

# An Axiomatic Approach to Link Prediction

Sara Cohen and Aviv Zohar

Rachel and Selim Benin School of Computer Science and Engineering,  
Hebrew University of Jerusalem,  
Jerusalem, Israel  
{sara, avivz}@cs.huji.ac.il

## Abstract

Link prediction functions are important tools that are used to predict the evolution of a network, to locate hidden or surprising links, and to recommend new connections that should be formed. Multiple link prediction functions have been developed in the past. However, their evaluation has mostly been based on experimental work, which has shown that the quality of a link prediction function varies significantly depending on the input domain. There is currently very little understanding of why and how a specific link prediction function works well for a particular domain. The underlying foundations of a link prediction function are often left informal—each function contains implicit assumptions about the dynamics of link formation, and about structural properties that result from these dynamics.

We draw upon the motivation used in characterizations of ranking algorithms, as well as other celebrated results from social choice, and present an axiomatic basis for link prediction. This approach seeks to deconstruct each function into basic axioms, or properties, that make explicit its underlying assumptions. Our framework uses “property templates” that can be considered as general choices made by a function designer, such as what score is assigned to a 2-vertex graph, which vertices are irrelevant to the score, how removing edges or contracting vertices affects the score, and more. Using this framework, we fully characterize four well known link prediction functions and show that they are in fact derived from different variants of a single basic set of property templates.

## 1 Introduction

Link prediction functions are widely used in a variety of fields where complex networks are naturally studied. Uses range from applications in social networks to discover hidden or surprising connections and to suggest possible “friends” to users, through collaborative filtering (Huang, Li, and Chen 2005), scientific collaboration networks, biological and chemical networks, and more.

Several link prediction functions exist, each inspired by different underlying principles of network formation and creation (Leskovec et al. 2008; Qiu, He, and Yen 2011). Such functions have mostly been evaluated empirically, and have usually been found to work well in practice on specific

data sets, while performing less favorably on others (Liben-Nowell and Kleinberg 2007). This is not surprising given that different dynamics often take part in the formation of networks from diverse domains. However, this also means that given a specific domain, there is currently no way to predict which functions are likely to work well over this domain. Some attempts have been made to train functions (Al Hasan et al. 2006) in order to automatically choose a well-performing function.

Except for a few notable exceptions (e.g. (Sarkar, Chakrabarti, and Moore 2011)), most link prediction functions do not have strong theoretical underpinnings, and it is not always clear which properties they uphold. In this work, we apply an *axiomatic approach* to the characterization of link prediction functions, in an attempt to formalize their underlying assumptions and basic components. This can be helpful in the future in determining whether a specific function would be a good choice for a specific domain (by seeing whether the underlying assumptions are compatible), and can also be useful as a starting point for defining new link prediction functions.

The axiomatic approach has been widely used in the context of social choice, including celebrated results such as Arrow’s impossibility result (Arrow 1951), and the Gibbard-Satterthwaite theorem (Gibbard 1973; Satterthwaite 1975). Axiomatic treatments of this flavor, that expose deep impossibilities, show us the limits of what voting rules and preference aggregation can achieve: certain sets of desirable properties (or axioms) cannot all be satisfied within a single voting rule. Compromises must therefore be made when choosing which rule to use.

A second flavor of axiomatization is to find characterization results for algorithms or rules (e.g., axiomatization of the Shapley value in cooperative game theory (Shapley 1953)). Such characterizations show how a set of properties uniquely determines the rule or algorithm to be used, implying again that any property not already satisfied by the algorithm will not be added without the loss of another.

A clear progression from these classic results extends to the domain of other algorithms that rank, for example, ranking the vertices in a graph in order to determine their centrality or “importance”. Such is the case with the PageRank algorithm, that was originally constructed to rank individual pages in the web-graph to aid in finding relevant search

results. An axiomatic treatment for PageRank was given in (Altman and Tennenholtz 2005). Other similar works include axiomatizations of selection from within the electorate, which is also formalized in a graph structure (Alon et al. 2011), personalized ranking systems that rank vertices differently from the perspective of the ranker (which is yet another node in the graph) (Altman and Tennenholtz 2010), tie strength in social networks (Gupte and Eliassi-Rad 2012), and trust based recommendation systems (Andersen et al. 2008). Finally, work has also been performed on axiomatic treatments of collaborative filtering (Pennock et al. 2000).

Our own work focuses, for the first time, on axiomatizations of link prediction functions. In contrast to previous work, these functions do not rank vertices, but rather pairs of vertices (potential edges), in order to ascertain which ones are likely to exist or to appear in the graph. As is standard, we use link prediction functions that provide a numerical score to each such pair. This, in turn, implies a ranking of vertices that is generated by sorting all pairs by their scores.

