frac{|\Gamma(q)||\Gamma(u)| - |\Gamma(q)\cap\Gamma(u)|(|\Gamma(q)|+|\Gamma(u)|+1)}{|\Gamma(q)||\Gamma(u)|} >$$
$$\frac{|\Gamma(q)||\Gamma(u)| - |\Gamma(q)\cap\Gamma(u)|(|\Gamma(q)|+|\Gamma(u)|+1) - 2(|\Gamma(q)\cap\Gamma(u)|+1)}{(|\Gamma(q)|+2)(|\Gamma(u)|+2)},$$

*which is true, since* $|\Gamma(q)||\Gamma(u)| < (|\Gamma(q)|+2)(|\Gamma(u)|+2)$ *and* $|\Gamma(q)\cap\Gamma(u)|+1 > 0$.

□

Next, we prove conditions for the Forward Random Walk.

**Lemma 33** *Forward Random Walk satisfies DR if and only if* $\theta > \frac{2D_{\tilde{c}\tilde{c}}D_{\hat{c}\hat{c}}}{3D_{\tilde{c}\tilde{c}}-D_{\hat{c}\hat{c}}}(f_W(q,u) - \frac{Q_{qq}(D_{\tilde{c}\tilde{c}}-D_{\hat{c}\hat{c}})}{2D_{\tilde{c}\tilde{c}}D_{\hat{c}\hat{c}}})$.

**Proof 33** *For DR to hold, the following must be true:*

$$f_{\hat{W}}(q,u) - f_W(q,u) > f_{\tilde{W}}(q,u) - f_{\hat{W}}(q,u)$$
$$2f_{\hat{W}}(q,u) > f_W(q,u) + f_{\tilde{W}}(q,u)$$
$$\frac{2\theta^2}{Q_{qq}+\theta}\frac{1}{D_{\hat{c}\hat{c}}} + \sum_{c\neq\hat{c},\tilde{c}}\frac{2W_{qc}}{Q_{qq}+\theta}\frac{W_{uc}}{D_{cc}} > \frac{\theta^2}{Q_{qq}+2\theta}\frac{1}{D_{\tilde{c}\tilde{c}}} + \frac{\theta^2}{Q_{qq}+2\theta}\frac{1}{D_{\hat{c}\hat{c}}} + \sum_{c\neq\hat{c},\tilde{c}}\frac{W_{qc}W_{uc}}{D_{cc}}(\frac{1}{Q_{qq}} + \frac{1}{Q_{qq}+2\theta})$$
$$\theta^2(\frac{2}{Q_{qq}+\theta}\frac{1}{D_{\hat{c}\hat{c}}} - \frac{1}{Q_{qq}+2\theta}\frac{1}{D_{\tilde{c}\tilde{c}}} - \frac{1}{Q_{qq}+2\theta}\frac{1}{D_{\hat{c}\hat{c}}}) > \sum_{c\neq\hat{c},\tilde{c}}\frac{W_{qc}W_{uc}}{D_{cc}}(\frac{1}{Q_{qq}} + \frac{1}{Q_{qq}+2\theta} - \frac{2}{Q_{qq}+\theta})$$
$$\theta^2\frac{(3D_{\tilde{c}\tilde{c}}-D_{\hat{c}\hat{c}})\theta + (D_{\tilde{c}\tilde{c}}-D_{\hat{c}\hat{c}})Q_{qq}}{D_{\tilde{c}\tilde{c}}D_{\hat{c}\hat{c}}(Q_{qq}+\theta)(Q_{qq}+2\theta)} > \sum_{c\neq\hat{c},\tilde{c}}\frac{W_{qc}W_{uc}}{D_{cc}}\frac{2\theta^2}{Q_{qq}(Q_{qq}+\theta)(Q_{qq}+2\theta)}$$
$$(\frac{3D_{\tilde{c}\tilde{c}}-D_{\hat{c}\hat{c}}}{2D_{\tilde{c}\tilde{c}}D_{\hat{c}\hat{c}}})\theta + \frac{Q_{qq}(D_{\tilde{c}\tilde{c}}-D_{\hat{c}\hat{c}})}{2D_{\tilde{c}\tilde{c}}D_{\hat{c}\hat{c}}} > f_W(q,u)$$
$$\theta > \frac{2D_{\tilde{c}\tilde{c}}D_{\hat{c}\hat{c}}}{3D_{\tilde{c}\tilde{c}}-D_{\hat{c}\hat{c}}}(f_W(q,u) - \frac{Q_{qq}(D_{\tilde{c}\tilde{c}}-D_{\hat{c}\hat{c}})}{2D_{\tilde{c}\tilde{c}}D_{\hat{c}\hat{c}}})$$

□

Next, we prove conditions for the Backward Random Walk.

