Explaining PageRank through Argumentation

Emanuele Albini[1 2 [0000−0003−2964−4638]], Pietro Baroni[1 [0000−0001−5439−9561]], Antonio Rago[2 [0000−0001−5323−7739]], and Francesca Toni[2 [0000−0001−8194−1459]]

[1] Dip.to di Ingegneria dell'Informazione, Università degli Studi di Brescia, IT
pietro.baroni@unibs.it
[2] Dept. of Computing, Imperial College London, UK
{emanuele.albini19,a.rago15,f.toni}@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. To this purpose we propose several types of explanation, each of which focuses on different aspects of the algorithm and uncovers information useful for the comprehension of its results.

Keywords: PageRank · Explainability · Bipolar Argumentation Framework.

1 Introduction

Providing explanations (XPs) of the outcomes of an algorithm is crucial for its users. For instance, they may allow users to understand *why an algorithm gives certain output* (e.g. attribution methods), to assess *which components of the input led to different outcomes* (e.g. contrastive XPs) or *how they could interact with an algorithm to change its output* (e.g. counterfactual XPs); for an overview see [1].

In the context of computational argumentation, one way to generate such explanations is through the *graphical* representation of the reasoning behind algorithms. We utilise *Quantitative Bipolar Argumentation Frameworks* [2] (QBAFs), essentially directed graphs whose nodes are arguments and edges represent attacks or supports between arguments. In this context, the argument graph structure is the basis of the numerical assessment of argument *strength* according to a *gradual semantics* [3, 2, 4].

In this paper, we focus our attention on *PageRank* [5] (PR), a popular ranking algorithm for web pages that, given its general formalisation for any kind of directed graph, has been applied to many other domains including citation networks, recommender systems, chemistry, biology, and neuroscience [6]. PR is based on a representation of the web as a directed graph, in which nodes are the pages and edges are the links, and assigns to each page a score that describes how relevant it is: the higher the score, the more important the page. These scores are determined through a mathematical model aiming at probabilistically estimating the number of users visiting the page. The assumption here is that the higher the number of links to (from) a page, the more it (the less each page linked by it, resp.) will be visited and hence the higher (lower, resp.) its PR score should be.

PR [5] can be directly interpreted, from an argumentation perspective, as a gradual semantics for a QBAF in which pages are arguments and links are supports. This naive approach does not give rise to a satisfactory outcome in terms of formal properties of the semantics and leads to counterintuitive results from a dialectical perspective. Thus, we defined a novel QBAF-based representation in which we introduce meta-arguments representing links and a gradual semantics reproducing the behavior of PR in this context [7]. This gradual semantics produces a *strength* value for each argument satisfying desirable theoretical properties and empowers the generation of various types of XPs for PR that emphasise different aspects of its underlying mechanism. For the sake of brevity, we omit the details of the semantics' formulation (see [7] for more details) and the formal definitions of the XPs. We focus instead on some examples of the XPs we generated illustrating their practical value and some empirical results. To explore the different notions of XPs in the context of PR, we used several datasets, focusing in this paper on the **Wikipedia** dataset[3] consisting of 965,748 pages and 7,388,700 links.

2 Argumentation-based Explanations for PageRank

In this section we evidence the limits of XPs for PR based on the original web graph and propose several novel XPs utilising the QBAF with meta-arguments mentioned above. We will highlight, when relevant, how the argumentative perspective allows us to leverage the QBAF representation and thus to generate XPs that give the user a better understanding of the algorithm. In particular we will consider XPs allowing the user to understand the reasons of a given score, providing hints on changes that can improve the score, or giving warnings on strong dependencies of the score on other pages.

The above mentioned analyses cannot be supported by explanations based on the web graph, which we call **basic explanations** in that they are solely based on the output of PR, i.e. the score of each page. Consider the problem of identifying which pages have a major role in determining the score of a given page one is interested in. Using the web graph and PR scores only, one could produce an XP like the one presented in Fig. 1.i. This is a magnification of the weighted view of the pages contributing to the score of the article *Nguyen Dynasty* in the web graph, with each page score being represented by the size of the relevant bubble. Notice how, looking at this XP, 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*). In fact, XPs 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*?'. To answer this question using the web graph representation a user should both have a deeper understanding of how PR works and be shown a larger part of the web graph, including all the pages linked by the supporters of the considered page. Only then the user might realize that *Hue*, instead of *Official Residence*, is the article providing most support to *Nguyen Dynasty*.