The main contributions of our paper are as follows:

- We present a set of axioms (properties), in the format of “templates,” allowing us to express natural and intuitive properties over specific types of graphs. Our axioms have a plug-and-play flavor, in that they can be instantiated in different manners, depending on the desired ranking properties. For example, an irrelevance property is defined using a function that trims away irrelevant portions of the graph that do not affect the score for a pair of vertices. Choosing different instantiations changes the property but maintains its primary function: to focus attention on the part of the graph that is influential and to specify which parts are not. We include a brief description of the intuition behind each axiom along with its definition.
- We show that different subsets of axioms can be used to fully characterize standard link prediction functions. Thus, while our axioms ostensibly consider limited types of graphs, they imply ranking functions over arbitrary graphs. In particular, we fully characterize four standard link prediction functions, namely, hitting time, the Katz score (Katz 1953), shortest path and reliability.

## 2 Formal Framework

**Graphs.** In the following,  $G$  denotes a simple directed graph (with no self-edges), with vertices  $V_G$  and edges  $E_G$ . We use  $\Gamma_G^i(v)$  and  $\Gamma_G^o(v)$  to denote the set of incoming and outgoing neighbors, respectively, of  $v$ . We denote  $\Gamma_G(v) = \Gamma_G^i(v) \cup \Gamma_G^o(v)$ . We say that  $v$  is a *source* if  $\Gamma_G^i(v) = \emptyset$  and a *sink* if  $\Gamma_G^o(v) = \emptyset$ .

We will sometimes also consider graphs with weights on the vertices and edges. Intuitively, a vertex weight  $\nu_G(v)$  is used to indicate properties of  $v$ , such as its importance or its likelihood of failure. An edge weight  $\varepsilon_G(u, v)$  is used to represent properties such as the closeness of the relationship between  $u$  and  $v$  or the probability of using this relationship when disseminating information.

We will say that  $G$  has *independent probabilistic* edge weights if  $0 \leq \varepsilon_G(u, v) \leq 1$  for all  $(u, v) \in E_G$ . We will say that  $G$  has *normalized probabilistic* edge weights if

(1)  $0 \leq \varepsilon_G(u, v) \leq 1$  for all  $u, v$ , and (2) for all  $u$  it holds that  $\sum_{v \in \Gamma_G^o(u)} \varepsilon_G(u, v) = 1$ . Intuitively, the former is used to allow each edge to be chosen independently of the others, while the latter models a probability distribution over outgoing edges.

**Link Prediction Functions.** *Link prediction functions* associate pairs of vertices in a graph with a number that indicates how likely it is for a link to form (or to exist) between these vertices. Some functions use lower values to infer a greater likelihood, while others use greater values. We use  $\mathbb{R}_+$  to denote the non-negative real numbers, and  $\mathbb{R}_+^\top$  to denote  $\mathbb{R}_+ \cup \{\top\}$ , where  $\top$  is a special symbol, such that for all  $c \in \mathbb{R}_+^\top$ , we have  $c + \top = \top + c = c \cdot \top = \top \cdot c = \top$ , as well as  $\min\{c, \top\} = c$  and  $\max\{c, \top\} = \top$ .

A *link prediction function*  $f$  associates every triple  $G, u, v$ , where  $u, v \in V_G$ , with a value in  $\mathbb{R}_+^\top$ . If  $f(G, u, v) \neq \top$ , we say that  $f$  is *finite* with respect to  $G, u, v$ . Otherwise, we say that  $f$  is *infinite* with respect to  $G, u, v$ . We use  $\text{fin}(f)$  to denote the (possibly infinite) set of triples for which  $f(G, u, v)$  is finite.

We demonstrate the notion of a link prediction function with several well-known examples.

**Function 2.1** (Weighted Distance). *One simple method to predict links from  $u$  to  $v$  in  $G$  is simply by examining the shortest weighted distance from  $u$  to  $v$ , denoted  $\text{wdist}(G, u, v)$ . Note that if  $v$  is not reachable from  $u$ , the value  $\top$  is returned, and also note that a lower value indicates that a link is more likely to form.*

**Function 2.2** (Reliability). *Another method for predicting links is based on the notion of network reliability. Each vertex  $v$  and each edge  $e$  is associated with an independent probability of non-failure  $\nu_G(v)$  and  $\varepsilon_G(e)$ , respectively. Note that  $G$  has independent probabilistic edge weights. Then, we can measure the likelihood of a link from  $u$  to  $v$  simply by computing the probability that a path exists from  $u$  to  $v$ , when  $u$  does not fail, i.e.,*

$$\text{rlbty}(G, u, v) = P(v \text{ is reachable from } u \mid u \text{ did not fail}).$$

*This function gives greater likelihood to link formation when it returns higher values.*

**Function 2.3** (Katz). *The Katz measure (Katz 1953) is a well-known method of measuring the closeness of a pair of vertices  $u$  and  $v$  in a graph  $G$ , and is used for link prediction.<sup>1</sup>*

$$\text{katz}^\beta(G, u, v) = \sum_{l=0}^{\infty} \beta^l \cdot |\text{Paths}^l(G, u, v)| \quad (1)$$

*where  $\text{Paths}^l(G, u, v)$  is the set of paths of length precisely  $l$  from  $u$  to  $v$  in  $G$ , and  $0 < \beta < 1$ . The parameter  $\beta$  is used to give lower weight to paths that are longer.*

<sup>1</sup>A more standard formulation of Katz has the summation starting at paths of length one. To simplify the exposition, we use the given version, with the summation starting at paths of length zero. We note, however, that it is straightforward to adapt our results to summations starting at one.