**Lemma 34** *Backward Random Walk satisfies DR if and only if* $\theta > \frac{2D_{\tilde{c}\tilde{c}}D_{\hat{c}\hat{c}}}{3D_{\tilde{c}\tilde{c}}-D_{\hat{c}\hat{c}}}(f_W(q,u) - \frac{Q_{uu}(D_{\tilde{c}\tilde{c}}-D_{\hat{c}\hat{c}})}{2D_{\tilde{c}\tilde{c}}D_{\hat{c}\hat{c}}})$.

**Proof 34** *For DR to hold, the following must be true:*

$$
\begin{aligned}
f_{\hat{W}}(q,u) - f_W(q,u) &> f_{\tilde{W}}(q,u) - f_{\hat{W}}(q,u) \\
2f_{\hat{W}}(q,u) &> f_W(q,u) + f_{\tilde{W}}(q,u) \\
\frac{2\theta^2}{Q_{uu}+\theta}\frac{1}{D_{\hat{c}\hat{c}}} + \sum_{c\neq\hat{c},\tilde{c}} \frac{2W_{uc}}{Q_{uu}+\theta}\frac{W_{qc}}{D_{cc}} &> \frac{\theta^2}{Q_{uu}+2\theta}\frac{1}{D_{\tilde{c}\tilde{c}}} + \frac{\theta^2}{Q_{uu}+2\theta}\frac{1}{D_{\hat{c}\hat{c}}} + \sum_{c\neq\hat{c},\tilde{c}} \frac{W_{uc}W_{qc}}{D_{cc}}(\frac{1}{Q_{uu}} + \frac{1}{Q_{uu}+2\theta}) \\
\theta^2(\frac{2}{Q_{uu}+\theta}\frac{1}{D_{\hat{c}\hat{c}}} - \frac{1}{Q_{uu}+2\theta}\frac{1}{D_{\tilde{c}\tilde{c}}} - \frac{1}{Q_{uu}+2\theta}\frac{1}{D_{\hat{c}\hat{c}}}) &> \sum_{c\neq\hat{c},\tilde{c}} \frac{W_{uc}W_{qc}}{D_{cc}}(\frac{1}{Q_{uu}} + \frac{1}{Q_{uu}+2\theta} - \frac{2}{Q_{uu}+\theta}) \\
\theta^2\frac{(3D_{\tilde{c}\tilde{c}} - D_{\hat{c}\hat{c}})\theta + (D_{\tilde{c}\tilde{c}} - D_{\hat{c}\hat{c}})Q_{uu}}{D_{\tilde{c}\tilde{c}}D_{\hat{c}\hat{c}}(Q_{uu}+\theta)(Q_{uu}+2\theta)} &> \sum_{c\neq\hat{c},\tilde{c}} \frac{W_{uc}W_{qc}}{D_{cc}}\frac{2\theta^2}{Q_{uu}(Q_{uu}+\theta)(Q_{uu}+2\theta)} \\
(\frac{3D_{\tilde{c}\tilde{c}} - D_{\hat{c}\hat{c}}}{2D_{\tilde{c}\tilde{c}}D_{\hat{c}\hat{c}}})\theta + \frac{Q_{uu}(D_{\tilde{c}\tilde{c}} - D_{\hat{c}\hat{c}})}{2D_{\tilde{c}\tilde{c}}D_{\hat{c}\hat{c}}} &> f_W(q,u) \\
\theta &> \frac{2D_{\tilde{c}\tilde{c}}D_{\hat{c}\hat{c}}}{3D_{\tilde{c}\tilde{c}} - D_{\hat{c}\hat{c}}}(f_W(q,u) - \frac{Q_{uu}(D_{\tilde{c}\tilde{c}} - D_{\hat{c}\hat{c}})}{2D_{\tilde{c}\tilde{c}}D_{\hat{c}\hat{c}}})
\end{aligned}
$$

□

Next, we prove conditions for the Mean Meeting Time.

