Undefined 1 (2014) 1–5
IOS Press

Interpreting and Explaining PageRank through Argumentation Semantics

Emanuele Albini [a,b,*], Pietro Baroni [a], Antonio Rago [b] and Francesca Toni [b]

[a] *Dip.to di Ingegneria dell'Informazione, Università degli Studi di Brescia, Via Branze, 38 25123 Brescia, Italy*
E-mail: pietro.baroni@unibs.it

[b] *Dept. of Computing, Imperial College London, South Kensington Campus SW7 2AZ London, UK*
E-mail: {emanuele, a.rago, ft}@imperial.ac.uk

Abstract. In this paper we show how re-interpreting PageRank as an argumentation semantics for a bipolar argumentation framework empowers its explainability. After showing that PageRank, naively re-interpreted as an argumentation semantics for support frameworks, fails to satisfy some generally desirable properties, we propose a novel approach able to reconstruct PageRank as a gradual semantics of a suitably defined bipolar argumentation framework, while satisfying these properties. We then show how the theoretical advantages afforded by this approach also enjoy an enhanced explanatory power: we propose several types of argument-based explanations for PageRank, each of which focuses on different aspects of the algorithm and uncovers information useful for the comprehension of its results.

Keywords: PageRank, Explainability, Gradual Argumentation Semantics, Quantitative Bipolar Argumentation Frameworks

1. Introduction

In the context of search engines, a user wants to find the (web) pages that are the most relevant to a search query, potentially among millions of them. The web has an essential feature: each piece of information (page) may link to other pieces of information (through hyperlinks), and therefore the web organisation can be regarded as a directed graph, where pages correspond to nodes and links to edges. This is the idea that in 1999 inspired the revolutionary PageRank (PR) algorithm [1]: a method for computing a ranking score for every page based on the graph structure of the web. Given its conceptual simplicity and general formalisation for any kind of directed graph, PR has been applied to many other domains where entities can be evaluated on the basis of their connections to other entities, including citation networks [2], recommendation systems [3], chemistry [4], biology [5] and neuroscience [6], and has been studied from several viewpoints including an axiomatic characterisation from a social choice theory perspective [7].

Graph-based representations are also pervasive in the field of computational argumentation. In particular Dung's abstract argumentation frameworks [8] are essentially directed graphs whose nodes are arguments and edges represent attacks. Dung's seminal proposal has been subsequently extended in several directions, e.g. bipolar argumentation frameworks [9] encompass also a notion of support, while in quantitative bipolar argumentation frameworks [10] a base score is assigned to each argument. In this context, the argument graph structure is the basis of the assessment of argument acceptability according to some *argumentation semantics* [11]: in Dung's traditional approach the evaluation is qualitative, while in further developments numerical argument assessments based on *gradual semantics* have been investigated [10,12]. Given the similarity between PR and gradual argumentation semantics as formal tools producing a numerical assessment of connected entities in a graph, it appears that exploring possible cross-fertilisation opportunities between

*Corresponding Author. Email: emanuele@imperial.ac.uk

0000-0000/14/$00.00 © 2014 – IOS Press and the authors. All rights reserved

the two areas represents on its own an interesting research direction.

But drawing bridges between the two areas possesses not only theoretical yields. In fact, reconstructing PR in an argumentative perspective opens the door to the use of such a re-interpretation to generate explanations, exploiting in particular the *graphical* representation of the reasoning behind the algorithm. Explanations are crucial for the users of an algorithm such as PR: they may allow them to understand *why the algorithm gives a certain output* (e.g. attribution methods such as LIME [13] or SHAP [14]), to assess *which components of the input led to different outcomes* (e.g. contrastive explanations such as those proposed in [15]) or to identify *which changes in the input could change the output* (e.g. counterfactual explanations such as those proposed in [16]); for an overview see [17]. In particular, argumentation-based explanation techniques have been proposed for many AI methods, e.g., neural networks [18,19], scheduling [20], Bayesian networks [21] and classifiers [22], query answering [23] and recommender systems [24].

In this paper we first explore how PR [1] can be directly interpreted, from an argumentation perspective, as a gradual semantics for *support argumentation frameworks* [25] in which pages are arguments and links are supports. We then evidence some limitations of this simplistic correspondence and propose the novel approach of reconstructing PR as a semantics in suitably constructed *quantitative bipolar argumentation frameworks* (QBAFs). Finally, we show how this gradual semantics produces a *strength* value for each argument satisfying desirable theoretical properties and empowering the generation of several types of explanations for PR, that emphasise different aspects of its underlying mechanism.

In a broader perspective, the contribution of the paper is two-fold. On one hand we define a new gradual semantics for QBAFs based on PageRank. On the other hand, we support the idea of using argumentation frameworks, not only to model dialectical debates, but also to describe the mechanism underpinning graph algorithms in order to present them in a dialectical form, with the main aim of generating explanations but possibly also enabling other practical applications.

The paper is organized as follows. In Section 2 we recall some background concepts on PR. In Section 3 we detail how PR can be directly interpreted as a gradual semantics in support argumentation frameworks, showing however that, as such, it does not satisfy some desirable properties for argumentation. In Section 4 we reconstruct PR as a gradual semantics of suitable QBAFs, achieving in this way the satisfaction of the above mentioned desirable properties. In Section 5 we first show the practical limitations of explanations based on the support argumentation framework introduced in Section 3 and then introduce four types of explanations for PR based on the gradual semantics of Section 4. In Section 6, using several datasets crawled from English and Irish universities' websites and the Wikipedia dataset, we evaluate the different notions of explanations we introduced along several dimensions including size and cognitive tractability. We conclude the paper and outline lines of future work in Section 7.

This article builds upon [26] and [27]. In particular, Sections 2, 3 and 4 are adapted and revised from [26] and 5 and 6 extensively expand the preliminary results presented in [27].

2. PageRank Background

We firstly recall the PR definition from the original paper [1], using a different but equivalent notation when necessary for our purposes.

We assume a set of pages/nodes $\mathcal{P} = \{u_1, u_2, ..., u_N\}$ and a set of links between the pages $\mathcal{L} \subseteq \mathcal{P} \times \mathcal{P}$, where $(u, v) \in \mathcal{L}$ indicates that there is a link from page u to page v and we call the directed graph $\langle \mathcal{P}, \mathcal{L} \rangle$ the *web graph*. We say $N = |\mathcal{P}| > 0$ is the total number of pages, $O_u = \{v \in \mathcal{P} : (u, v) \in \mathcal{L}\}$ is the set of pages u points to and $I_u = \{v \in \mathcal{P} : (v, u) \in \mathcal{L}\}$ is the set of pages that point to u. We assume that $\forall u \in \mathcal{P}, \nexists (u, u) \in \mathcal{L}$, i.e. self-loops are ignored to prevent the manipulation of PR. We also assume that $\forall u \in \mathcal{P}, |O_u| > 0$, i.e. there are no *dangling* pages, that is, no pages without outgoing links (in practice, if such a page is found it is treated as having links towards all other pages as in [28]).

A *random surfer model* is used, which is based on the assumption that a user can either reach a page from a link in another page with probability $d \in]0; 1[$, referred to as *damping factor*, or land on a page directly with probability $1 - d$. Unless otherwise specified, we assume the value suggested in [1] of $d = 0.85$ and a uniform probability of directly landing on a page (i.e. we focus on *non-personalized PR*). In Section 7 we discuss how in future works these assumptions could be changed.

Definition 1. *[1] The* PageRank *(PR) of a set of pages is an assignment* $R : \mathcal{P} \to]0, 1]$ *to the pages which*

satisfies:

$$R(u) = (1-d)\cdot\frac{1}{N} + d\cdot\sum_{v\in I_u}\frac{R(v)}{|O_v|} \quad \forall u \in \mathcal{P}.$$

Note that R is the solution of a system of linear equations derived from Definition 1 (we refer to R as both the assignment and the vector resulting from it). Notice also that, as described in [28], R is unique and $||R||_1 = 1$, i.e. the L_1 norm of R is 1.

The aim of PR is to assign to every page a score that describes how relevant it is: the higher the score, the more important the page, since the score is intended to approximate the amount of users visiting the page. The latter is calculated through a mathematical model aiming at probabilistically estimating the number of user visits. The assumption here is therefore that the higher the number of links to (from) a page, the more it (the less each page linked by it, respectively) will be visited and hence the higher (lower, respectively) its PR score should be.

3. PageRank as a Gradual Semantics

In this section we show how PR may be interpreted directly as a gradual argumentation semantics and examine its ability to satisfy some desirable properties. First, we recall in Definition 2 some necessary formal notions from [10,29].

Definition 2. *A* Quantitative Bipolar Argumentation Framework *(QBAF) is a 4-tuple $\langle \mathcal{X}, \mathcal{R}^-, \mathcal{R}^+, \tau\rangle$, comprising:*

- *a finite set of arguments $\mathcal{X}$,*
- *a binary attack relation between arguments $\mathcal{R}^- \subseteq \mathcal{X}\times\mathcal{X}$,*
- *a binary support relation between arguments $\mathcal{R}^+ \subseteq \mathcal{X}\times\mathcal{X}$,*
- *a total function $\tau : \mathcal{X} \to \mathbb{I}$, with $\tau(\alpha)$ the* base score *of α, where $\mathbb{I}$ is a set equipped with a pre-order $\leq$ where, as usual, $a<b$ denotes $a\leq b$ and $b\not\leq a$.*

Given a QBAF, a total function $\sigma : \mathcal{X} \to \mathbb{I}$, called a gradual semantics*, may be used to assign a* strength *to each argument.*

We define an sQBAF as a QBAF such that $\mathcal{R}^- = \emptyset$. Finally, we let $\mathcal{R}^-(\alpha) = \{\beta \in \mathcal{X} : (\beta,\alpha) \in \mathcal{R}^-\}$ and $\mathcal{R}^+(\alpha) = \{\beta \in \mathcal{X} : (\beta,\alpha) \in \mathcal{R}^+\}$, and similarly $\mathfrak{R}^-(\alpha) = \{\beta \in \mathcal{X} : (\alpha,\beta) \in \mathcal{R}^-\}$ and $\mathfrak{R}^+(\alpha) = \{\beta \in \mathcal{X} : (\alpha,\beta) \in \mathcal{R}^+\}$.

A *web graph* $\langle \mathcal{P}, \mathcal{L}\rangle$ can be interpreted as an sQBAF where the pages (nodes) are arguments and the links between them (edges) are supports, as follows.

Definition 3. *Given a set of pages $\mathcal{P}$ and a set of links $\mathcal{L}$, a* PageRank Argumentation Framework *(PRAF) is an sQBAF defined as $PR = \langle \mathcal{X}, \emptyset, \mathcal{R}^+, \tau\rangle$, where:*

- *$\mathcal{X} = \mathcal{P}$ is the set of arguments corresponding to the set of pages,*
- *$\mathcal{R}^+ = \mathcal{L}$ is the set of supports corresponding to the set of links between pages,*
- *$\tau : \mathcal{X} \mapsto \mathbb{I} = [\frac{1-d}{|\mathcal{X}|}, 1]$ is the base score, defined as a constant function:*

$$\tau(\alpha) = \frac{1-d}{|\mathcal{X}|} \quad \forall \alpha \in \mathcal{X}.$$

Given Definition 1 and the notes on loops and dangling nodes in Section 2, Remark 1 can be trivially derived.

Remark 1. *Given a PRAF it always holds that:*

- *each argument has at least one outgoing link: $|\mathfrak{R}^+(\alpha)| > 0,\ \forall \alpha \in \mathcal{X}$;*
- *there are no self-supports: $\nexists(\alpha,\alpha) \in \mathcal{R}^+,\ \forall \alpha \in \mathcal{X}$.*

We then interpret PR as a gradual semantics for sQBAFs.

Definition 4. *The PageRank semantics is a gradual semantics $\sigma : \mathcal{X} \mapsto \mathbb{I}$ such that:*

$$\sigma(\alpha) = \tau(\alpha) + d\cdot\sum_{\beta\in\mathcal{R}^+(\alpha)}\frac{\sigma(\beta)}{|\mathfrak{R}^+(\beta)|} \quad \forall \alpha \in \mathcal{X}.$$

The following remark is directly derived from Definition 4.

Remark 2. *The codomain of σ is $\mathbb{I} = [\frac{1-d}{|\mathcal{X}|}, 1]$*

In order to formally assess PR as an argumentation semantics, we now review some desirable properties for argument strength, called *group properties* (GPs) in [10,29], as they imply groups of other properties. Some preliminary definitions need to be recalled first. Given a QBAF $\langle \mathcal{X}, \mathcal{R}^-, \mathcal{R}^+, \tau\rangle$ and a gradual semantics σ, for any $A \subseteq \mathcal{X}$, we refer to the multiset $\{\sigma(\beta) : \beta \in A\}$ as A_σ. Given $A, B \subseteq \mathcal{X}$, A is *strength equivalent to* B, denoted $A \overset{\sigma}{=} B$, iff $A_\sigma = B_\sigma$; A is *at least as strong as* B, denoted $A \overset{\sigma}{\geq} B$, iff there exists an injective mapping f from B to A such that $\forall \alpha \in B$,

$\sigma(f(\alpha)) \geq \sigma(\alpha)$; and A is *stronger than* B, denoted $A \overset{\sigma}{>} B$, iff $A \overset{\sigma}{\geq} B$ and $B \overset{\sigma}{\not\geq} A$.