Note that  $\text{katz}^\beta$  may be undefined for pairs of vertices  $u, v$  in a graph  $G$ . In particular, this occurs when the summation does not converge, in which case we define  $\text{katz}^\beta(G, u, v) = \top$ . We note that these cases can easily be characterized, in the following fashion (Liben-Nowell and Kleinberg 2007). Let  $G$  be a graph, and  $u, v$  be vertices in  $G$ . Let  $G^{u,v}$  be the subgraph of  $G$  containing  $u, v$ , and precisely those vertices and edges that participate in some path from  $u$  to  $v$ . Let  $A^{u,v}$  be the adjacency matrix for  $G^{u,v}$  and  $I^{u,v}$  be the identity matrix with the same dimension as  $A^{u,v}$ . Then, it is well known that  $\text{katz}^\beta(G, u, v) \neq \top$  if and only if  $I^{u,v} - \beta A^{u,v}$  is invertible, i.e.,

$$\text{fin}(\text{katz}^\beta) = \{(G, u, v) \mid I^{u,v} - \beta A^{u,v} \text{ is invertible}\}.$$

Moreover, if  $(G, u, v) \in \text{fin}(\text{katz}^\beta)$ , then the matrix

$$K = (I^{u,v} - \beta A^{u,v})^{-1} \quad (2)$$

satisfies that  $K_{i,j} = \text{katz}^\beta(G, v_i, v_j)$ , where  $v_i$  and  $v_j$  are the  $i$ -th and  $j$ -th vertices in  $G^{u,v}$ .<sup>2</sup>

**Function 2.4 (Hitting Time).** Hitting time is another well known link prediction function. Intuitively, a random walk starting at vertex  $u$  iteratively moves to a neighbor chosen uniformly at random. The hitting time  $\text{hit}(G, u, v)$  from  $u$  to  $v$  is the expected number of steps required for a random walk starting at  $u$  to first reach  $v$ . A slightly more expressive version of hitting time, which we will use in this paper, considers weighted graphs with probabilities associated with edges, i.e., graphs have normalized probabilistic edge weights. A random walk now chooses a neighbor  $w$  of  $u$  at random, with probability  $\varepsilon_G(u, w)$ .

We say that  $v$  is strongly reachable from  $u$  in  $G$ , written  $u \rightsquigarrow_G v$ , if for every node  $w$  that is reachable from  $u$  without traversing  $v$ , it holds that  $v$  is reachable from  $w$ . Intuitively, if  $u \rightsquigarrow_G v$ , then every path starting at  $u$  can eventually reach  $v$ .

It is immediately obvious that

- If  $u \not\rightsquigarrow_G v$ , then the hitting time from  $u$  to  $v$  is not finite, written  $\text{hit}(G, u, v) = \top$ .
- Otherwise,  $\text{hit}(G, u, v)$  is

$$\begin{cases} 0 & u = v \\ 1 + \sum_{w \in \Gamma_G^o(u)} \varepsilon_G(u, w) \cdot \text{hit}(G, w, v) & u \neq v \end{cases}$$

### 3 Link Prediction Axioms

In this paper, we will focus on *axioms* (also called *properties*) that link prediction functions may satisfy. These axioms typically consider special cases of graphs, and specify how to compute a link prediction function based on a reduction to a smaller (or simpler) graph. Interestingly, we will show that although the axioms only explicitly deal with special cases, they can often completely characterize the behavior of a link prediction function on an arbitrary graph. This characterization by means of special cases allows a deeper

<sup>2</sup>Since the standard formation of Katz sums paths starting from length one, this formula appears in the literature as  $(I - \beta A)^{-1} - I$ .

understanding of the underlying principles of various functions, and can also be a source of inspiration for additional functions.

In the following  $f$  is a link prediction function. We will consider graphs  $G$  for which both  $\nu_G$  and  $\varepsilon_G$  are defined.

**Pair Graphs.** We first focus on graphs  $G$  containing only a pair of vertices  $u, v$ , and either a single edge from  $u$  to  $v$ , or no edge at all. We call  $G$  a *pair graph*. The Pair Graph property defines the scoring function for the simplest non-trivial case, and so imbues meaning to the edge and vertex weights, as well as to the existence of an edge.

Suppose there is an edge from  $u$  to  $v$ . In this case, it is natural to use the likelihood of taking edge  $(u, v)$ , along with the value of the vertex  $v$ , to define  $f(G, u, v)$ . On the other hand, if there is no edge in  $G$ , then there is no indication in the graph that a link should form from  $u$  to  $v$ . In this case,  $f(G, u, v)$  is given the worst possible score. We observe that the notion of a “worst” possible score can differ from function to function, as there are those for which higher scores indicate greater likelihood, while there are those who behave in the opposite manner. Therefore, the following property is parameterized by a value  $\square \in \mathbb{R}_+^\top$  representing this worst possible score.