**Lemma 35** *Mean Meeting Time satisfies DR if and only if* $3\theta^2 + (Q_{qq}+Q_{uu})\theta > (2\theta^2 + 3(Q_{qq}+Q_{uu})\theta + Q_{qq}^2 + Q_{uu}^2 + Q_{qq}Q_{uu})f_W(q,u)$.

**Proof 35** *For DR to hold, the following must be true:*

$$
\begin{aligned}
f_{\hat{W}}(q,u) - f_W(q,u) &> f_{\tilde{W}}(q,u) - f_{\hat{W}}(q,u), \\
2f_{\hat{W}}(q,u) &> f_W(q,u) + f_{\tilde{W}}(q,u), \\
2(\sum_{c\neq\hat{c},\tilde{c}} \frac{W_{qc}}{Q_{qq}+\theta}\frac{W_{uc}}{Q_{uu}+\theta}) + 2\frac{\theta}{Q_{qq}+\theta}\frac{\theta}{Q_{uu}+\theta} &> \\
(\sum_{c\neq\hat{c},\tilde{c}} \frac{W_{qc}}{Q_{qq}}\frac{W_{uc}}{Q_{uu}}) + (\sum_{c\neq\hat{c},\tilde{c}} \frac{W_{qc}}{Q_{qq}+2\theta}\frac{W_{uc}}{Q_{uu}+2\theta}) + 2\frac{\theta}{Q_{qq}+2\theta}\frac{\theta}{Q_{uu}+2\theta}&, \\
\frac{2\theta^2(3\theta^2 + (Q_{qq}+Q_{uu})\theta)}{(Q_{qq}+\theta)(Q_{uu}+\theta)(Q_{qq}+2\theta)(Q_{uu}+2\theta)} &> \\
\sum_{c\neq\hat{c},\tilde{c}} W_{qc}W_{uc}(\frac{1}{Q_{qq}Q_{uu}} + \frac{1}{(Q_{qq}+2\theta)(Q_{uu}+2\theta)} - \frac{2}{(Q_{qq}+\theta)(Q_{uu}+\theta)})&, \\
3\theta^2 + (Q_{qq}+Q_{uu})\theta > \sum_{c\neq\hat{c},\tilde{c}} \frac{W_{qc}W_{uc}}{Q_{qq}Q_{uu}}(2\theta^2 + 3(Q_{qq}+Q_{uu})\theta + Q_{qq}^2 + Q_{uu}^2 + Q_{qq}Q_{uu})&, \\
3\theta^2 + (Q_{qq}+Q_{uu})\theta > (2\theta^2 + 3(Q_{qq}+Q_{uu})\theta + Q_{qq}^2 + Q_{uu}^2 + Q_{qq}Q_{uu})f_W(q,u)&.
\end{aligned}
$$

□

Finally, we show that RWS satisfies Diminishing Return(DR) axiom.

**Lemma 36** *RWS satisfies DR if and only if* $s > 2\alpha - Q_{qq}$ *where* $\alpha = \sum_c W_{qc} \frac{W_{uc}}{D_{cc}}$.

**Proof 36** *From the definition of RWS in Section 5,*

$$RWS_W(q,u) = \frac{\alpha}{s + Q_{qq}},$$

$$RWS_{\hat{W}}(q,u) = \frac{\alpha + \frac{\theta}{2}}{s + Q_{qq} + \theta},$$

$$RWS_{\tilde{W}}(q,u) = \frac{\alpha + \theta}{s + Q_{qq} + 2\theta}.$$

*DR is satisfied if and only if*

$$RWS_{\hat{W}}(q,u) - RWS_W(q,u) > RWS_{\tilde{W}}(q,u) - RWS_{\hat{W}}(q,u)$$
$$\frac{2\alpha + \theta}{s + Q_{qq} + \theta} > \frac{\alpha}{s + Q_{qq}} + \frac{\alpha + \theta}{s + Q_{qq} + 2\theta}$$

*The lemma follows immediately by substituting and rearranging terms.* □