GPs are then defined as follows (some being reformulated in more general or more specific ways wrt [10,29], where useful for our present purposes):

GP1. If $\mathcal{R}^-(\alpha)=\emptyset$ and $\mathcal{R}^+(\alpha)=\emptyset$ then $\sigma(\alpha) = \tau(\alpha)$.
GP2. If $\mathcal{R}^-(\alpha)\neq\emptyset$ and $\mathcal{R}^+(\alpha)=\emptyset$ then $\sigma(\alpha) < \tau(\alpha)$.
GP3. If $\mathcal{R}^-(\alpha)=\emptyset$ and $\mathcal{R}^+(\alpha)\neq\emptyset$ then $\sigma(\alpha) > \tau(\alpha)$.
GP4. If $\sigma(\alpha) < \tau(\alpha)$ then $\mathcal{R}^-(\alpha) \neq \emptyset$.
GP5. If $\sigma(\alpha) > \tau(\alpha)$ then $\mathcal{R}^+(\alpha) \neq \emptyset$.
GP6. If $\mathcal{R}^-(\alpha) \overset{\sigma}{=} \mathcal{R}^-(\beta)$, $\mathcal{R}^+(\alpha) \overset{\sigma}{=} \mathcal{R}^+(\beta)$ and $\tau(\alpha) = \tau(\beta)$ then $\sigma(\alpha) = \sigma(\beta)$.
GP7. If $\mathcal{R}^-_\sigma(\alpha) \subsetneq \mathcal{R}^-_\sigma(\beta)$, $\mathcal{R}^+(\alpha) \overset{\sigma}{=} \mathcal{R}^+(\beta)$ and $\tau(\alpha) = \tau(\beta)$ then $\sigma(\beta) < \sigma(\alpha)$.
GP8. If $\mathcal{R}^-(\alpha) \overset{\sigma}{=} \mathcal{R}^-(\beta)$, $\mathcal{R}^+_\sigma(\alpha) \subsetneq \mathcal{R}^+_\sigma(\beta)$ and $\tau(\alpha) = \tau(\beta)$ then $\sigma(\alpha) < \sigma(\beta)$.
GP9. If $\mathcal{R}^-(\alpha) \overset{\sigma}{=} \mathcal{R}^-(\beta), \mathcal{R}^+(\alpha) \overset{\sigma}{=} \mathcal{R}^+(\beta)$ and $\tau(\alpha) < \tau(\beta)$ then $\sigma(\alpha) < \sigma(\beta)$.
GP10. If $\mathcal{R}^-(\alpha) \overset{\sigma}{<} \mathcal{R}^-(\beta)$, $\mathcal{R}^+(\alpha) \overset{\sigma}{=} \mathcal{R}^+(\beta)$ and $\tau(\alpha) = \tau(\beta)$ then $\sigma(\beta) < \sigma(\alpha)$.
GP11. If $\mathcal{R}^-(\alpha) \overset{\sigma}{=} \mathcal{R}^-(\beta)$, $\mathcal{R}^+(\alpha) \overset{\sigma}{>} \mathcal{R}^+(\beta)$ and $\tau(\alpha) = \tau(\beta)$ then $\sigma(\beta) < \sigma(\alpha)$.

In [10,29], two general principles (and their strict counterparts) were also identified as a more synthetic way of describing the desirable (group) properties of a gradual semantics.

The intuition for the first principle is that a difference in an argument's strength and base score must correspond to an imbalance in its attackers' and supporters' strengths.

Principle 1. *[10,29] A gradual semantics σ is* balanced *iff for any $\alpha \in \mathcal{X}$:*

1. *if* $\mathcal{R}^-(\alpha) \overset{\sigma}{=} \mathcal{R}^+(\alpha)$ *then* $\sigma(\alpha) = \tau(\alpha)$;
2. *if* $\mathcal{R}^-(\alpha) \overset{\sigma}{>} \mathcal{R}^+(\alpha)$ *then* $\sigma(\alpha) < \tau(\alpha)$;
3. *if* $\mathcal{R}^-(\alpha) \overset{\sigma}{<} \mathcal{R}^+(\alpha)$ *then* $\sigma(\alpha) > \tau(\alpha)$.

A gradual semantics σ is strictly balanced *iff σ is balanced and for any $\alpha \in \mathcal{X}$:*

4. *if* $\sigma(\alpha) < \tau(\alpha)$ *then* $\mathcal{R}^-(\alpha) \overset{\sigma}{>} \mathcal{R}^+(\alpha)$;
5. *if* $\sigma(\alpha) > \tau(\alpha)$ *then* $\mathcal{R}^-(\alpha) \overset{\sigma}{<} \mathcal{R}^+(\alpha)$.

In [10,29] it is shown that if σ is balanced then it satisfies GP1 to GP3 and if it is strictly balanced then it satisfies GP1 to GP5.

The second principle requires that the strength of an argument depends monotonically on its base score and on the strengths of its attackers and supporters. To introduce this principle formally, we first recall the notion of shaping triple of an argument [10,29], where for any $\alpha \in \mathcal{X}$, the *shaping triple* of α is $(\tau(\alpha), \mathcal{R}^+(\alpha), \mathcal{R}^-(\alpha))$, denoted $\mathcal{ST}(\alpha)$. Given $\alpha, \beta \in \mathcal{X}$, $\mathcal{ST}(\beta)$ is said to be: *as boosting as* $\mathcal{ST}(\alpha)$, denoted as $\mathcal{ST}(\alpha) \simeq \mathcal{ST}(\beta)$, iff $\tau(\alpha) = \tau(\beta)$, $\mathcal{R}^+(\alpha) \overset{\sigma}{=} \mathcal{R}^+(\beta)$, and $\mathcal{R}^-(\beta) \overset{\sigma}{=} \mathcal{R}^-(\alpha)$; *at least as boosting as* $\mathcal{ST}(\alpha)$, denoted as $\mathcal{ST}(\alpha) \preceq \mathcal{ST}(\beta)$, iff $\tau(\alpha) \leq \tau(\beta)$, $\mathcal{R}^+(\alpha) \overset{\sigma}{\leq} \mathcal{R}^+(\beta)$, and $\mathcal{R}^-(\beta) \overset{\sigma}{\leq} \mathcal{R}^-(\alpha)$; or *strictly more boosting than* $\mathcal{ST}(\alpha)$, denoted as $\mathcal{ST}(\alpha) \prec \mathcal{ST}(\beta)$, iff $\mathcal{ST}(\alpha) \preceq \mathcal{ST}(\beta)$ and $\mathcal{ST}(\beta) \not\preceq \mathcal{ST}(\alpha)$. (See [10,29] for intuitions and illustrations.)

Principle 2. *[10,29] A gradual semantics σ is* monotonic *iff:*

1. *for any $\alpha, \beta \in \mathcal{X}$, if $\mathcal{ST}(\alpha) \simeq \mathcal{ST}(\beta)$ then $\sigma(\alpha) = \sigma(\beta)$;*
2. *if $\mathcal{ST}(\alpha) \preceq \mathcal{ST}(\beta)$ then $\sigma(\alpha) \leq \sigma(\beta)$.*

A gradual semantics σ is strictly monotonic iff σ is monotonic and:

3. *for any $\alpha, \beta \in \mathcal{X}$, if $\mathcal{ST}(\alpha) \prec \mathcal{ST}(\beta)$ then $\sigma(\alpha) < \sigma(\beta)$.*

In [10,29] it is shown that if σ is (strictly) monotonic then it satisfies GP6 to GP11.

We will now show that the PR semantics σ satisfies some, but not all, of these desirable properties for gradual semantics. We will consider whether or not the properties are satisfied by the semantics σ when applied to a generic QBAF, in Propositions 1 and 2, or when applied to a PRAF (denoted as $\langle PR, \sigma \rangle$), in Propositions 3 and 4 (see Table 1 for a compact summary). Note that in the first case, if attacks are present in the QBAF, they are simply ignored by the definition of the semantics, and some of the properties may not hold for this mere reason.

Proposition 1. *σ satisfies GP1, GP3, GP4, GP5 but not GP2, and thus is* not *balanced.*

Proof. GP1 holds as when $\mathcal{R}^+(\alpha) = \emptyset$, $\sigma(\alpha) = \tau(\alpha)$. GP3 and GP5 hold as $\sigma(\alpha) > \tau(\alpha)$ is true iff $\sum_{\beta \in \mathcal{R}^+(\alpha)} \frac{\sigma(\beta)}{|\mathcal{R}^+(\beta)|} > 0$ that in turn is true iff $\mathcal{R}^+(\alpha) \neq \emptyset$ because if $\exists \beta \in \mathcal{R}^+(\alpha)$ then, by Remark 1, $|\mathcal{R}^+(\beta)| > 0$ and, by Remark 2, $\sigma(\beta) > 0$. GP4 holds because its preconditions cannot be verified: by Remark 2, $\forall \alpha \in \mathcal{X}, \sigma(\alpha) \geq \tau(\alpha)$. GP2 does not hold as when $\mathcal{R}^+(\alpha) = \emptyset$, $\sigma(\alpha) = \tau(\alpha)$ independently of $\mathcal{R}^-(\alpha)$, which is ignored in the definition of σ. □

Proposition 2. *σ satisfies GP8 and GP9 but not GP6, GP7, GP10 and GP11, and thus is* not *monotonic.*

Proof. GP8 holds as if $\tau(\alpha) = \tau(\beta)$ and $\mathcal{R}^+_\sigma(\alpha) \subset \mathcal{R}^+_\sigma(\beta)$ and we assume by contradiction that $\sigma(\alpha) \geq \sigma(\beta)$, then $\sum_{\gamma \in \mathcal{R}^+(\alpha)} \frac{\sigma(\gamma)}{|\mathcal{R}^+(\gamma)|} \geq \sum_{\gamma \in \mathcal{R}^+(\beta)} \frac{\sigma(\gamma)}{|\mathcal{R}^+(\gamma)|}$, but this is not possible, by Remark 2 because $\nexists \gamma$ such that $\sigma(\gamma) \leq 0$. GP9 holds because its preconditions cannot be verified: by Definition 3 τ is a constant, thus $\nexists \alpha, \beta \in \mathcal{X} : \tau(\alpha) \neq \tau(\beta)$. GP6: in the framework in Figure 1, we have $\mathcal{R}^+(\beta) \overset{\sigma}{=} \mathcal{R}^+(\delta)$ but $\sigma(\beta) \neq \sigma(\delta)$. GP7 and GP10 cannot hold as attackers do not affect σ. GP11: in the framework in Figure 1, we have $\mathcal{R}^+(\zeta) \overset{\sigma}{>} \mathcal{R}^+(\eta)$ but $\sigma(\zeta) < \sigma(\eta)$. □

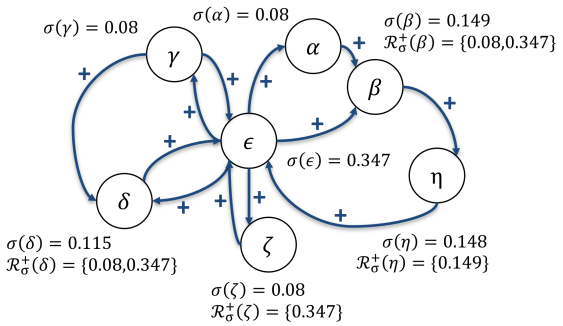

Fig. 1. Counter-example to GP6 and GP11 for the PR semantics σ in Proposition 2.

Proposition 3. *$\langle PR, \sigma \rangle$ is strictly balanced and thus satisfies GP1 to GP5.*

Proof. For balance, Point 1 holds as, by Definition 4, if $\mathcal{R}^+(\alpha) = \emptyset$ then $\sigma(\alpha) = \tau(\alpha)$. Point 2 holds trivially because its preconditions cannot be satisfied by an sQBAF since $\nexists \alpha$ such that $\mathcal{R}^+(\alpha) \overset{\sigma}{<} \emptyset$. Points 3 and 5 hold as if $\mathcal{R}^+(\alpha) \overset{\sigma}{>} \emptyset$ then $\mathcal{R}^+(\alpha) \neq \emptyset$ and we already proved in Proposition 1 for GP3 and GP5 that $\sigma(\alpha) > \tau(\alpha)$ iff $\mathcal{R}^+(\alpha) \neq \emptyset$. For strict balance, Point 4 holds because, by Remark 2, $\nexists \alpha$ such that $\sigma(\alpha) < \tau(\alpha)$. □

Proposition 4. *$\langle PR, \sigma \rangle$ satisfies GP7 to GP10 but not GP6 or GP11 and thus is* not *monotonic.*

Proof. GP6 and GP11 can be shown not to hold with the same counterexamples given in Proposition 2. GP8 and GP9 hold as, by Proposition 2, they hold for σ in general. GP7 and GP10 hold because their preconditions cannot be verified: $\forall \alpha \in \mathcal{X}, \mathcal{R}^-(\alpha) = \emptyset$, thus trivially $\mathcal{R}^-(\alpha) \overset{\sigma}{=} \emptyset$. □

We have thus shown that directly interpreting PR as a gradual semantics for an sQBAF does not give rise to a satisfactory outcome in terms of formal properties. Indeed, while using PR as a semantics is somehow straightforward, it does not appear fully appropriate from a modeling perspective, as it does not provide a suitable argumentative counterpart to some key aspects of PR. In particular, note that, as a consequence of the PR definition, the strength of each node depends not only on the strengths of its supporters but also on the cardinality of their outgoing supports. This has quite counter-intuitive effects from an argumentation perspective which could also affect explanations generated from this sQBAF. For example, consider the situation where two nodes have the same strength $\sigma(\alpha) = \sigma(\beta)$, but α has one outgoing support, while β has ten: the latter's support to each of its children is actually ten times 'less powerful' (i.e. it transfers $1/10$ of the strength) than the former's. It follows that a node γ supported by α only and a node δ supported by β only would have different strengths even if their supporters appear to be equivalent (formally the shaping triples of γ and δ are the same). This is the main reason for the lack of several desirable properties and calls for an alternative approach, which we introduce next.