**Property 1. PAIR GRAPH: ( $\square$ -PG).** We say that  $f$  satisfies the  $\square$ -PG property if whenever  $G$  is a pair graph,

$$f(G, u, v) = \begin{cases} \varepsilon_G(u, v) \cdot \nu_G(v) & (u, v) \in E_G \\ \square & E_G = \emptyset \end{cases}$$

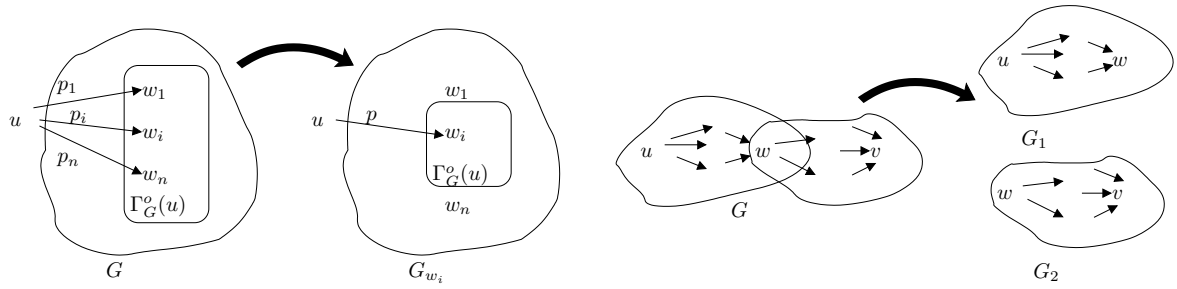
**Alternatives for Source Vertices.** We now consider the special case in which we are interested in computing  $f(G, u, v)$  when  $u$  is a source vertex. This case essentially defines the meaning of alternative edges emanating from the source vertex. Is the best score among the different routes chosen (as in the case of a message taking the shortest path to its destination)? Or is a weighted combination of scores considered (for example in case of an undirected message travelling at random)?

We will provide two different properties that are used to model functions in which edge weights are normalized, or are arbitrary.

Suppose that  $w \in \Gamma_G^o(u)$ . We use  $G_w$  to denote the graph derived by (1) removing all outgoing edges from  $u$ , other than  $(u, w)$  (2) defining  $\nu_{G_w}$  to be equal to  $\nu_G$ , and (3) defining  $\varepsilon_{G_w}$  to be equal to  $\varepsilon_G$ . We call  $G_w$  the *same probability alternative graph* for  $G, u, w$ .

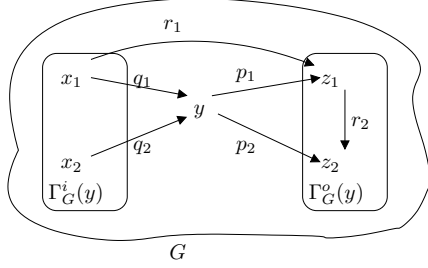
We say that  $G_w'$  is the *propagated probability alternative graph* for  $G, u, w$ , if  $G_w'$  is defined in the same manner as  $G_w$ , except that we have  $\varepsilon_{G_w'}(u, w) = 1$ . Intuitively,  $G_w'$  represents the graph in which we have “chosen” to use some edge  $(u, w)$  to reach  $v$  from  $u$ , and, hence, the probability is modified. Figure 1 (a) demonstrates these graph transformations.

The following properties express  $f(G, u, v)$  as a function of same probability or propagated probability alternative graphs. Since there are many different useful ways in which these values can be combined to derive  $f(G, u, v)$ ,

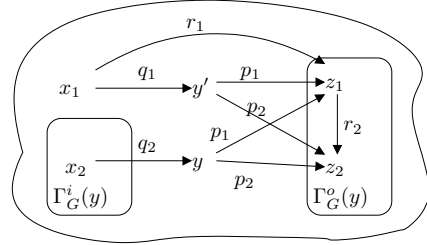


(a) Alternatives Transformation. For same probability alternative graphs,  $p = p_i$ . For propagated probability alternative graphs,  $p = 1$ .

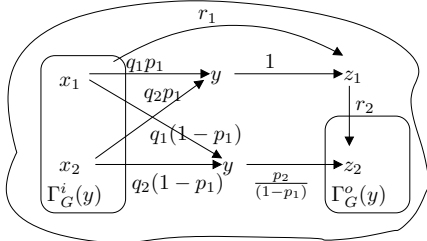
(b) Series Transformation



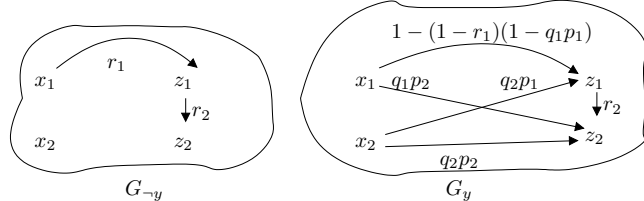
(c) Example Graph  $G$



(d) Result of In-Splt on  $(x_1, y)$ ,  $y$  and graph of (c)



(e) Result of PropOut-Splt on  $(y, z_1)$ ,  $y$  and graph of (c)



(f) Result of RemV on graph of (c) with  $y$

Figure 1: Illustration of Properties.

our properties are parameterized by an aggregate function  $\alpha$ . (An aggregate function receives a set of values and returns a single value.)