Attribution explanations based on QBAFs solve this problem, and focus the attention 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?'. Fig. 1.iii, shows an example of attribution explanation for the article *Nguyen Dynasty* as an excerpt of the QBAF

[3] *Wikipedia simple* from Wikipedia Dumps `https://dumps.wikimedia.org/`

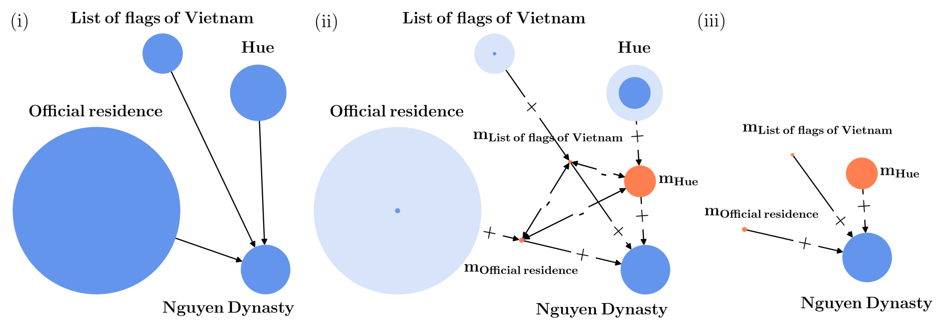

Fig. 1. Transition, for the article *Nguyen Dynasty*, from its basic explanation (i), to the excerpt of the QBAF including it and its supporters, to eventually its attribution explanation (iii). Each bubble represents an argument and its size is proportional the the strength of the argument. In (ii) the opaque bubbles highlight the actual contribution of an argument to the *Nguyen Dynasty* page. $-$ and $+$ indicate attacks and supports, resp..

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 one. In this representation it is clear that the contribution of *Hue* to the score of *Nguyen Dynasty* is bigger than the one of *Official Residence*, despite its lower PR score. Although the full set of meta-arguments potentially included in the attribution explanation of a page may be very large (in the order of hundreds); our experiments showed that considering only a limited subset is enough to produce a satisfactory explanation: 10 meta-arguments are enough to explain on average 96.24% of the score of a page in the Wikipedia dataset. This property allows our XPs to fulfil the desideratum of simplicity, avoiding overwhelming the user with too much information when the number of supporters is large.

Besides evidencing the actual contribution of each supporting page, attribution explanations may address other kinds of user queries, e.g. counterfactual questions: 'What would happen if a given link is suppressed?'. In this context, meta-arguments strengths directly show the portion of the score that a page would lose if a link were removed.

When two (or more) pages have many shared supporters, understanding why the pages have different scores through attribution explanation is not trivial. **Contrastive attribution explanations** tackle this issue in a different manner to the attribution explanation: they show for each page of interest the nodes contributing exclusively to its score, ignoring the ones they share. Thus, these XPs answer questions of the kind: 'What are the links that make pages p and q have different scores?'. XPs 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(s), comparing their non-shared supporters sorted by their strengths. Fig. 2 shows an example of contrastive attribution explanation for the Wikipedia articles *Calorimeter* and *Spectrophotometer*. Using this XP users can focus on the differences in the supporters of two pages rather than on their common supporters that in this example amount to 40 nodes. Our experiments on 100,000 pairs

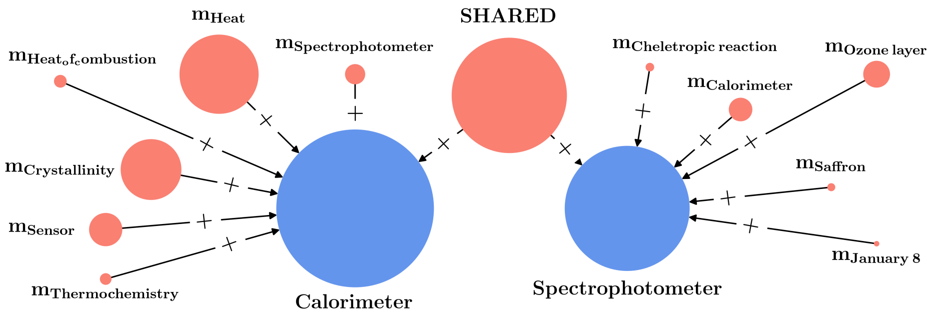

Fig. 2. Contrastive attribution explanation for the articles *Calorimeter* and *Spectrophotometer* from the Wikipedia dataset. The bubble labeled *SHARED* encompasses the contributions from the 40 shared supporters.