4. PageRank as a Gradual Semantics in a Meta-Argumentation Framework

In this section, we introduce an alternative approach to capture PageRank as an argumentation semantics. To this purpose we transform the sQBAF corresponding to a set of linked pages into a QBAF including additional meta-arguments and attacks between them. The underlying intuition is that each additional meta-argument can be understood as a vehicle of support from one page to another and that supports from the same page are in mutual conflict as they 'compete' in drawing strength from the same source.

In particular, as shown in Figure 2, we add a meta-argument on every support relationship in the original PRAF, and all the meta-arguments supported by the same page attack each other. While the 'regular' arguments still represent the pages, these new meta-arguments correspond to the links between them. This increases the expressivity of the representation, as it includes attacks between the meta-arguments corresponding to links from the same page in order to describe the fact that they 'compete' for conveying strength, as mentioned above. As a consequence, the

more links originating from the same page, the lower the strength transferred through each of them.

Definition 5. *Given a PRAF* $PR = \langle \mathcal{X}, \emptyset, \mathcal{R}^+, \tau \rangle$*, the* PageRank Meta-Argumentation Framework *(MPRAF) derived from* PR *is a QBAF* $\langle \mathcal{X} \cup \mathcal{M}, \widehat{\mathcal{R}}^-, \widehat{\mathcal{R}}^+, \widehat{\tau} \rangle$*, where:*

- $\mathcal{M} = \{m_{\alpha,\beta}{:}(\alpha,\beta) \in \mathcal{R}^+\}$ *is the set of meta-arguments,*
- $\widehat{\mathcal{R}}^+ = \{(\alpha, m_{\alpha,\beta}), (m_{\alpha,\beta}, \beta){:}\alpha,\beta \in \mathcal{X}, m_{\alpha,\beta} \in \mathcal{M}\}$ *is the set of supports,*
- $\widehat{\mathcal{R}}^- = \{(m_{\alpha,\beta}, m_{\alpha,\gamma}) \in \mathcal{M}\times\mathcal{M}{:}(\alpha,\beta),(\alpha,\gamma) \in \mathcal{R}^+\}$ *is the set of attacks,*
- $\widehat{\tau} : \mathcal{X} \cup \mathcal{M} \mapsto \widehat{\mathbb{I}} = [0,1[$ *is the base score defined as the function:*

$$\widehat{\tau}(\alpha) = \begin{cases} 0 & \textit{if } \alpha \in \mathcal{M} \\ \frac{1-d}{|\mathcal{X}|} & \textit{if } \alpha \in \mathcal{X}. \end{cases}$$

Figure 2 illustrates the transformation of a PRAF into an MPRAF: the supports go from a 'regular' argument to another through an intermediate meta-argument. The following remarks illustrate some of the properties of MPRAFs $\langle \mathcal{X} \cup \mathcal{M}, \widehat{\mathcal{R}}^-, \widehat{\mathcal{R}}^+, \widehat{\tau} \rangle$.

Remark 3. *For any* $\alpha \in \mathcal{X}$*,* $\widehat{\mathcal{R}}^-(\alpha) = \emptyset$*.*

Remark 4. *For any* $m_{\alpha,\beta} \in \mathcal{M}$*,* $\exists!\alpha \in \widehat{\mathcal{R}}^+(m_{\alpha,\beta})$*,* $\exists!\beta \in \widehat{\mathfrak{R}}^+(m_{\alpha,\beta})$ *,* $\alpha \in \mathcal{X}$ *and* $\beta \in \mathcal{X}$*.*

Remark 5. *For any* $m_{\alpha,\beta} \in \mathcal{M}$*,* $|\widehat{\mathcal{R}}^-(m_{\alpha,\beta})| + 1 = |\mathfrak{R}^+(\alpha)| = |\widehat{\mathfrak{R}}^+(\alpha)|$*.*

Remark 6. *For any* $\alpha \in \mathcal{X}$ *such that* $\exists!\ m_{\alpha,\beta} : (\alpha, m_{\alpha,\beta}) \in \widehat{\mathcal{R}}^+$*,* $\widehat{\mathcal{R}}^-(m_{\alpha,\beta}) = \emptyset$*.*

With reference to MPRAFs, we now define a gradual semantics $\widehat{\sigma}$, whose outcomes on 'regular' arguments coincide with the score produced by PR, as proved in Theorem 1.

Definition 6. *The* Meta-PageRank semantics *(M-PR) is a gradual semantics* $\widehat{\sigma} : \mathcal{X} \cup \mathcal{M} \mapsto \widehat{\mathbb{I}}$ *such that:*

$$\widehat{\sigma}(\alpha) = \widehat{\tau}(\alpha) + \sqrt{d} \cdot \frac{\sum_{\beta \in \widehat{\mathcal{R}}^+(\alpha)} \widehat{\sigma}(\beta)}{|\widehat{\mathcal{R}}^-(\alpha)| + 1} \quad \forall \alpha \in \mathcal{X} \cup \mathcal{M}.$$

We now prove that, given a PRAF and corresponding MPRAF, for any $\alpha \in \mathcal{X}$, the strength $\widehat{\sigma}(\alpha)$ according to Definition 6 is the same as the strength $\sigma(\alpha)$ according to Definition 4, i.e. to the PR score.

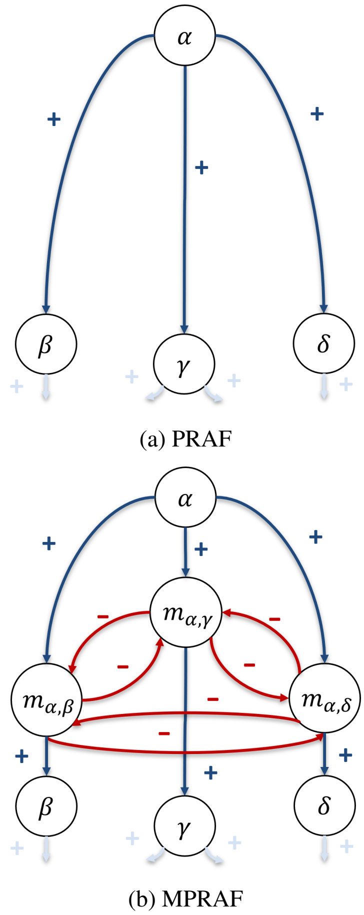

Fig. 2. Example of a transformation from a PRAF to an MPRAF.

Theorem 1 (Equivalence of σ-$\widehat{\sigma}$)**.** *Given a PRAF* $\langle \mathcal{X}, \emptyset, \mathcal{R}^+, \tau \rangle$*, denoted as* PR*, and the corresponding MPRAF* $\langle \mathcal{X} \cup \mathcal{M}, \widehat{\mathcal{R}}^-, \widehat{\mathcal{R}}^+, \widehat{\tau} \rangle$*, denoted as* $\widehat{PR}$*, with the semantics* σ *for* PR *and* $\widehat{\sigma}$ *for* $\widehat{PR}$*, for any argument* $\alpha \in \mathcal{X}$ *it holds that* $\sigma(\alpha) = \widehat{\sigma}(\alpha)$*.*

Proof. $\widehat{\sigma}(\alpha) = \frac{1-d}{|\mathcal{X}|} + \sqrt{d} \cdot \frac{\sum_{\gamma \in \widehat{\mathcal{R}}^+(\alpha)} \widehat{\sigma}(\gamma)}{|\widehat{\mathcal{R}}^-(\alpha)|+1}$ by Definition 6. By hypothesis $\alpha \in \mathcal{X}$, thus if $\gamma \in \widehat{\mathcal{R}}^+(\alpha)$ then $\gamma \in \mathcal{M}$, so we can rewrite γ as $m_{\beta,\alpha}$ where $\beta \in \mathcal{R}^+(\alpha)$. By the same hypothesis, we can derive, by Remark 3, that $|\widehat{\mathcal{R}}^-(\alpha)| = 0$. This means that $\widehat{\sigma}(\alpha)$ can be rewritten as $\frac{1-d}{|\mathcal{X}|} + \sqrt{d} \cdot \sum_{m_{\beta,\alpha} \in \widehat{\mathcal{R}}^+(\alpha)} \widehat{\sigma}(m_{\beta,\alpha})$. Expliciting $\widehat{\sigma}(m_{\beta,\alpha})$ by Definition 6 and recalling that, by Definition 5, $\tau(m_{\beta,\alpha}) = 0$ because $m_{\beta,\alpha}$ is a meta-argument, $\widehat{\sigma}(\alpha) =$

$= \frac{1-d}{|\mathcal{X}|} + \sqrt{d} \cdot \sum_{m_{\beta,\alpha} \in \widehat{\mathcal{R}}^+(\alpha)} \left(\sqrt{d} \cdot \frac{\sum_{\beta \in \widehat{\mathcal{R}}^+(m_{\beta,\alpha})} \widehat{\sigma}(\beta)}{|\widehat{\mathcal{R}}^-(m_{\beta,\alpha})|+1} \right)$.

We recall that, by Remark 4, $\exists!\beta : \beta \in \widehat{\mathcal{R}}^+(m_{\beta,\alpha})$ be-

cause $m_{\beta,\alpha} \in \mathcal{M}$. Furthermore, we know by Remark 5 that $|\widehat{\mathcal{R}}^-(m_{\beta,\alpha})| + 1 = |\mathfrak{R}^+(\beta)|$. Thus, $\widehat{\sigma}(\alpha) = \frac{1-d}{|\mathcal{X}|} + d \cdot \sum_{m_{\beta,\alpha} \in \widehat{\mathcal{R}}^+(\alpha)} \frac{\widehat{\sigma}(\beta)}{|\mathfrak{R}^+(\beta)|}$. This is equivalent to $\widehat{\sigma}(\alpha) = \frac{1-d}{|\mathcal{X}|} + d \cdot \sum_{\beta \in \mathcal{R}^+(\alpha)} \frac{\widehat{\sigma}(\beta)}{|\mathfrak{R}^+(\beta)|} = \sigma(\alpha)$. □

Proposition 5 proves that the codomain of $\widehat{\sigma}$ is $\widehat{\mathbb{I}}$.

Proposition 5. *The codomain of $\widehat{\sigma}$ on an MPRAF $\langle \mathcal{X} \cup \mathcal{M}, \widehat{\mathcal{R}}^-, \widehat{\mathcal{R}}^+, \widehat{\tau} \rangle$ is $\widehat{\mathbb{I}} =]0, 1]$. Moreover, for any $\alpha \in \mathcal{X} \cup \mathcal{M}$, if $\alpha \in \mathcal{X}$ then $\widehat{\sigma}(\alpha) \geq \frac{1-d}{|\mathcal{X}|}$, otherwise $\widehat{\sigma}(\alpha) > 0$.*

Proof. By Definition 6, $\widehat{\sigma}(\alpha)$ is the sum of $\widehat{\tau}(\alpha)$ and positive values. Hence if $\alpha \in \mathcal{X}$ then $\widehat{\sigma}(\alpha) \geq \frac{1-d}{|\mathcal{X}|} > 0$. Otherwise, if $\alpha \in \mathcal{M}$ then, by Definitions 5 and 6, $\widehat{\sigma}(\alpha) = \sqrt{d} \cdot \frac{\sum_{\beta \in \widehat{\mathcal{R}}^+(\alpha)} \widehat{\sigma}(\beta)}{|\widehat{\mathcal{R}}^-(\alpha)|+1} \geq \sqrt{d} \cdot \sum_{\beta \in \widehat{\mathcal{R}}^+(\alpha)} \widehat{\sigma}(\beta)$, and since $\beta \in \mathcal{X}$ then $\widehat{\sigma}(\beta) > 0 \quad \forall \beta$, hence $\widehat{\sigma}(\alpha) > 0$. By Theorem 1 and by Remark 2, we have that if $\alpha \in \mathcal{X}$ then $\widehat{\sigma}(\alpha) \leq 1$. Otherwise, if $\alpha \in \mathcal{M}$ then, by Remark 4, $\widehat{\mathcal{R}}^+(\alpha) = \{\beta\}$ and $\beta \in \mathcal{X}$, hence by Definition 6, $\widehat{\sigma}(\alpha) = \sqrt{d} \cdot \frac{\widehat{\sigma}(\beta)}{|\widehat{\mathcal{R}}^-(\alpha)|+1} \leq 1$. □

The next proposition sheds light on the intuition behind our MPRAFs, in that the support from non-meta-arguments is partitioned among the meta-arguments. Meta-arguments supported by the same 'regular' argument all have the same strength since according to the random surfer model the probability of clicking on links is uniform.