**Property 2. ALTERNATIVE GRAPHS:** ( $\alpha$ -SAMEALT,  $\alpha$ -PROPALT). We say that  $f$  satisfies the  $\alpha$ -SameAlt property if whenever  $u$  is a source vertex in  $G$ , then

$$f(G, u, v) = \alpha(\{f(G_w, u, v) \mid w \in \Gamma_G^o(v)\}),$$

where  $G_w$  is the same probability alternative graph for  $G, u, w$ .

We say that  $f$  satisfies the  $\alpha$ -PropAlt property if whenever  $u$  is a source vertex in  $G$ , then

$$f(G, u, v) = \alpha(\{\varepsilon_G(u, w) \cdot f(G_w, u, v) \mid w \in \Gamma_G^o(v)\}),$$

where  $G_w$  is the propagated probability alternative graph for  $G, u, w$ .

**Series Graphs.** Sometimes it is possible to split a graph  $G$  with vertices  $u, v$  into two graphs connected in sequence

through a single vertex. Decomposing the graph in this way reveals the effect of chaining graphs (and paths). Does the likelihood of a connection decay as distance is added, such as in the case of rumor travelling in a social network? Is the effect additive or multiplicative?

Formally, let  $G_1$  and  $G_2$  be graphs such that:

- $u$  is in  $G_1$  and  $v$  is in  $G_2$ ;
- there is a single vertex  $w$  in both  $G_1$  and  $G_2$ , and  $w$  is a sink in  $G_1$  or a source in  $G_2$ .

When  $w$  is a sink (source) in  $G_1$ , we will say that  $G_1$  and  $G_2$  are  $w$ -sink ( $w$ -source) series graphs covering  $G, u, v$ .

Since  $w$  is, in essence, a single point of division between  $u$  and  $v$ , it is natural to express  $f(G, u, v)$  as a function of  $f(G_1, u, w)$  and  $f(G_2, w, v)$ . Different such functions can be useful, and thus, the following properties are parameterized by a binary operation  $\oplus$ .

**Property 3. SINK (SOURCE) SERIES:** ( $\oplus$ -SNK,  $\oplus$ -SRC).

We say that  $f$  satisfies the  $\oplus$ -Snk property (resp. the  $\oplus$ -Src property) if whenever  $G_1$  and  $G_2$  are  $w$ -sink (resp.  $w$ -source) series graphs covering  $G$ ,  $u, v$ , then

$$f(G, u, v) = f(G_1, u, w) \oplus f(G_2, w, v).$$

This property is demonstrated in Figure 1 (b).

**Splitting Vertices.** We now consider graphs in which there is some vertex  $y$  (different from  $u$  and  $v$ ) that has multiple incoming or outgoing edges. The following properties express  $f(G, u, v)$  in terms of  $f(G', u, v)$  where  $G'$  is a graph derived from  $G$  in which  $y$  has been split into two vertices. A single incoming (resp. outgoing) edge of the original vertex is reconnected to only one copy, while the other outgoing (resp. incoming) edges are replicated and connected to both vertices.

Intuitively, this property reveals the process occurring at each vertex. Suppose for example, messages (or rumors, or infections) come in through several incoming edges. Is this message identical? Does it get sent out through all outgoing edges (as is the case with infections)? Or does it get forwarded along a single outgoing edge? We determine this by removing an edge from the vertex and preserve as much as possible of the structure. The basic operation used to derive  $G'$  is called an *incoming* or *outgoing split*, and is defined next.

We start with incoming splits. Let  $e = (x, y)$  be an edge in  $G$ . We say that  $G'$  is the  $e, y$ -split of  $G$  if the following three conditions hold. First,  $V_{G'} = V_G \cup \{y'\}$  where  $y'$  is a new vertex. Second,  $E_{G'}$  is derived from  $E_G$  by replacing edge  $e$  with  $(x, y')$  (thereby leaving  $y$  with less incoming edges than before), and adding an edge  $(y', z)$  for every  $z \in \Gamma_G^o(y)$ . Finally, if  $\pi : V_{G'} \rightarrow V_G$  is the identity on  $V_G$  and maps  $y'$  to  $y$ , then for all vertices  $w$  and edges  $e'$ , we define  $\nu_{G'}(w) = \nu_G(\pi(w))$  and  $\varepsilon_{G'}(e') = \varepsilon_G(\pi(e'))$ . Figure 1 (d) demonstrates this transformation on the graph in Figure 1 (c).