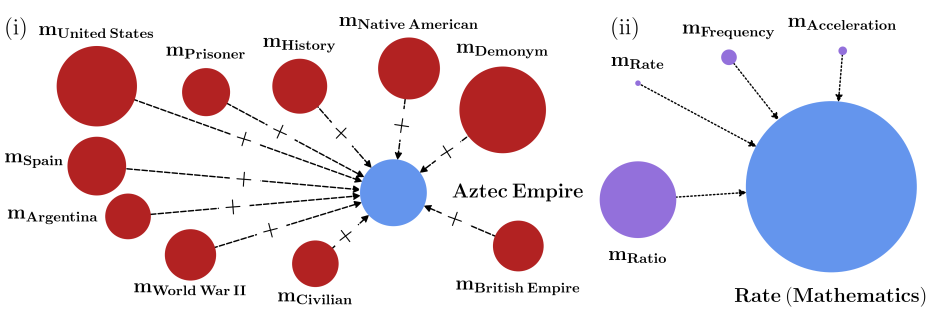

Fig. 3. Additive counterfactual explanation for the article *Aztec Empire* (i) and edit-sensibility counterfactual explanation for the article *Rate (Mathematics)* (ii).

of pages with links to each other showed that on average 42% of the supporters of the page having the fewest number of them are shared between the two pages.

Additive counterfactual explanations extend our effort in answering counter-factual questions in that they provide the user with information on links, not currently present, that if added would increase the score of a page of interest. These XPs answer the question 'In which pages a link to p could be added to maximize the increment of its score?'. A typical usage scenario of this XP is searching for pages that could be modified to increase the score of a specific page. Clearly the set of pages satisfying the condition we just mentioned is potentially huge. This means that, for this XP to be useful in practice, we have to define some relevance condition that arguments must satisfy to be included in the XP. As a preliminary choice, 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. This allowed us to select on average 170 meta-arguments: a reasonable amount when compared to the 965,748 nodes in the dataset. Fig. 3.i shows an example of this type of XP for the article *Aztec Empire* that visualizes a graph of the 10 most influential meta-arguments (selected from 515) that could be added to increase the most the score of *Aztec Empire*.

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 XP answers the question 'If an outgoing link is added to page q (a supporter of page p), how much the score of p will change?'. In other words, this XP highlights how much a page score is "exposed" to endogenous changes in the "link structure" of other pages; our experiments showed that a single change to the "most sensitive" supporter of a page leads to a reduction of its score of up to 50% and on average of about 2.15%. Fig. 3.ii shows an example of this XP for the article *Rate (Mathematics)*. Here, the sizes of the supporting meta-arguments are proportional to the score loss that they would experience (and therefore also that of the page *Rate (Mathematics)*) if another link were to be added to their parent page. This means that, for instance, a *single* new link from the article *Ratio* to another page would significantly change the score of *Rate (Mathematics)*, reducing it by almost 20%.

3 Conclusion

In this paper we have shown how using an approach able to reconstruct PR as a gradual argumentation semantics of a suitably defined bipolar argumentation framework enables the generation of better explanations of PR scores to end users. As to 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.

Our proposal can be extended mainly in two directions. On one hand, the development of other types of explanations for PR can be investigated. On the other hand, the QBAF-based representation with meta-arguments might be applied to other algorithms designed for directed graphs. In particular, the extension of this approach to variants of PR for domains other than the web [6] might be explored.

References

1. B. D. Mittelstadt, C. Russell, and S. Wachter. Explaining explanations in AI. In *Proc. of the 2nd Conference on Fairness, Accountability, and Transparency, FAT* 2019*, pages 279–288. ACM, 2019.
2. P. Baroni, A. Rago, and F. Toni. From fine-grained properties to broad principles for gradual argumentation: A principled spectrum. *Int. J. of Approximate Reasoning*, 105:252–286, 2019.
3. C. Cayrol and M.-C. Lagasquie-Schiex. Graduality in Argumentation. *J. of Artificial Intelligence Research*, 23:245–297, 2005.
4. P. Baroni, M. Caminada, and M. Giacomin. An introduction to argumentation semantics. *Knowledge Engineering Review*, 26(4):365–410, 2011.
5. L. Page, S. Brin, R. Motwani, and T. Winograd. The PageRank Citation Ranking: Bringing Order to the Web. Technical report, Stanford InfoLab, 1998.
6. D. F. Gleich. PageRank beyond the web. *SIAM Review*, 57(3):321–363, 2015.
7. E. Albini, P. Baroni, A. Rago, and F. Toni. PageRank as an argumentation semantics. In *Proc. of the 8th International Conference on Computational Models of Argument, COMMA*, 2020.