Proposition 6. *In an MPRAF $\langle \mathcal{X} \cup \mathcal{M}, \widehat{\mathcal{R}}^-, \widehat{\mathcal{R}}^+, \widehat{\tau} \rangle$, if a meta-argument $\alpha \in \mathcal{M}$ has attackers then $\widehat{\sigma}(\alpha) = \widehat{\sigma}(\gamma), \ \forall \gamma \in \widehat{\mathcal{R}}^-(\alpha)$.*

Proof. By Definition 5, $\forall \gamma \in \widehat{\mathcal{R}}^-(\alpha) \quad \gamma \in \mathcal{M}$ and by Definition 5 and Remark 4 $\forall \gamma \in \widehat{\mathcal{R}}^-(\alpha) \quad \widehat{\mathcal{R}}^+(\alpha) = \widehat{\mathcal{R}}^+(\gamma) = \{\beta\}$ where $\beta \in \mathcal{X}$ is the single supporter of α. By Definition 6, $\widehat{\sigma}(\alpha) = \widehat{\tau}(\alpha) + \sqrt{d} \cdot \frac{\sum_{\beta \in \widehat{\mathcal{R}}^+(\alpha)} \widehat{\sigma}(\beta)}{|\widehat{\mathcal{R}}^-(\alpha)|+1}$, and by Definition 5 and Remark 4, $\widehat{\sigma}(\alpha) = \sqrt{d} \cdot \frac{\widehat{\sigma}(\beta)}{|\widehat{\mathcal{R}}^-(\alpha)|+1}$, and the same is true for any $\gamma \in \widehat{\mathcal{R}}^-(\alpha)$: $\widehat{\sigma}(\gamma) = \sqrt{d} \cdot \frac{\widehat{\sigma}(\beta)}{|\widehat{\mathcal{R}}^-(\gamma)|+1}$. By construction α and the elements of $\widehat{\mathcal{R}}^-(\alpha)$ all attack each other, thus $|\widehat{\mathcal{R}}^-(\alpha)| = |\widehat{\mathcal{R}}^-(\gamma)| \ \forall \gamma \in \widehat{\mathcal{R}}^-(\alpha)$, and the result follows. □

We now assess this framework and semantics with respect to the desirable properties.

Proposition 7. *$\widehat{\sigma}$ satisfies GP1, GP4, GP5, GP6, GP8, GP9 and GP11.*

Proof. GP1: by Definition 6, if $\widehat{\mathcal{R}}^+(\alpha) = \emptyset$ and $\widehat{\mathcal{R}}^-(\alpha) = \emptyset$ then the second term of the sum is always 0, therefore $\sigma(\alpha) = \tau(\alpha)$. GP4 holds because the GP's preconditions cannot be verified: by Proposition 5, $\forall \alpha \in \mathcal{X} \quad \widehat{\sigma}(\alpha) \geq \widehat{\tau}(\alpha)$. GP5: by Definition 6, $\widehat{\sigma}(\alpha) > \widehat{\tau}(\alpha)$ iff $\sum_{\beta \in \widehat{\mathcal{R}}^+(\alpha)} \widehat{\sigma}(\beta) > 0$. Thus, it must be the case that $\exists \beta \in \widehat{\mathcal{R}}^+(\alpha) : \widehat{\sigma}(\beta) > 0$, therefore $\widehat{\mathcal{R}}^+(\alpha) \neq \emptyset$ GP6: follows directly from Definition 6. GP8: if $\widehat{\mathcal{R}}^-(\alpha) \overset{\sigma}{=} \widehat{\mathcal{R}}^-(\beta)$ then $|\widehat{\mathcal{R}^-_\sigma}(\alpha)| = |\widehat{\mathcal{R}^-_\sigma}(\beta)|$ and if $\widehat{\mathcal{R}^+_\sigma}(\alpha) \subsetneq \widehat{\mathcal{R}^+_\sigma}(\beta)$ then $\sum_{\gamma \in \widehat{\mathcal{R}}^+(\alpha)} \widehat{\sigma}(\gamma) < \sum_{\gamma \in \widehat{\mathcal{R}}^+(\beta)} \widehat{\sigma}(\gamma)$. The result follows from Definition 6. GP9: if $\widehat{\mathcal{R}}^-(\alpha) \overset{\sigma}{=} \widehat{\mathcal{R}}^-(\beta)$ then $|\widehat{\mathcal{R}^-_\sigma}(\alpha)| = |\widehat{\mathcal{R}^-_\sigma}(\beta)|$ and if $\widehat{\mathcal{R}}^+(\alpha) \overset{\sigma}{=} \widehat{\mathcal{R}}^+(\beta)$ then $\sum_{\gamma \in \widehat{\mathcal{R}}^+(\alpha)} \widehat{\sigma}(\gamma) = \sum_{\gamma \in \widehat{\mathcal{R}}^+(\beta)} \widehat{\sigma}(\gamma)$. The result follows from Definition 6. GP11: if $\widehat{\mathcal{R}}^-(\alpha) \overset{\sigma}{=} \widehat{\mathcal{R}}^-(\beta)$ then $|\widehat{\mathcal{R}^-_\sigma}(\alpha)| = |\widehat{\mathcal{R}^-_\sigma}(\beta)|$ and if $\widehat{\mathcal{R}}^+(\alpha) \overset{\sigma}{>} \widehat{\mathcal{R}}^+(\beta)$ then $\sum_{\gamma \in \widehat{\mathcal{R}}^+(\alpha)} \widehat{\sigma}(\gamma) > \sum_{\gamma \in \widehat{\mathcal{R}}^+(\beta)} \widehat{\sigma}(\gamma)$. The result follows from Definition 6. □

Proposition 8. *$\langle \widehat{PR}, \widehat{\sigma} \rangle$ is balanced and thus satisfies GP1 to GP3. However, $\langle \widehat{PR}, \widehat{\sigma} \rangle$ is not strictly balanced.*

Proof. For balance, Point 1: (A) If $\widehat{\mathcal{R}}^-(\alpha) \overset{\sigma}{=} \widehat{\mathcal{R}}^+(\alpha) = \emptyset$ then the result follows by Definition 6. (B) Otherwise, if $\widehat{\mathcal{R}}^-(\alpha) \neq \emptyset$ then $\alpha \in \mathcal{M}$ and thus it has a single supporter β. There are two possible scenarios, which turn out to be impossible, as they contradict the hypothesis. (B.i) $\exists! \gamma \in \mathcal{M} : (\beta, \alpha), (\beta, \gamma) \in \widehat{\mathcal{R}}^+$, then we get $\{\beta\} = \widehat{\mathcal{R}}^+(\alpha) \overset{\sigma}{>} \widehat{\mathcal{R}}^-(\alpha) = \{\gamma\}$ (which contradicts the hypothesis) because by Definition 6 $\widehat{\sigma}(\alpha) = \widehat{\sigma}(\gamma) < \widehat{\sigma}(\beta)$ (B.ii) $\exists_{>1} \gamma_1, ..., \gamma_n \in \mathcal{M} : (\beta, \alpha), (\beta, \gamma_1), ..., (\beta, \gamma_n) \in \widehat{\mathcal{R}}^+$, hence $|\widehat{\mathcal{R}}^-(\alpha)| > 1$, therefore it cannot hold that $\{\gamma_1, ..., \gamma_n\} = \widehat{\mathcal{R}}^-(\alpha) \overset{\sigma}{=} \widehat{\mathcal{R}}^+(\alpha) = \{\beta\}$, since by Definition 6 it holds again $\widehat{\sigma}(\alpha) = \widehat{\sigma}(\gamma_1) = ... = \widehat{\sigma}(\gamma_n) < \widehat{\sigma}(\beta)$, hence there cannot be any injective mapping $f : \widehat{\mathcal{R}}^-(\alpha) \to \widehat{\mathcal{R}}^+(\alpha) : \forall \alpha \in \widehat{\mathcal{R}}^-(\alpha), \sigma(f(\alpha)) \geq \sigma(\alpha)$, and thus there is no strength-equivalence relationship between $\widehat{\mathcal{R}}^-(\alpha)$ and $\widehat{\mathcal{R}}^+(\alpha)$, contradicting the hypothesis. Point 2. For $\widehat{\mathcal{R}}^-(\alpha) \overset{\sigma}{>} \widehat{\mathcal{R}}^+(\alpha)$ to hold $\widehat{\mathcal{R}}^-(\alpha) \neq \emptyset$, thus $\alpha \in \mathcal{M}$. Hence, we are in the same situation of (B) in the proof of Point 1, and therefore the precondition cannot hold and the result follows. Point 3. By Proposition 5, $\widehat{\sigma}(\alpha) > 0$ and if $\widehat{\mathcal{R}}^-(\alpha) \overset{\sigma}{<} \widehat{\mathcal{R}}^+(\alpha)$ then $\widehat{\mathcal{R}}^+(\alpha) \neq \emptyset$. Hence by Definition 6, $\widehat{\sigma}(\alpha) > \widehat{\tau}(\alpha)$, thus $\langle \widehat{PR}, \widehat{\sigma} \rangle$ is balanced. For strict balance, Point 4 holds because $\nexists \alpha : \widehat{\sigma}(\alpha) <$

$\widehat{\tau}(\alpha)$. But, Point 5 does not hold. For example, consider the framework in Figure 2.b and in particular $m_{\alpha,\gamma} \in \mathcal{M}$ that it is supported by $\alpha \in \mathcal{X}$ and attacked by $m_{\alpha,\beta}, m_{\alpha,\delta} \in \mathcal{M}$. By Definition 5 and Proposition 5, we have that $\widehat{\sigma}(m_{\alpha,\gamma}) \leq \widehat{\sigma}(\alpha)$ and $\widehat{\sigma}(m_{\alpha,\gamma}) = \widehat{\sigma}(m_{\alpha,\beta}) = \widehat{\sigma}(m_{\alpha,\delta}) > 0$. Hence, $\widehat{\sigma}(m_{\alpha,\gamma}) > \widehat{\tau}(m_{\alpha,\gamma})$, but $\widehat{\mathcal{R}}^+(m_{\alpha,\gamma}) \overset{\sigma}{\not\geq} \widehat{\mathcal{R}}^-(m_{\alpha,\gamma})$ because no injective mapping exists from $\widehat{\mathcal{R}}^-(m_{\alpha,\gamma})$ to $\widehat{\mathcal{R}}^+(m_{\alpha,\gamma})$. Thus $\widehat{\mathcal{R}}^+(m_{\alpha,\gamma}) \overset{\sigma}{\not>} \widehat{\mathcal{R}}^-(m_{\alpha,\gamma})$ and therefore $\langle \widehat{PR}, \widehat{\sigma} \rangle$ is not strictly balanced. □

Proposition 9. *$\langle \widehat{PR}, \widehat{\sigma} \rangle$ is strictly monotonic and thus satisfies GP6 to GP11.*

Proof. Point 1: if $\widehat{\mathcal{R}}^-(\alpha) \overset{\sigma}{=} \widehat{\mathcal{R}}^-(\beta)$ then $|\widehat{\mathcal{R}}^-(\alpha)| = |\widehat{\mathcal{R}}^-(\beta)|$ and if $\widehat{\mathcal{R}}^+(\alpha) \overset{\sigma}{=} \widehat{\mathcal{R}}^+(\beta)$ then $\sum_{\gamma \in \widehat{\mathcal{R}}^+(\alpha)} \widehat{\sigma}(\gamma) = \sum_{\gamma \in \widehat{\mathcal{R}}^+(\beta)} \widehat{\sigma}(\gamma)$. The result follows from Definition 6. Point 3: if $\alpha, \beta \in \mathcal{X}$ then $\widehat{\tau}(\alpha) = \widehat{\tau}(\beta)$ and $\widehat{\mathcal{R}}^-(\beta) \overset{\sigma}{=} \widehat{\mathcal{R}}^-(\alpha) = \emptyset$, hence $|\widehat{\mathcal{R}}^-(\alpha)| = |\widehat{\mathcal{R}}^-(\beta)|$. If $\widehat{\mathcal{R}}^+(\alpha) \overset{\sigma}{<} \widehat{\mathcal{R}}^+(\beta)$ then $\sum_{\gamma \in \widehat{\mathcal{R}}^+(\alpha)} \widehat{\sigma}(\gamma) < \sum_{\gamma \in \widehat{\mathcal{R}}^+(\beta)} \widehat{\sigma}(\gamma)$. Thus, by Definition 6, $\widehat{\sigma}(\alpha) < \widehat{\sigma}(\beta)$. If $\alpha \in \mathcal{M}$ and $\beta \in \mathcal{X}$ then $\widehat{\tau}(\alpha) < \widehat{\tau}(\beta)$ and $\widehat{\mathcal{R}}^-(\beta) = \emptyset$. If $\widehat{\mathcal{R}}^-(\alpha) \overset{\sigma}{\geq} \emptyset$ then $|\widehat{\mathcal{R}}^-(\alpha)| \geq |\widehat{\mathcal{R}}^-(\beta)| = 0$. If $\widehat{\mathcal{R}}^+(\alpha) \overset{\sigma}{\leq} \widehat{\mathcal{R}}^+(\beta)$ then $\sum_{\gamma \in \widehat{\mathcal{R}}^+(\alpha)} \widehat{\sigma}(\gamma) \leq \sum_{\gamma \in \widehat{\mathcal{R}}^+(\beta)} \widehat{\sigma}(\gamma)$. Thus, by Definition 6, $\widehat{\sigma}(\alpha) < \widehat{\sigma}(\beta)$. If $\alpha, \beta \in \mathcal{M}$ then $\widehat{\tau}(\alpha) = \widehat{\tau}(\beta)$. If $\widehat{\mathcal{R}}^-(\beta) \overset{\sigma}{\leq} \widehat{\mathcal{R}}^-(\alpha)$ then $|\widehat{\mathcal{R}}^-(\alpha)| \geq |\widehat{\mathcal{R}}^-(\beta)|$. If $\widehat{\mathcal{R}}^+(\alpha) \overset{\sigma}{\leq} \widehat{\mathcal{R}}^+(\beta)$ then $\sum_{\gamma \in \widehat{\mathcal{R}}^+(\alpha)} \widehat{\sigma}(\gamma) \leq \sum_{\gamma \in \widehat{\mathcal{R}}^+(\beta)} \widehat{\sigma}(\gamma)$. Hence, by Definition 6, $\widehat{\sigma}(\alpha) \leq \widehat{\sigma}(\beta)$. For $\mathcal{ST}(\beta) \not\subseteq \mathcal{ST}(\alpha)$ to hold, either:

- $\widehat{\mathcal{R}}^-(\beta) \overset{\sigma}{<} \widehat{\mathcal{R}}^-(\alpha)$ and $\widehat{\mathcal{R}}^+(\alpha) \overset{\sigma}{=} \widehat{\mathcal{R}}^+(\beta)$, or
- $\widehat{\mathcal{R}}^-(\beta) \overset{\sigma}{=} \widehat{\mathcal{R}}^-(\alpha)$ and $\widehat{\mathcal{R}}^+(\alpha) \overset{\sigma}{<} \widehat{\mathcal{R}}^+(\beta)$, or
- $\widehat{\mathcal{R}}^-(\beta) \overset{\sigma}{<} \widehat{\mathcal{R}}^-(\alpha)$ and $\widehat{\mathcal{R}}^+(\alpha) \overset{\sigma}{<} \widehat{\mathcal{R}}^+(\beta)$.