**Property 4. INCOMING SPLIT: (IN-SPLT).** We say that  $f$  satisfies the In-Splt property if whenever  $G'$  is the  $e, y$ -split of  $G$  for some  $y \neq u, y \neq v$  and incoming edge  $e$  of  $y$ , then

$$f(G, u, v) = f(G', u, v).$$

Outgoing splits are somewhat more intricate, due to the different ways in which the probability weights attached to edges can change. We consider two different options: leaving the probabilities unchanged, or propagating probabilities so that the probabilities of entire paths being taken do not change. Let  $e = (y, z)$  be an edge in  $G$ . Let  $G'$  be the graph such that (1)  $V_{G'} = V_G \cup \{y'\}$  where  $y'$  is a new vertex and (2)  $E_{G'}$  is derived from  $E_G$  by replacing edge  $e$  with  $(y', z)$  (thereby leaving  $y$  with less outgoing edges than before), and adding an edge  $(x, y')$  for every  $x \in \Gamma_G^i(y)$ . Let  $\pi : V_{G'} \rightarrow V_G$  be the identity on  $V_G$  and map  $y'$  to  $y$ .

We say that  $G'$  is the *same probability  $e, y$ -split* of  $G$  if for all vertices  $w$  and edges  $e'$ , we have  $\nu_{G'}(w) = \nu_G(\pi(w))$  and  $\varepsilon_{G'}(e') = \varepsilon_G(\pi(e'))$ . We say that  $G'$  is the *propagated probability  $e, y$ -split* of  $G$  if for all vertices  $w$  and edges  $e' = (w_1, w_2)$ , we have  $\nu_{G'}(w) = \nu_G(\pi(w))$  and

- if  $w_1 = y'$  (and thus  $w_2 = z$ ), then  $\varepsilon_{G'}(w_1, w_2) = 1$ ;
- if  $w_1 = y$ , then  $\varepsilon_{G'}(w_1, w_2) = \varepsilon_G(w_1, w_2)/(1 - \varepsilon_G(y, z))$ ;
- if  $w_2 = y'$ , then  $\varepsilon_{G'}(w_1, w_2) = \varepsilon_G(w_1, y) \cdot \varepsilon_G(y, z)$ ;
- if  $w_2 = y$ , then  $\varepsilon_{G'}(w_1, w_2) = \varepsilon_G(w_1, y) \cdot (1 - \varepsilon_G(y, z))$ ;
- otherwise  $\varepsilon_{G'}(e') = \varepsilon_G(\pi(e'))$ .

where  $e = (y, z)$ . The result of a propagated probability  $e, y$  split of the graph of Figure 1 (c) appears in Figure 1 (e).

Using these two types of outgoing splits, we define additional properties.

**Property 5. OUTGOING SPLIT: (SAMEOUT-SPLT, PROPOUT-SPLIT).** We say that  $f$  satisfies the SameOut-Splt property (resp. the PropOut-Split property) if whenever  $G'$  is the same probability (resp. propagated probability)  $e, y$ -split of  $G$  for some  $y \neq u, y \neq v$  and outgoing edge  $e$  of  $y$ , then

$$f(G, u, v) = f(G', u, v).$$

**Removing Vertices.** While the previous property considered splitting vertices, we now consider removing vertices completely from a graph. This is useful, in particular, when the weights on vertices and edges indicate their probability of failure, or some probability that they will transmit an infection further.

Let  $G$  be a graph and  $y$  be a vertex in  $G$ . We use  $G_{-y}$  to denote the graph derived by removing  $y$  and all its adjacent edges from  $G$ . We use  $G_y$  to denote the graph derived by following the same process as for  $G_{-y}$ , and then adding in the edge  $(x, z)$  for all  $x \in \Gamma_G^i(y), z \in \Gamma_G^o(y)$  with  $\varepsilon_{G_y}(x, z) = \varepsilon_G(x, y) \cdot \varepsilon_G(y, z)$ . In order to remain with a simple graph, we do two final corrections to  $G_y$ . First, we remove any self edges that have been created. Second, if there are two edges  $e_1, e_2$ , with weights  $p_1, p_2$ , between the same two vertices, we only retain a single edge  $e$ , and define  $\varepsilon_{G_y}(e) = 1 - (1 - p_1) \cdot (1 - p_2)$ . Figure 1 (f) demonstrates this transformation on the graph in Figure 1 (c).

**Property 6. REMOVING VERTICES: (REMV).** We say that  $f$  satisfies the RemV property if whenever  $y \neq u, y \neq v$  the following holds:

$$f(G, u, v) = \nu_G(y) \cdot f(G_y, u, v) + (1 - \nu_G(y)) \cdot f(G_{-y}, u, v).$$

**Irrelevant Portions of Graph.** Our final property is used to state that certain portions (vertices and/or edges) are irrelevant when computing  $f(G, u, v)$ . This property is important as it can focus the attention on a smaller portion of the graph. Different link prediction functions may deem different portions of the graph as irrelevant. For example,  $\text{katz}^\beta$  can ignore edges that are not on any path from  $u$  to  $v$ , while  $\text{wdist}$  can ignore edges that are not on any simple path from  $u$  to  $v$ . Thus, the following property is parameterized by a *relevance function*  $\Upsilon(G, u, v)$  which returns a subgraph of  $G$  that should not be ignored when computing  $f(G, u, v)$ . We note that the weights on the subgraph  $\Upsilon(G, u, v)$  will be renormalized, if  $G$  has normalized probabilistic edge weights.

Every link prediction function will satisfy this property when the relevance function is chosen as the identity function on graphs. However, some link prediction functions will satisfy this property with richer functions. Furthermore, relevance functions are selected to be almost “obvious”, for example, unreachable vertices, or vertices that are not on any path between the pair we are scoring.

**Property 7. RELEVANT SUBGRAPH: ( $\Upsilon$ -REL).** We say that  $f$  satisfies the  $\Upsilon$ -Rel property if for all  $G, u, v$ , it holds that  $f(G, u, v) = f(\Upsilon(G, u, v), u, v)$ .

## 4 Characterizing Link Prediction Functions

We now show that the axioms presented can be used to uniquely characterize several standard link prediction functions. For this we will need to instantiate the property parameters with a variety of values. We use *min* and *max* to denote the aggregation function minimum and maximum, respectively. We also define the following relevance functions:

- $\Upsilon_{\text{path}}(G, u, v)$ : Returns the subgraph of  $G$  containing  $u, v$  and all vertices and edges that participate in a path from  $u$  to  $v$ ;
- $\Upsilon_{\text{spath}}(G, u, v)$ : Returns the subgraph of  $G$  containing  $u, v$  and all vertices and edges that participate in a *simple* path from  $u$  to  $v$ ;
- $\Upsilon_{\text{fpath}}(G, u, v)$ : Returns the subgraph of  $G$  containing  $u, v$  and all vertices and edges that participate in a path from  $u$  to  $v$  that does not contain  $v$  more than once.

Due to lack of space, proofs are omitted.

**Theorem 4.1.** Let  $G$  be a graph with weights on edges and the constant weight function of 1 on vertices. Let  $u, v$  be vertices in  $G$ . There is a single link prediction function  $f$  satisfying the properties  $\top$ -PG, *min*-SameAlt,  $+$ -Snk and  $\Upsilon_{\text{spath}}$ -Rel, and this function is precisely  $\text{wdist}(G, u, v)$ .

We note that  $\text{wdist}$  also satisfies  $+$ -Src, In-Splt and SameOut-Splt, but these properties are not needed to characterize  $\text{wdist}$ . (An alternative characterization to the one given in Theorem 4.1 can use  $+$ -Src instead of  $+$ -Snk.)

**Theorem 4.2.** Let  $G$  be a graph with probabilities on vertices and independent probabilistic weights on edges. Let  $u, v$  be vertices in  $G$ . There is a single link prediction function  $f$  satisfying the properties 0-PG, RemV and  $\Upsilon_{\text{spath}}$ -Rel, and this function is precisely  $\text{rlbty}$ .

The function  $\text{rlbty}$  also satisfies *max*-Alt, but this property is not required for the characterization. For the special case where all probabilities are 1 or 0, we observe that  $\text{rlbty}$  becomes  $\text{rchblty}$  (reachability) which gives a score of 1 if  $v$  is reachable from  $u$ , and 0 otherwise. For this reason,  $\text{rchblty}$  satisfies the properties from Theorem 4.2 (as well as *max*-Alt). It also satisfies  $\times$ -Snk,  $\times$ -Src, In-Splt and SameOut-Splt. An alternative characterization to that given here for  $\text{rlbty}$  uses an additional axiom to reduce the graph to one with 0/1 weights, and then uses the above-mentioned axioms for  $\text{rchblty}$ . We omit further details due to lack of space.

**Theorem 4.3.** Let  $G$  be a graph with the constant weight function of 1 on vertices and the constant weight function of  $\beta$  on edges. Let  $u, v$  be vertices in  $G$ . There is a single link prediction function  $f$  satisfying the properties 0-PG,  $\sum$ -SameAlt,  $\times$ -Snk,  $\times$ -Src, In-Splt, SameOut-Splt,  $\Upsilon_{\text{path}}$ -Rel, and this function is precisely  $\text{katz}^\beta(G, u, v)$ .

**Theorem 4.4.** Let  $G$  be a graph with normalized probabilistic edge weights and the constant weight function of 1 on vertices. Let  $u, v$  be vertices in  $G$ . There is a single link prediction function  $f$  satisfying the properties  $\top$ -PG,  $\sum$ -PropAlt,  $+$ -Snk, In-Splt, PropOut-Splt,  $\Upsilon_{\text{fpath}}$ -Rel, and this function is precisely  $\text{hit}(G, u, v)$ .

Theorems 4.1–4.4 are proven in two parts. First we show that each of the link prediction functions of interest satisfies the given properties (e.g., we show that  $\text{wdist}$  satisfies  $\top$ -PG, *min*-SameAlt,  $+$ -Snk). This is usually not difficult and requires a careful analysis of the behavior of the link ranking functions on the the special graphs specified in the properties.