In the first case, by construction of the framework PR, $|\widehat{\mathcal{R}}^-(\alpha)| < |\widehat{\mathcal{R}}^-(\beta)|$, thus $\widehat{\sigma}(\alpha) < \widehat{\sigma}(\beta)$. In the second case, $\sum_{\gamma \in \widehat{\mathcal{R}}^+(\alpha)} \widehat{\sigma}(\gamma) < \sum_{\gamma \in \widehat{\mathcal{R}}^+(\beta)} \widehat{\sigma}(\gamma)$, thus $\widehat{\sigma}(\alpha) < \widehat{\sigma}(\beta)$. In the third case, $\sum_{\gamma \in \widehat{\mathcal{R}}^+(\alpha)} \widehat{\sigma}(\gamma) < \sum_{\gamma \in \widehat{\mathcal{R}}^+(\beta)} \widehat{\sigma}(\gamma)$ and $|\widehat{\mathcal{R}}^-(\alpha)| \leq |\widehat{\mathcal{R}}^-(\beta)|$, thus $\widehat{\sigma}(\alpha) < \widehat{\sigma}(\beta)$. Point 3 implies Point 2, thus the result follows. □

We have thus proven that, through MPRAF, in exchange for a little structural addition, it is possible to ensure equivalence with PR while at the same time satisfying more desirable properties from an argumentation semantics perspective.

Table 1 shows a summary of the properties that the M-PR semantics applied on MPRAFs satisfies, including in particular *monotonicity*. This means that, from a dialectical viewpoint, the strength of an argument depends exclusively on its intrinsic strength, the reasons supporting it and the reasons against it, and any strengthening/weakening of these will affect the argument's strength in an intuitive way.

The satisfaction of monotonicity is achieved through the role ascribed to meta-arguments and is a key factor for exploiting MPRAFs for practical applications, such as the generation of intuitive explanations of the PR score of a page. In this scenario, *monotonicity* is clearly a crucial factor because it allows a user to identify direct dependencies between the strengths of arguments according to the attacks and supports linking them in the graph structure of the MPRAF.

5. Argumentation-based Explanations for PageRank

In this section we first evidence the limits of argumentative explanations for PR based on PRAFs and propose several novel explanations utilising the QBAF with meta-arguments introduced in Section 4. In particular we will consider explanations allowing the user to understand the reasons for a given PR score, providing hints on changes that can improve the score, or giving warnings on strong dependencies of the score on other pages. Throughout this section we assume as given a PRAF $\langle \mathcal{X}, \emptyset, \mathcal{R}^+, \tau \rangle$ and its MPRAF counterpart $\langle \mathcal{X} \cup \mathcal{M}, \widehat{\mathcal{R}}^-, \widehat{\mathcal{R}}^+, \widehat{\tau} \rangle$. For ease of comprehension, we will also propose some examples generated from the Wikipedia web graph that we will introduce in more detail in Section 6.

5.1. Types of explanations: singular explanations and plural explanations

Explanations may have different scopes and levels of abstraction. In this paper, we focus on local explanations concerning the score of a page or a group of pages, rather than global explanations concerning the overall PR score assignment. In particular, we consider two families of explanations:

- *singular explanations* concerning the score of a single page, and
- *plural explanations* concerning the scores of a set of pages.

Table 1

Satisfaction (✓) or not (×) of GPs and principles (**B**alance, **S**trict **B**alance, **M**onotonicity, **S**trict **M**onotonicity) by σ, $\langle PR, \sigma \rangle$, $\widehat{\sigma}$ and $\langle \widehat{PR}, \widehat{\sigma} \rangle$

	GP1	GP2	GP3	GP4	GP5	GP6	GP7	GP8	GP9	GP10	GP11	B	SB	M	SM
σ	✓	×	✓	✓	✓	×	×	✓	✓	×	×	×	×	×	×
$\langle PR, \sigma \rangle$	✓	✓	✓	✓	✓	×	✓	✓	✓	✓	×	✓	✓	×	×
$\widehat{\sigma}$	✓	×	×	✓	✓	✓	×	✓	✓	×	✓	×	×	×	×
$\langle \widehat{PR}, \widehat{\sigma} \rangle$	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	×	✓	✓

We understand a singular explanation as a set of (meta-)arguments in the PRAF or in the MPRAF, accompanied by a function describing their importance w.r.t. the page whose score is explained.

Definition 7. *A* singular explanation *for a page $\alpha \in \mathcal{X}$ is a pair $E(\alpha) = (\mathcal{E}, i)$ where:*

- *$\mathcal{E} \subseteq \mathcal{X} \cup \mathcal{M}$ is the set of* explaining arguments*, and*
- *$i : \mathcal{E} \mapsto \mathbb{R}$ is the* importance function*.*

In the case of plural explanations, the explanation concerns the scores of a set of pages: for each of these pages a set of explaining arguments is given and a function describes the importance of all explaining arguments.

Definition 8. *A* plural explanation *for a set of pages $A \subseteq \mathcal{X}$ is a pair $E(A) = (\{\mathcal{E}_\alpha : \alpha \in A\}, i)$ where:*

- *$\mathcal{E}_\alpha \subseteq \mathcal{X} \cup \mathcal{M}$ is the set of* explaining arguments *for any page $\alpha \in A$, and*
- *$i : \bigcup_{\alpha \in A} \mathcal{E}_\alpha \mapsto \mathbb{R}$ is the* importance function*.*

We will consider several instances of singular explanations and plural explanations, obtained by specific choices of explaining arguments and importance functions, drawn from the underlying PRAF or MPRAF.

5.2. PRAF-based explanations

Consider the problem of identifying *which pages have a major role in determining the score of a given page one is interested in*. If we were to answer this query using only PRAFs we could return the set of supporters of the page of interest: we call this form of explanation a *basic explanation*, in that it is solely based on the PRAF.

Definition 9. *A* basic explanation *for page $\alpha \in \mathcal{X}$ is a singular explanation $E_{\leftarrow}(\alpha) = (\mathcal{R}^+(\alpha), \sigma)$.*

Basic explanations essentially provide a magnification of the original PRAF focused on an argument (page) and its supporters (pages linking to it, whose importance coincides with their score). The result is an explanation like the one presented in Fig. 3.i. Here, the score of the page *Nguyen Dynasty* is explained showing its supporters, with each page score being represented by the size of the relevant bubble. Notice how, looking at this basic explanation, a user might (erroneously) deduce that the score of *Nguyen Dinasty* is mostly determined by *Official Residence*, which is actually not the case (due to the high number of outgoing links from *Official Residence*). Thus, basic explanations have clear limitations. The extent to which basic explanations could be misleading can be quite large, as shown by the excerpt of the MPRAF in Figure 3.ii, where we can see how the actual contributions of the supporters of *Nguyen Dinasty* (the opaque bubbles) compare with their strengths (transparent bubbles).

In fact, basic explanations of this kind answer the question '*Which are the pages with the highest score with a link to a page p?*' but this is different from answering the question '*Which are the pages with the highest contribution to the score of a page p?*' or, more concisely, '*Why does page p have this score?*'. To answer this question using the PRAF representation a user should both have a deeper understanding of how PR works and be shown a larger part of the PRAF, including all the pages linked by the supporters of the considered page. Only then might the user realise that *Hue*, instead of *Official Residence*, is the Wikipedia article providing most support to *Nguyen Dynasty*.

5.3. MPRAF-based explanations

Attribution explanation. The unsuitability of PRAFs as explanatory tools can be overcome by *attribution explanations* based on MPRAFs that focus the atten-

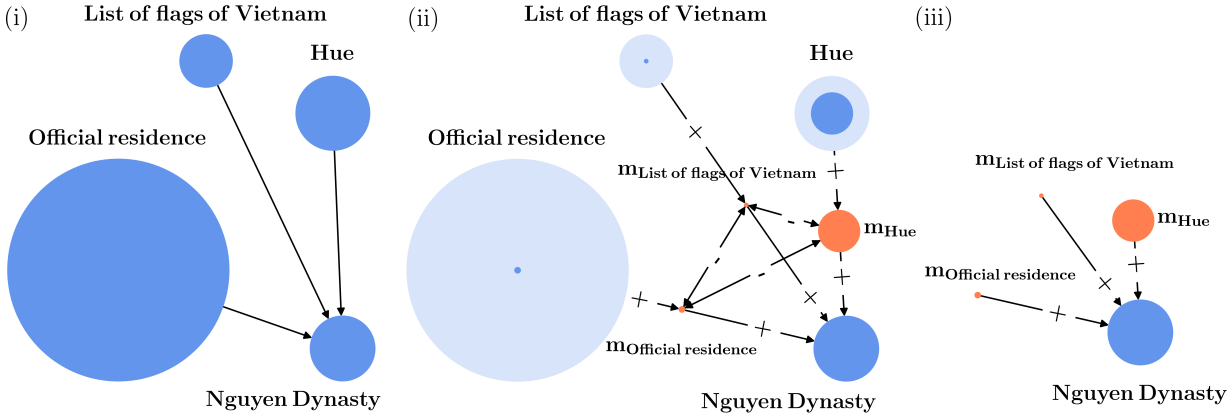

Fig. 3. Transition, for the Wikipedia article *Nguyen Dynasty*, from its basic explanation (i), to the excerpt of the QBAF including it and its supporters, to eventually its attribution explanation (iii). Each bubble represents an argument and its size is proportional to the strength of the argument. In (ii) the opaque bubbles highlight the actual contribution of an argument to the *Nguyen Dynasty* page. Labels $-$ and $+$ indicate, respectively, attacks and supports.

tion of the user only on the meta-arguments supporting the page of interest, thus truly answering the question '*Why does page p have this score?*'.

Definition 10. *An* attribution explanation *for page $\alpha \in \mathcal{X}$ is a singular explanation $E_{\leftarrow}(\alpha) = (\mathcal{R}^{+}(\alpha), \widehat{\sigma})$.*

Fig. 3.iii, shows an example of an attribution explanation for the Wikipedia article *Nguyen Dynasty* as an excerpt of the QBAF comprising the argument of *Nguyen Dynasty* and its supporting meta-arguments. Intuitively, the strength assigned to each meta-argument by our novel semantics corresponds to the support actually flowing from one page to another. In this representation it is clear that the contribution of *Hue* to the score of *Nguyen Dynasty* is bigger than that of *Official Residence*, despite the former's lower PR score.

Besides better supporting explanations of the reasons behind the PR of a page, attribution explanations appear to enable answering other kinds of user queries, like counterfactual questions of the kind: '*What would happen if a given link is suppressed?*'. In this context, meta-arguments' strengths directly show an approximation[1] of the portion of the score that a page would lose if a link were removed. For example, in the attribution explanation in Figure 3.iii, if we remove from the supporters of *Nguyen Dynasty* the page *Hue* then its PR will reduce considerably since *Hue* is the supporter contributing the most to the strength of *Nguyen Dynasty*. Although the full set of meta-arguments potentially included in attribution explanations of a page may be very large (in the order of hundreds) we will show in Section 6 that considering only a limited subset is enough to produce a satisfactory explanation. This means that our explanations fulfil the desideratum of simplicity, avoiding overwhelming the user with too much information when the number of supporters is large.

Contrastive attribution explanation. When two (or more) pages have many shared supporters, understanding why the pages have different scores through attribution explanations is not trivial. *Contrastive attribution explanations* tackle this issue: given a set of pages of interest they show for each of them the nodes contributing exclusively to its score, ignoring the ones they share. Thus, these explanations answer questions of the kind: '*Which are the links that make pages p and q have different scores?*'. explanations of this kind find their typical usage scenario in the assessment of the reasons behind a page having a higher (or lower) score than other pages, comparing their non-shared supporters sorted by their strengths.

Definition 11. *A* contrastive attribution explanation *for a set of pages $A \subseteq \mathcal{X}$ is a plural explanation $E_{\leftrightarrow}(A) = (\{\widehat{\mathcal{R}}^{+}_{\setminus A}(\alpha) : \alpha \in A\}, \widehat{\sigma})$ where $\widehat{\mathcal{R}}^{+}_{\setminus A}(\alpha)$ is the set of exclusive supporters of α in A i.e., $\widehat{\mathcal{R}}^{+}_{\setminus A}(\alpha) = \widehat{\mathcal{R}}^{+}(\alpha) \setminus \{m_{\beta,\alpha} \in \widehat{\mathcal{R}}^{+}(\alpha) : \beta \in \bigcup_{\gamma \in A \setminus \{\alpha\}} \mathcal{R}^{+}(\gamma)\}$.*

Fig. 4 shows an example of a contrastive attribution explanation for the Wikipedia articles *Calorimeter* and

[1] If there are cycles in the MPRAF, the removal of a link could indirectly strengthen or weaken the other incoming links, e.g., because the page of interest is cyclically supporting one of its supporting pages. However, in most cases this gives rise to negligible changes due to PageRank's design.