The second part of each proof is to show that the set of given properties uniquely defines the specific link prediction function (i.e., there exist no additional functions satisfying the set of properties). This type of proof is typically more intricate. For  $\text{wdist}$ , we prove the required by induction on the length of the longest simple path from  $u$  to  $v$ . For  $\text{rlbty}$ , we show uniqueness by induction on the number of vertices in  $G$ . The proof for  $\text{katz}^\beta$  involves applying a series of graph transformations and an analysis that formulates the score returned by link prediction functions satisfying the properties of Theorem 4.3 as a matrix equation that coincides with the definition of  $\text{katz}^\beta$ , i.e.,  $(I^{u,v} - \beta A^{u,v})^{-1}$ . Finally, to show the second part of  $\text{hit}$ , we again apply a series of graph transformations and analysis to derive a function coinciding with  $\text{hit}$ .

## 5 Conclusions and Future Work

We have presented an axiomatization framework for a class including several link prediction functions and have shown their characterizations. These axioms shed light on the underlying principles of the various functions. Thus, they can also be seen as a starting point to develop additional link prediction functions (or to enrich those that have been studied in the past).

In the future, we intend to develop new link prediction functions using instantiations of the properties in different ways, and to test their quality on a variety of domains, with different principles of network formation. We will also look for new and appealing properties, and check for impossibility results for certain combinations of properties. In addition, many other link prediction functions are known in the literature, and further axiomatic treatment for these is of interest. Additional link prediction functions to be considered include, among others, preferential attachment, triadic closure, and Adamic-Adar, as well as network-link scoring domains coming from sociology, such as balance theory.

## 6 Acknowledgements

Sara Cohen was supported by the Israel Science Foundation (Grant 1467/13) and the Ministry of Science and Technology (Grant 3-8710). Aviv Zohar was supported by the Israel Science Foundation (Grants 616/13, and 1773/13), and by the Israel Smart Grid (ISG) Consortium.

## References

- Al Hasan, M.; Chaoji, V.; Salem, S.; and Zaki, M. 2006. Link prediction using supervised learning. In *SDM06: Workshop on Link Analysis, Counter-terrorism and Security*.
- Alon, N.; Fischer, F.; Procaccia, A.; and Tennenholtz, M. 2011. Sum of us: Strategyproof selection from the selectors. In *Proceedings of the 13th Conference on Theoretical Aspects of Rationality and Knowledge*, 101–110. ACM.
- Altman, A., and Tennenholtz, M. 2005. Ranking systems: the pagerank axioms. In *Proceedings of the 6th ACM conference on Electronic commerce*, 1–8. ACM.
- Altman, A., and Tennenholtz, M. 2010. An axiomatic approach to personalized ranking systems. *Journal of the ACM (JACM)* 57(4):26.
- Andersen, R.; Borgs, C.; Chayes, J.; Feige, U.; Flaxman, A.; Kalai, A.; Mirrokni, V.; and Tennenholtz, M. 2008. Trust-based recommendation systems: an axiomatic approach. In *Proceedings of the 17th international conference on World Wide Web*, 199–208. ACM.
- Arrow, K. J. 1951. 1963. social choice and individual values. *Wiley, New York*.
- Gibbard, A. 1973. Manipulation of voting schemes: a general result. *Econometrica: journal of the Econometric Society* 587–601.
- Gupte, M., and Eliassi-Rad, T. 2012. Measuring tie strength in implicit social networks. In *Web Science 2012*, 109–118.
- Huang, Z.; Li, X.; and Chen, H. 2005. Link prediction approach to collaborative filtering. In *Proceedings of the 5th ACM/IEEE-CS joint conference on Digital libraries*, 141–142. ACM.
- Katz, L. 1953. A new status index derived from sociometric analysis. *Psychometrika* 18(1):39–43.
- Leskovec, J.; Backstrom, L.; Kumar, R.; and Tomkins, A. 2008. Microscopic evolution of social networks. In *Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '08*, 462–470. New York, NY, USA: ACM.
- Liben-Nowell, D., and Kleinberg, J. 2007. The link-prediction problem for social networks. *J. Am. Soc. Inf. Sci. Technol.* 58(7):1019–1031.
- Pennock, D. M.; Horvitz, E.; Giles, C. L.; et al. 2000. Social choice theory and recommender systems: Analysis of the axiomatic foundations of collaborative filtering. In *AAAI/IAAI*, 729–734.
- Qiu, B.; He, Q.; and Yen, J. 2011. Evolution of node behavior in link prediction. In *AAAI*.
- Sarkar, P.; Chakrabarti, D.; and Moore, A. W. 2011. Theoretical justification of popular link prediction heuristics. In *IJCAI Proceedings-International Joint Conference on Artificial Intelligence*, volume 22, 2722.
- Satterthwaite, M. A. 1975. Strategy-proofness and arrow's conditions: Existence and correspondence theorems for voting procedures and social welfare functions. *Journal of economic theory* 10(2):187–217.
- Shapley, L. S. 1953. A value for n-person games. contribution to the theory of games. *Annals of Mathematics Studies* 2:307–317.