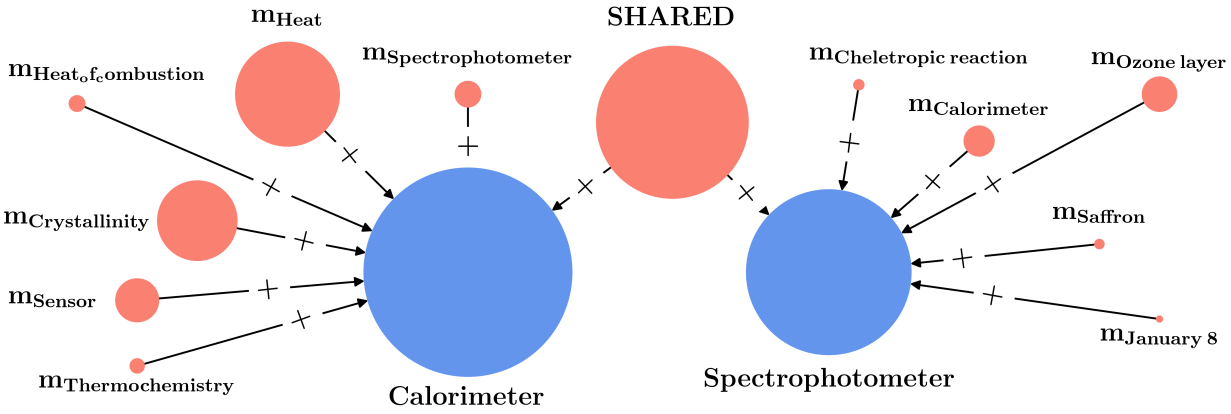

Fig. 4. Contrastive attribution explanation for the Wikipedia articles *Calorimeter* and *Spectrophotometer* from the Wikipedia dataset. The size of the bubbles is proportional to the arguments' strengths. The bubble labeled *SHARED* encompasses the contributions from the 40 shared supporters.

Spectrophotometer. Using this explanation, users can focus on the differences in the supporters of two pages rather than on their common supporters, which in this example amount to 40 nodes. As we will show in Section 6, the number of shared supporters typically increases if one of the two pages supports the other, and even more so if they mutually support each other.

Additive counterfactual explanations. This form of explanation extends our effort in answering counterfactual questions in that *additive counter-factual explanations* provide the user with information on links, not currently present, that if added would increase the score of a page of interest. These explanations answer the question '*To which pages could a link to p be added to maximize the increment of its score?*'. A typical usage scenario of this explanation is searching for pages that could be modified to increase the score of a specific page.

In order to formally define additive counter-factual explanation we will now introduce the definition of the *MPRAF with a link addition or removal*.

Definition 12. *Given two arguments $\alpha, \beta \in \mathcal{X}$, then: if $(\alpha, \beta) \notin \mathcal{R}^+$, we define the* PRAF with the addition of the support (α, β), *denoted by $PR_{+(\alpha,\beta)}$, as $\langle \mathcal{X}, \emptyset, \mathcal{R}^+_{+(\alpha,\beta)}, \tau \rangle$ where $\mathcal{R}^+_{+(\alpha,\beta)} = \mathcal{R}^+ \cup \{(\alpha, \beta)\}$. We denote with $\widehat{PR}_{+(\alpha,\beta)}$, $\widehat{\sigma}_{+(\alpha,\beta)}$ the MPRAF corresponding to $PR_{+(\alpha,\beta)}$ and the semantics $\widehat{\sigma}$ on $\widehat{PR}_{+(\alpha,\beta)}$, respectively.*

We now formally define additive counter-factual explanations.

Definition 13. *An additive counter-factual explanation for page β is a singular explanation $E_{\leftrightarrow}(\beta) = (\widehat{\mathcal{R}}_2^+(\beta), \widehat{\sigma}_{+(\alpha,\beta)})$ where $\widehat{\mathcal{R}}_2^+(\alpha)$ is the set of meta-arguments that are* not *supporters of α or supported by α with backward hop-distance of 2 from α, i.e., $\widehat{\mathcal{R}}_2^+(\alpha) = \{m_{x,\alpha} : x \notin \mathcal{R}^+(\alpha) \wedge x \notin \mathfrak{R}^+(\alpha) \wedge x \in \bigcup_{\beta \in \mathcal{R}^+(\alpha)} \mathcal{R}^+(\beta)\}$.*

Note that, in principle, the set of pages from which one could draw an additional link is potentially very large, thus some restriction is needed, also to ensure that the considered additions are somehow meaningful. For this reason, for this form of explanation to be useful in practice, we opted to include only meta-arguments (links) from pages with backward hop-distance of 2 to the page of interest in the web graph. As we will show in Section 6 this allowed us to select a smaller but "more relevant" portion of meta-arguments.

Fig. 5.i shows an example of this type of explanation for the Wikipedia article *Aztec Empire*, visualizing the 10 most (potentially) influential meta-arguments (selected from 515).

Edit-sensibility counterfactual explanation. While an additional incoming link positively affects the newly linked page, this addition will negatively affect the score of all the other pages linked by the same source. Edit-sensibility counterfactual explanations aim to inform the user about this aspect, giving information on how sensitive the score of a page is to changes in the supporting pages. This type of explanation answers the question '*If an outgoing link is added to page q (a supporter of page p), how much will the score of p change?*'.

To formally define edit-sensibility counterfactual explanations we first define the concept of the *sensi-*

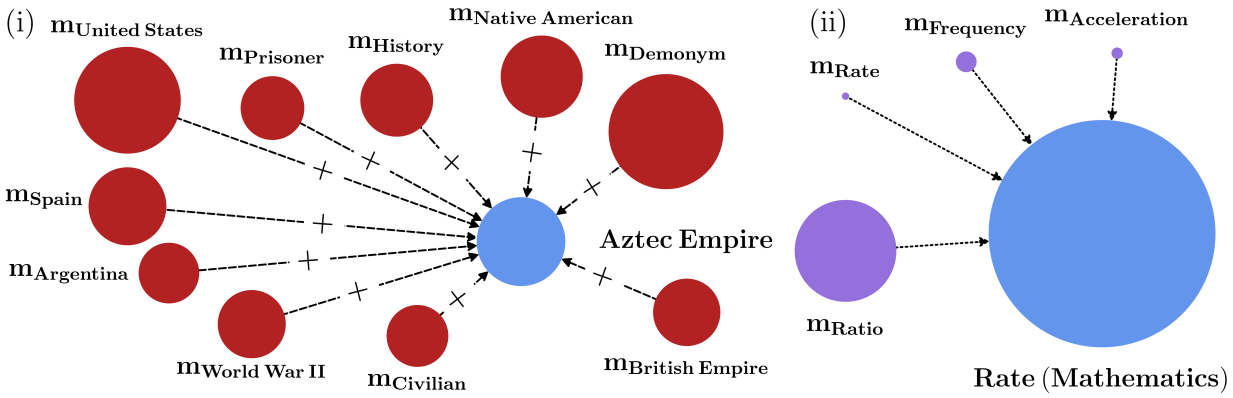

Fig. 5. Additive counterfactual explanation for the Wikipedia article *Aztec Empire* (i) and edit-sensibility counterfactual explanation for the Wikipedia article *Rate (Mathematics)* (ii). Sizes are proportional to the arguments' strengths for *blue* bubbles, and to the importance of the meta-arguments according to the form of explanation for the red and purple bubbles.

tivity of an argument, describing the extent to which a page is susceptible to the change of its supporters.

Definition 14. *Given* $\alpha, \beta, \delta \in \mathcal{X}$, $(\alpha, \beta) \notin \mathcal{R}^+$, *and* $(\alpha, \delta) \in \mathcal{R}^+$, *the* sensitivity to addition *of node* $m_{\alpha,\delta} \in \mathcal{M}$ *is defined as :*

$$\phi(m_{\alpha,\delta}) = \widehat{\sigma}(m_{\alpha,\delta}) - \widehat{\sigma}_{+(\alpha,\beta)}(m_{\alpha,\delta})$$

We can now define edit-sensibility counterfactual explanations.

Definition 15. *An* edit-sensibility counterfactual explanation *for page* α *is a singular explanation* $\underset{\leftarrow}{E}_{\Phi?}(\alpha) = (\widehat{\mathcal{R}}^+(\alpha), \phi)$.

Essentially, this form of explanation highlights how much a page score is "exposed" to endogenous changes in the "link structure" of other pages. Fig. 5.ii shows an example of this explanation for the Wikipedia article *Rate (Mathematics)*. Here, the sizes of the supporting meta-arguments (including that of the page *Rate (Mathematics)*) are proportional to the sensitivity to addition (ϕ), that is the score loss that they would experience if another outgoing link were to be added to their parent page. This means that, for instance, a *single* new link from the Wikipedia article *Ratio* to another page would significantly change the PageRank score of *Rate (Mathematics)*, reducing it by almost 20%.

5.4. Computational Approximations of Explanations

The counterfactual explanations that we introduced require the values of $\widehat{\sigma}_{+(\alpha,\beta)}$ to be computed. In particular, to generate the explanations for a node, the PR scores of the supporters of the node have to be assessed on the web graph with a single link changed, a computationally expensive operation that should be avoided when not necessary. To this purpose, with Proposition 10 we show that, under certain assumptions, it is not necessary to re-run PR on the whole web graph to compute the values of $\widehat{\sigma}_{+(\alpha,\beta)}$.

Figure 6 provides a graphical support to the proposition. It shows the relationships between the nodes α, β and δ_i used in the proposition itself and in its proof.

Proposition 10. *For* $\widehat{PR}_{+(\alpha,\beta)}$ *and* $\forall \delta_i \in \mathfrak{R}^+(\alpha)$ *it holds that:*

1. $|\widehat{\mathcal{R}}^-_{+(\alpha,\beta)}(m_{\alpha,\delta_i})| = |\widehat{\mathcal{R}}^-(m_{\alpha,\delta_i})| + 1$.
2. *If there is no support path*[2] *from* β *to* α *it holds that* $\widehat{\sigma}_{+(\alpha,\beta)}(m_{\alpha,\delta_i}) = \widehat{\sigma}(m_{\alpha,\delta_i}) - \frac{\widehat{\sigma}(m_{\alpha,\delta_i})}{|\widehat{\mathcal{R}}^-(m_{\alpha,\delta_i})|+2}$
3. *If there is no support path also from* β *to* δ *then it holds that* $\widehat{\sigma}_{+(\alpha,\beta)}(\delta_i) = \widehat{\sigma}(\delta_i) - d \cdot \frac{\widehat{\sigma}(\alpha)}{|\widehat{\mathcal{R}}^-(m_{\alpha,\delta_i})|\cdot(|\widehat{\mathcal{R}}^-(m_{\alpha,\delta_i})|+1)}$

Proof. Point 1. By Definition 5, it is immediate that $\forall \delta_i \in \mathfrak{R}^+(\alpha), |\widehat{\mathcal{R}}^-_{+(\alpha,\beta)}(m_{\alpha,\delta_i})| = |\widehat{\mathcal{R}}^-(m_{\alpha,\delta_i})| + 1$. *Point 2.* Given that, by Remark 3, α does not have any attacker and that there is no support path from β to α then the strength value $\widehat{\sigma}(\alpha)$ does not change when adding (α, β) to $\mathcal{R}^+$ because by Definition 6 $\widehat{\sigma}(\alpha)$ depends only on the strengths of its support-

[2]For any $\alpha, \beta \in \mathcal{X}$ there exists a support path from α to β iff $\exists \gamma_1, \ldots, \gamma_l$ such that $\gamma_1 = \alpha, \gamma_l = \beta, l > 0$ and $\forall i \in \{1, \ldots, l-1\}, (\gamma_i, \gamma_{i+1}) \in \mathcal{R}^+$. Note that, by construction of the MPRAF, this is equivalent to requiring $(\gamma_i, \gamma_{i+1}) \in \widehat{\mathcal{R}}^+$.

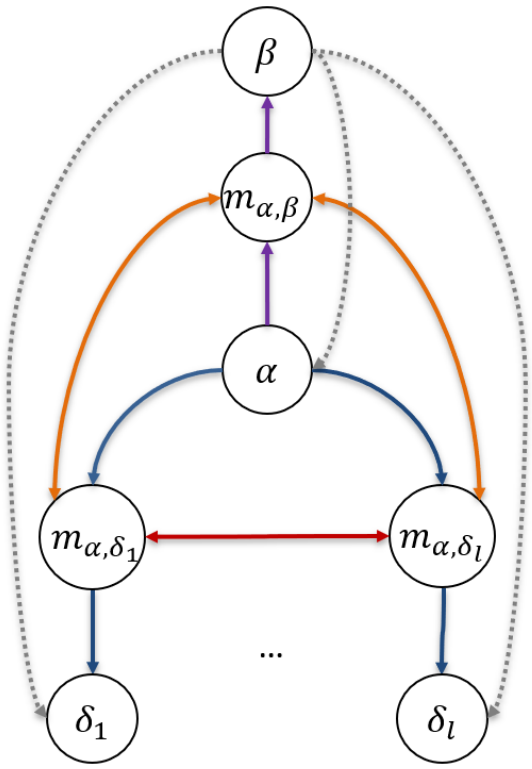

Fig. 6. Excerpt of an argumentation framework in support of the proof of Proposition 10. Red and blue arrows represent, respectively, attacks and supports; orange and violet represent, respectively, additional attacks and supports; grey arrows highlight forbidden paths. n_c is the number of children of α.

ers and attackers. Thus, given Definition 6 and Remark 4 (meta-arguments only have a single support), we can write $\widehat{\sigma}(m_{\alpha,\delta_i}) = \sqrt{d} \cdot \frac{\widehat{\sigma}(\alpha)}{|\widehat{\mathcal{R}}^-(m_{\alpha,\delta_i})|+1}$ and $\widehat{\sigma}_{+(\alpha,\beta)}(m_{\alpha,\delta_i}) = \sqrt{d} \cdot \frac{\widehat{\sigma}(\alpha)}{|\widehat{\mathcal{R}}^-_{+(\alpha,\beta)}(m_{\alpha,\delta_i})|+1}$. Now isolating $\widehat{\sigma}(\alpha)$ from the former and substituting it back into the latter we get $\widehat{\sigma}_{+(\alpha,\beta)}(m_{\alpha,\delta_i}) = \widehat{\sigma}(m_{\alpha,\delta_i}) \cdot \frac{|\widehat{\mathcal{R}}^-(m_{\alpha,\delta_i})|+1}{|\widehat{\mathcal{R}}^-_{+(\alpha,\beta)}(m_{\alpha,\delta_i})|+1}$. Using what we just proved in point 1 then $\widehat{\sigma}_{+(\alpha,\beta)}(m_{\alpha,\delta_i}) = \widehat{\sigma}(m_{\alpha,\delta_i}) \cdot \frac{|\widehat{\mathcal{R}}^-(m_{\alpha,\delta_i})|+1}{|\widehat{\mathcal{R}}^-(m_{\alpha,\delta_i})|+2}$ that is also equivalent to $\widehat{\sigma}_{+(\alpha,\beta)}(m_{\alpha,\delta_i}) = \widehat{\sigma}(m_{\alpha,\delta_i}) - \frac{\widehat{\sigma}(m_{\alpha,\delta_i})}{|\widehat{\mathcal{R}}^-(m_{\alpha,\delta_i})|+2}$. *Point 3*. Given Definition 6 and the hypothesis that $\nexists$ path also from β to δ_i, we can write $\widehat{\sigma}_{+(\alpha,\beta)}(\delta_i) = \frac{1-d}{|\mathcal{X}|} + +\sqrt{d} \cdot \frac{\sum_{m_{\zeta,\delta_i} \in \widehat{\mathcal{R}}^+_{+(\alpha,\beta)}(\delta_i)} \widehat{\sigma}_{+(\alpha,\beta)}(m_{\zeta,\delta_i})}{|\widehat{\mathcal{R}}^-(\delta_i)|+1}$. Given Remark 3 (non-meta-arguments have no attacks), we can rewrite the previous as $\widehat{\sigma}_{+(\alpha,\beta)}(\delta_i) = \frac{1-d}{|\mathcal{X}|} + +\sqrt{d} \cdot \sum_{m_{\zeta,\delta_i} \in \widehat{\mathcal{R}}^+_{+(\alpha,\beta)}(\delta_i)} \widehat{\sigma}_{+(\alpha,\beta)}(m_{\zeta,\delta_i})$. Now, if we isolate α's contribution to δ_i in the summation, we get $\widehat{\sigma}_{+(\alpha,\beta)}(\delta_i) = \frac{1-d}{|\mathcal{X}|} + \sqrt{d} \cdot \left[\left(\sum_{m_{\zeta,\delta_i} \in \widehat{\mathcal{R}}^+(\delta_i)} \widehat{\sigma}(m_{\zeta,\delta_i}) \right) - -\widehat{\sigma}(m_{\alpha,\delta_i}) + \widehat{\sigma}_{+(\alpha,\beta)}(m_{\alpha,\delta_i}) \right]$. And given that $\widehat{\sigma}(\delta_i) = \frac{1-d}{|\mathcal{X}|} + \sqrt{d} \cdot \sum_{m_{\zeta,\delta_i} \in \widehat{\mathcal{R}}^+(\delta_i)} \widehat{\sigma}(m_{\zeta,\delta_i})$ then it holds that $\widehat{\sigma}_{+(\alpha,\beta)}(\delta_i) = \widehat{\sigma}(\delta_i) + \sqrt{d} \cdot [-\widehat{\sigma}(m_{\alpha,\delta_i}) + +\widehat{\sigma}_{+(\alpha,\beta)}(m_{\alpha,\delta_i})]$. Using what we proved in point 2, we can rewrite as $\widehat{\sigma}_{+(\alpha,\beta)}(\delta_i) = \widehat{\sigma}(\delta_i) + +\sqrt{d} \cdot \left(-\widehat{\sigma}(m_{\alpha,\delta_i}) + \widehat{\sigma}(m_{\alpha,\delta_i}) \cdot \frac{|\widehat{\mathcal{R}}^-(m_{\alpha,\delta_i})|+1}{|\widehat{\mathcal{R}}^-(m_{\alpha,\delta_i})|+2} \right)$ or equivalently as $\widehat{\sigma}_{+(\alpha,\beta)}(\delta_i) = \widehat{\sigma}(\delta_i) - -\widehat{\sigma}(m_{\alpha,\delta_i}) \cdot \frac{\sqrt{d}}{|\widehat{\mathcal{R}}^-(m_{\alpha,\delta_i})|+2}$. □

In practice, this proposition has a twofold use. On the one hand, it provides a computationally efficient procedure to compute $\widehat{\sigma}_{+(\alpha,\beta)}$ under certain assumptions. On the other hand, the same procedure can possibly be used as an estimation method when those assumptions do not hold. We denote such estimator as $\widehat{\sigma}^e_{+(\alpha,\beta)}$.

$$\widehat{\sigma}^e_{+(\alpha,\beta)} = \widehat{\sigma}(\delta) - d \cdot \frac{\widehat{\sigma}(\alpha)}{|\widehat{\mathcal{R}}^-(m_{\alpha,\delta})| \cdot (|\widehat{\mathcal{R}}^-(m_{\alpha,\delta})| + 1)}$$

In Section 6 will use this procedure to generate additive counter-factual explanations and we will also show empirically that this is a good estimator for $\widehat{\sigma}_{+(\alpha,\beta)}$ in terms of approximation error.

6. Experiments

In this section we evaluate the proposed explanations on the Wikipedia web graph and several other web graphs crawled from some British and Irish universities, see Table 2 for information. In particular, we aim to address the following research questions:

- Misleading basic explanations: do basic explanation provide a misleading picture of the pages contributing the most to the PR score of a page?
- Cognitive tractability of explanations: Are explanations cognitively tractable? In particular:
  - (A) What is the size of explanations?
  - (B) How many arguments must be included in explanations to best explain page scores?
- Contrastive attribution explanations usefulness: which portion of supporters is shared between two pages?

Table 2

Characteristics of the web graphs used in the experiments. (†) These web graphs result from a partial crawl, i.e. the crawling of these websites was stopped before it completed after running for more than 1 month.

Web Graph	Website	Number of pages	Number of links	Average number of links per page
Leeds Trinity University	www.leedstrinity.ac.uk	1,119	35,011	31.28
Homerton College	www.homerton.cam.ac.uk	1,261	55,903	44.33
National University of Ireland	www.nui.ie	1,295	11,632	8.98
University of East Anglia	www.uea.ac.uk	2,871	120,735	42.05
Cardiff Metropolitan University	www.cardiffmet.ac.uk	4,467	107,354	24.03
University of Leeds	www.leeds.ac.uk	25,824	180,196	6.97
Queen's University Belfast	www.qub.ac.uk	64,659	451,882	6.98
University College Dublin	www.ucd.ie	81,893	1,129,103	13.78
University of Exeter †	www.exeter.ac.uk	118,005	1,859,492	15.75
Imperial College London †	www.imperial.ac.uk	146,125	4,758,430	32.56
University of Reading	www.reading.ac.uk	302,130	7,063,264	23.37
London School of Economics †	www.lse.ac.uk	426,434	2,622,280	6.14
University of Oxford †	www.ox.ac.uk	430,490	5,429,420	12.61
Wikipedia	simple.wikipedia.org	965,748	7,388,700	7.65

Table 3

Average divergence ratio of the strength of arguments (in the basic explanations) and meta-arguments (in attribution explanations).

	Leeds Trinity University	Homerton College	National University of Ireland	University of East Anglia	Cardiff Metropolitan University	University of Leeds	Queen's University Belfast	University College Dublin	University of Exeter	Imperial College London	University of Reading	London School of Economics	University of Oxford	Wikipedia
$r_{\%}$	35.2	41	99.2	40.2	17.2	25.6	701.4	51	523.7	1324.9	97.9	34.3	56.3	73.8

– Approximation: is the estimator $\widehat{\sigma}^{e}_{+(\alpha,\beta)}$ based on Proposition 10 a good approximation for $\widehat{\sigma}_{+(\alpha,\beta)}$, i.e., for the M-PR semantics on the MPRAF with the addition of a support?

Note that, when conducting experiments on contrastive attribution explanations and additive counterfactual explanations, we randomly sampled 500,000 and 200,000 pairs of pages, respectively, for performance reasons; in all other experiments we used all the pages instead.

Misleading basic explanations. To assess if and how often basic explanations provide a misleading picture of the pages contributing to the score of a page of interest, we checked the divergence ratio of the strengths of arguments in basic explanations and meta-arguments in attribution explanations for the same page. We denote with $\sigma_{\%}(\beta,\alpha)$ and $\widehat{\sigma}_{\%}(\beta,\alpha)$ the contribution to page $\alpha \in \mathcal{X}$ from a supporter $\beta \in \mathcal{R}^{+}(\alpha)$ according to, respectively, the basic explanation and the attribution explanation, i.e., $\sigma_{\%}(\beta,\alpha) = \frac{\sigma(\beta)}{\sum_{\gamma \in E_b} \sigma(\gamma)}$ and $\widehat{\sigma}_{\%}(\beta,\alpha) = \frac{\widehat{\sigma}(m_{\beta,\alpha})}{\sum_{\gamma \in E_{\leftarrow}} \widehat{\sigma}(m_{\gamma,\alpha})}$. We then define the divergence ratio, denoted with $r_{\%}$, as $r_{\%}(\beta,\alpha) = \left|\frac{\sigma_{\%}(\beta,\alpha)}{\widehat{\sigma}_{\%}(\beta,\alpha)} - 1\right|$, describing how much the contribution of a supporter β is under or over-estimated in a basic explanation of α wrt to its actual contribution in the attribution explanation. Table 3 shows that the divergence ratio ranges between 35% and 1324% in our experiments. This means that the picture portrayed by basic explanation can be very misleading in some web graphs.

Table 4

Average size of attribution explanations ($E_{\leftarrow}$), contrastive attribution explanations ($E_{\leftrightarrow}$), additive counter-factual explanations ($E_{\underset{\leftarrow}{+?}}$) and edit-sensibility counterfactual explanations ($E_{\underset{\leftarrow}{\Phi?}}$). (†) Additive counterfactual explanations' sizes are equal to those of attribution explanations.

Explanation	$E_{\leftarrow}$	$E_{\leftrightarrow}$	$E_{\underset{\leftarrow}{+?}}$	$E_{\underset{\leftarrow}{\Phi?}}$
Leeds Trinity University	177.1	166.1	259.3	†
Homerton College	228	223.9	347.8	†
National University of Ireland	229.8	228.5	245	†
University of East Anglia	456.4	448.8	531.1	†
Cardiff Metropolitan University	868	849.4	979.8	†
University of Leeds	3887	3884.8	3973.1	†
Queen's University Belfast	8710.8	8704.2	9353.1	†
University College Dublin	12605.3	12605.9	12765.6	†
University of Exeter	18495.1	18493.5	19310.4	†
Imperial College London	21072.1	21070.6	22313.4	†
University of Reading	47246.9	47237.1	37557.4	†
London School of Economics	61911	61663.7	33907.1	†
University of Oxford	63804.7	63778.4	36124.3	†
Wikipedia	123495.6	59340	38594.3	†

Table 5

Percentages of PageRank score explained by the top meta-arguments in explanations according to their importances.

Explanation	Supporters	Leeds Trinity University	Homerton College	National University of Ireland	University of East Anglia	Cardiff Metropolitan University	University of Leeds	Queen's University Belfast	University College Dublin	University of Exeter	Imperial College London	University of Reading	London School of Economics	University of Oxford	Wikipedia
$E_{\leftarrow}$	top-1	73.4	83.7	88.2	84.5	76.9	82.2	86	79.8	82.1	79.1	67.2	90.5	80.6	82.3
	top-3	81	87.4	97.1	92.1	90	94.6	94.1	90.4	91.3	88.9	85.6	97.7	92.6	92.9
	top-5	83.9	88	98	92.6	93	96.2	95.9	92.6	93.7	91	92.3	98.8	95.7	95
	top-10	84.8	89.1	98.9	93.1	96.2	97.8	97.5	95.2	96	93.1	96.3	99.4	97.8	96.9
$E_{\underset{\leftarrow}{\Phi?}}$	top-1	74.6	86.2	90.7	85.9	77	83.6	87.1	80.9	83.5	80.8	69.4	92.3	82.9	85.7
	top-3	81.4	89.7	97.2	92.6	89.9	94.9	94.5	90.8	91.9	89.2	87.3	98.2	93.8	94.4
	top-5	84	90.3	98.1	93.1	93.1	96.4	96.1	93	94.2	91.1	93.6	99	96.4	96.1
	top-10	84.9	91.2	99	93.6	96.3	98	97.7	95.4	96.3	93.1	96.9	99.4	98	97.5
$E_{\leftrightarrow}$	top-1	96.8	90	90.7	91.4	85.5	92.6	93.1	83.3	88.9	75.3	66.1	88.9	84.3	90.2
	top-3	97.9	91.1	98.9	94.7	95.9	97.7	97.5	94.4	96	86.8	83.5	96.7	96.1	98.5
	top-5	98.1	91.4	99.4	95.2	97.9	98.6	98.4	95.3	97.5	89	91	97.9	98.2	99.2
	top-10	98.2	91.9	99.5	95.4	99	99.7	99.1	96.4	98.3	91.6	95.2	98.6	99.4	99.7
$E_{\underset{\leftarrow}{+?}}$	top-1	73.4	83.7	88.2	84.5	76.9	82.2	86	79.8	82.1	79.1	67.2	90.5	80.6	82.3
	top-3	81	87.4	97.1	92.1	90	94.6	94.1	90.4	91.3	88.9	85.6	97.7	92.6	92.9
	top-5	83.9	88	98	92.6	93	96.2	95.9	92.6	93.7	91	92.3	98.8	95.7	95
	top-10	84.8	89.1	98.9	93.1	96.2	97.8	97.5	95.2	96	93.1	96.3	99.4	97.8	96.9

Table 6

Percentages of shared supporters in contrastive attribution explanations of (1) two random pages, (2) a page and one of its supporters and (3) two pages mutually supporting each other.

	Leeds Trinity University	Homerton College	National University of Ireland	University of East Anglia	Cardiff Metropolitan University	University of Leeds	Queen's University Belfast	University College Dublin	University of Exeter	Imperial College London	University of Reading	London School of Economics	University of Oxford	Wikipedia
Random pages (%)	45.6	25.7	21.1	20.7	3.7	0.5	3.2	0.6	4.3	1.8	0.1	0.1	0.2	0
Supporting pages (%)	27.1	17.1	9.7	25.4	18.7	13.7	8.8	14.7	20.8	12.5	25.4	5.4	10.7	12.1
Mutually supporting pages (%)	96.2	63.4	50.3	70.3	59.1	47.4	49.4	62	55.5	38.7	67.3	37.7	55	42

Table 7

Approximation error of $\widehat{\sigma}^{e}_{+(\alpha,\beta)}$, i.e., $\left|\frac{\widehat{\sigma}^{e}_{+(\alpha,\beta)}-\widehat{\sigma}_{+(\alpha,\beta)}}{\widehat{\sigma}}\right|$.

	Leeds Trinity University	Homerton College	National University of Ireland	University of East Anglia	Cardiff Metropolitan University	University of Leeds	Queen's University Belfast	University College Dublin	University of Exeter	Imperial College London	University of Reading	London School of Economics	University of Oxford	Wikipedia
Average (%)	0.006	0.002	0.006	0.005	0.014	0.008	0.012	0.024	0.007	0.019	0.006	0	0.002	0.005
Maximum (%)	0.415	0.009	0.066	0.242	0.906	0.25	0.633	1.255	0.091	0.566	0.126	0.009	0.06	0.105

Cognitive tractability. In order to assess the cognitive tractability of the explanations we proposed, we checked the size of explanations, in terms of overall number of arguments and the percentage of a score explained by the top arguments according to their importance in the explanation. Table 4 and Table 5 show the results on the sizes of the explanations and the percentage of score explained by the top arguments, respectively. We note that: (1) the average explanation size ranges from some hundreds of arguments for the smaller web graphs, to hundreds of thousands for the bigger ones, significantly increasing with the number of pages and links in the web graph; (2) selecting only arguments with backward hop distance of 2 in additive counter-factual explanation considerably reduces the number of arguments in the explanations to an amount similar to that of other types of explanations. This is a reasonable amount when compared to the number of nodes in the web graph that would have been otherwise included. (3) Although the full set of meta-arguments potentially included in the explanations of a page may be very large, considering only a limited subset is enough to produce a satisfactory explanation. In fact, 10 meta-arguments are enough to explain on average between 84.8% and 99.7% of the score of a page depending on the web graph and the type of explanation.

Contrastive attribution explanation usefulness. Contrastive attribution explanations are useful only when the amount of shared supporters is not negligible. We checked therefore the average number of shared supporters in different scenarios. Table 6 shows the results. We note that: (1) the average number of shared supporters of two random pages can be more than 45% in some (smaller) datasets; (2) despite the small average number of shared supporters for two random pages in some datasets, in other scenarios where one of the two pages supports the other or they support one an-

other, the number of shared supporters increases up to 96.2%.

Approximation of $\widehat{\sigma}_{+(\alpha,\beta)}$. We checked whether $\widehat{\sigma}^{e}_{+(\alpha,\beta)}$ is a good estimator for $\widehat{\sigma}_{+(\alpha,\beta)}$ on 50 random pairs of pages in each dataset. As shown by Table 7, the average approximation error is at most 0.024% and the maximum error ranges between 0.009% and 1.255% depending on the dataset. These low values confirm that the approximation provided by $\widehat{\sigma}^{e}_{+(\alpha,\beta)}$ for $\widehat{\sigma}_{+(\alpha,\beta)}$ is satisfactory.

7. Conclusions

In this paper we have investigated connections between PageRank (PR) and formal argumentation. Firstly, we have introduced a novel approach capable of reconstructing PR as a gradual argumentation semantics of a suitably defined bipolar argumentation framework, while ensuring the satisfaction of a set of generally desirable properties. Secondly, we have shown how using this approach enables the generation of better explanations of PR scores to end-users, proposing four different types of explanation.

To the best of our knowledge, the investigation of the relationships between PR and argumentation semantics has not been previously considered in the literature. The work in [30] explores the application of PR to rank the relevance of arguments available on the web to support or attack a given stance. This is an interesting but different goal: in [30] PR is not related to any semantics notion and the links have a different meaning, relating the conclusion of an argument with the premises of another one. On a different but related line, some works, e.g. [31], have explored connections between argumentation semantics and matrix representations from network theory, whose relationships with our approach are worth future investigation. To the best of our knowledge, the generation of explanations based on argumentation for PR has not been previously considered in the literature. We have illustrated the promise of our method in helping users to better understand PR, a popular algorithm for ranking pages, but leave user evaluations to future work.

Our proposal can be extended mainly in two directions. The role of bipolar argumentation framework representation with meta-arguments in enhancing the explainability of graph-based algorithms could be further investigated. In this regard, understanding how other algorithms designed for directed graphs could be re-interpreted in an argumentative perspective and developing other types of explanations from their argumentative counterparts represent two interesting research possibilities. Another fruitful direction would be the investigation of the relation between PR and argumentation semantics could be expanded. In this respect, firstly, the investigation of PR-inspired gradual semantics for various kinds of argumentation frameworks could be pursued. For example, it would be interesting to consider *weighted* versions of PR where a node's strength can be distributed unevenly to its children and, more generally, to the variants of PR considered in various domains [6]. Secondly, one can notice that PR is essentially a mechanism to produce a score based on a relation of support, but it could be considered that, in several domains where PR is applied, also other relations, in particular attack, could be relevant for a proper scoring. Also, in the web domain, one could argue that the absence of a link from one page to another (where this link could instead be expected according to some criterion) could be interpreted as an attack diminishing the relevance of the non-linked page. Given the strong tradition on attack-based and bipolar evaluations in argumentation semantics, this suggests that the study of argumentation-inspired variants of PR may also represent a fruitful research direction.

References

[1] Page L, Brin S, Motwani R, Winograd T. The PageRank Citation Ranking: Bringing Order to the Web. World Wide Web Internet And Web Information Systems. 1998;54(1999-66):1–17.

[2] Ma N, Guan J, Zhao Y. Bringing PageRank to the citation analysis. Information Processing and Management. 2008 3;44(2):800–810.

[3] Gori M, Pucci A. ItemRank: A Random-Walk Based Scoring Algorithm for Recommender Engines. In: Proceedings of the 20th International Joint Conference on Artificial Intelligence (IJCAI); 2007. p. 2766–2771.

[4] Hudelson M, Mooney BL, Clark AE. Determining polyhedral arrangements of atoms using PageRank. Journal of Mathematical Chemistry. 2012 9;50(9):2342–2350.

[5] Morrison JL, Breitling R, Higham DJ, Gilbert DR. GeneRank: Using search engine technology for the analysis of microarray experiments. BMC Bioinformatics. 2005;6(1):233.

[6] Gleich DF. PageRank beyond the web. SIAM Review. 2015;57(3):321–363.

[7] Altman A, Tennenholtz M. Ranking systems: the PageRank axioms. In: Proceedings of the 6th ACM Conference on Electronic Commerce (EC); 2005. p. 1–8.

[8] Dung PM. On the Acceptability of Arguments and its Fundamental Role in Nonmonotonic Reasoning, Logic Programming and n-Person Games. Artificial Intelligence. 1995;77(2):321–358.

[9] Cayrol C, Lagasquie-Schiex MC. On the Acceptability of Arguments in Bipolar Argumentation Frameworks. In: Proceedings of the 8th European Conference on Symbolic and Quantitative Approaches to Reasoning with Uncertainty (ECSQARU); 2005. p. 378–389.

[10] Baroni P, Rago A, Toni F. From fine-grained properties to broad principles for gradual argumentation: A principled spectrum. International Journal of Approximate Reasoning. 2019;105:252–286.

[11] Baroni P, Caminada M, Giacomin M. An introduction to argumentation semantics. Knowledge Engineering Review. 2011;26(4):365–410.

[12] Cayrol C, Lagasquie-Schiex MC. Graduality in Argumentation. Journal of Artificial Intelligence Research. 2005;23:245–297.

[13] Ribeiro MT, Singh S, Guestrin C. "Why Should I Trust You?". In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD); 2016. p. 1135–1144.

[14] Lundberg SM, Allen PG, Lee SI. A Unified Approach to Interpreting Model Predictions. In: Proceedings of the 31st International Conference on Neural Information Processing Systems (NeurIPS); 2017. p. 4768–4777.

[15] Dhurandhar A, Chen PY, Luss R, Tu CC, Ting PS, Shanmugam K, et al. Explanations based on the Missing: Towards Contrastive Explanations with Pertinent Negatives. In: Proceedings of the 31st Conference on Neural Information Processing Systems (NeurIPS); 2018. p. 590–601.

[16] Wachter S, Mittelstadt B, Russell C. Counterfactual Explanations Without Opening the Black Box: Automated Decisions and the GDPR. SSRN Electronic Journal. 2017:1–52.

[17] Mittelstadt BD, Russell C, Wachter S. Explaining Explanations in AI. In: Proceedings of the 2nd Conference on Fairness, Accountability, and Transparency (FAT*); 2019. p. 279–288.

[18] Albini E, Lertvittayakumjorn P, Rago A, Toni F. DAX: Deep Argumentative eXplanation for Neural Networks. ArVix; 2020. Available from: http://arxiv.org/abs/2012.05766.

[19] Dejl A, He P, Mangal P, Mohsin H, Surdu B, Voinea E, et al. Argflow: A Toolkit for Deep Argumentative Explanations for Neural Networks. In: Proceedings of the 20th International Conference on Autonomous Agents and Multiagent Systems (AAMAS); 2021. .

[20] Cyras K, Letsios D, Misener R, Toni F. Argumentation for Explainable Scheduling. In: Proceedings of the 33rd AAAI Conference on Artificial Intelligence (AAAI); 2019. p. 2752–2759.

[21] Timmer ST, Meyer JC, Prakken H, Renooij S, Verheij B. Explaining Bayesian Networks Using Argumentation. In: Proceedings of the 13th European Conference on Symbolic and Quantitative Approaches Reasoning with Uncertainty (ECSQARU); 2015. p. 83–92.

[22] Rago A, Albini E, Baroni P, Toni F. Influence-Driven Explanations for Bayesian Network Classifiers. ArVix; 2020. Available from: http://arxiv.org/abs/2012.05773.

[23] Arioua A, Tamani N, Croitoru M. Query Answering Explanation in Inconsistent Datalog +/- Knowledge Bases. In: Chen Q, Hameurlain A, Toumani F, Wagner R, Decker H, editors. Proceedings of the 26th International Conference on Database and Expert Systems Applications (DEXA). vol. 9261 of Lecture Notes in Computer Science; 2015. p. 203–219.

[24] Rago A, Cocarascu O, Bechlivanidis C, Toni F. Argumentation as a Framework for Interactive Explanations for Recommendations. In: Proceedings of the 17th International Conference on Principles of Knowledge Representation and Reasoning (KR); 2020. p. 805–815.

[25] Amgoud L, Ben-Naim J. Evaluation of Arguments from Support Relations: Axioms and Semantics. In: Proceedings of the 25th International Joint Conference on Artificial Intelligence (IJCAI); 2016. p. 900–906.

[26] Albini E, Baroni P, Rago A, Toni F. PageRank as an argumentation semantics. In: Proceedings of the 8th International Conference on Computational Models of Argument (COMMA); 2020. p. 55–66.

[27] Albini E, Baroni P, Rago A, Toni F. Explaining PageRank through Argumentation. In: Workshop on Explainable Logic-Based Knowledge Representation (Co-located with the 17th International Conference on Principles of Knowledge Representation and Reasoning, KR); 2020. .

[28] Langville AN, Meyer CD. Deeper inside PageRank. Internet Mathematics. Internet Mathematics. 2004;1(3):335–380.

[29] Baroni P, Rago A, Toni F. How many properties do we need for gradual argumentation? In: Proceedings of the 32nd AAAI Conference on Artificial Intelligence (AAAI); 2018. p. 1736–1743.

[30] Wachsmuth H, Stein B, Ajjour Y. "PageRank" for Argument Relevance. In: Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics (EACL); 2017. p. 1117–1127.

[31] Corea C, Thimm M. Using Matrix Exponentials for Abstract Argumentation. In: Proceedings of the 1st International Workshop on Systems and Algorithms for Formal Argumentation (SAFA); 2016. p. 10–21.