An Algorithm for Ranking Authors

Muhammad Nur Yanhaona[1⋆], Farig Yousuf Sadeque[1⋆⋆], Mohammad Kaykobad[1⋆⋆⋆], and M Tanvir Kaykobad[2†]

[1] BRAC University, Dhaka, Bangladesh
[2] Department of Computing, Queen's University, Kingston, ON, Canada

Abstract. In this paper, we introduce an algorithm for ranking authors and researchers based on their achievements in terms of significant awards and published articles. Authors are usually credited with the number of scientific citations they attract from other researchers. Unfortunately, this metric does not truly reflect scientific achievements. There are areas of scientific research that are too difficult for many researchers to assimilate and cite. Again, there are fields in which it is not difficult to cite published articles without really appreciating the matter to the depth. Hence, the citation count metric may be confusing. With a growing number of journals and conferences and their online versions, including the presence of predatory journals, it has also become very difficult to maintain standards of publication. Thus, citation counts of papers can be arbitrarily boosted. All of these suggest the need for a better research ranking mechanism. Fortunately, in almost all fields; recognitions, prizes, and awards are massively sought after by researchers. We are introducing a metric that will be induced to all researchers through initially benchmarking renowned researchers like Nobel Laureates, Fields medalists, Turing award winners, Abel and other prize recipients, quality of whose contributions are beyond questions. Then other researchers will get some weight through the esteemed researchers by being cited by the latter. We are in search of a metric in which we can avoid the curse of too many citations for publications yet to warrant any credit. At the same time, accomplished researchers like the first ever female Fields Medalist, who attracted barely a thousand citations at a time when she received the Fields Medal, should be adequately recognized.

Keywords: Ranking Algorithm

1 Introduction

With so much information being generated every year, surpassing the total information generated by mankind, it is becoming increasingly difficult to find

⋆ nur.yanhaona@bracu.ac.bd
⋆⋆ farig.sadeque@bracu.ac.bd
⋆⋆⋆ Corresponding Author: kaykobad@bracu.ac.bd
† m.kaykobad@queensu.ca

information worthy of processing. With rapid technological progresses, bandwidth of information is increasing every year. As new fields of research and study are introduced, more knowledge workers are needed to process them. Many researchers are needed to dedicate their time and effort on intersecting fields. Journals are published and conferences are being held with unprecedented frequencies. It has become tougher to maintain quality of research and discover repetition of research or plagiarism. With the presence of more and more researchers, it has become necessary to have a metric to make comparisons among them. Sometimes, we see the list of 'most cited authors' in computer science, or physics, or other fields of research. Occasionally we also come across papers that have been highly cited. Unfortunately, however, these statistics may not well represent the impact a researcher or a paper has on the progress of science. There are many individuals with remarkable number of citations. Nevertheless, no scientific community could recognize them with prizes. There are papers that naturally attract many scientists. For example, papers concerning general and wider audience, like, poverty, environment, and economics. Works in these fields may well be cited without actually comprehensively understanding the related material to the depth. Whereas, understanding the works of, say, the first ever female Fields medalist Maryam Mirzakhani must be a very daunting task even for mathematicians of significant orders. This is why when Maryam Mirzakhani won Fields Medal, her citation count was even less than 1000.

In this paper, we propose an alternative metric for researcher ranking based on the presumed quality of their published work. In each field, there are significant honours and prizes that are well recognized and sought after by the scientific community. For example, Turing award and ACM Fellowships are very honored in the field of computers. We can consider these honors and prizes as sources of citation weights. Whoever obtains these honours can be considered as being vetted by these sources, and some weights will flow from the sources to the scientists being recognized with these accolades and prizes. The weights gained by the award recipients will flow to scientists referred by these scientists. Then from these publications to their coauthors and other work the publications cite, and so on. As this process continues, elements of the entire citation network will be infused with weights reflecting a qualitative measure of the publications and their authors.

Some Useful Resources

1. [1] has a database of all Nobel laureates' publications in the past century in chemistry, physics, and medicine.
2. [2] has a list of all Nobel Laureates in different fields.
3. [3] has a list of field medal winners.
4. [4] has a formatted list of Turing Award winners.
5. [5] has the list of winners of the Abel prize. There are other websites that have a tabular list that is easy to export.

2 The Ranking Model

Our proposed ranking approach can be applied either to the researcher collaboration network or the research citation network. We will explain how to apply the algorithm in both cases.

2.1 Ranking in the Researcher Collaboration Network

To seamlessly integrate ranking in the researcher collaboration network, we will treat each considered award/prize/recognition ρ as a dummy researcher with a positive weight w_ρ indicating the significance (equally, the rank) of the award/prize/recognition. A scientist σ receiving the award ρ is modeled as ρ citing σ. Then the rank of the scientist is expressed as:

$$rank_\sigma = w_\sigma = \sum_{\rho=1}^{N} \frac{w_\rho c_{\rho\sigma}}{c_\rho} \tag{1}$$

Here, c_ρ is the total number of scientists jointly received award ρ for their work in a particular field and $c_{\rho\sigma}$ is 1 if σ receives ρ and 0 otherwise.

Since an award ρ may be given to different scientists in different years, Equation 1 ends up devaluating the weight contribution of awards that are older than those that are new due to the fact that there will be more recipients for older awards. To overcome this problem, we treat each delivery of the award as separate and update Equation 1 as follows to incorporate the time-dependent aspect of award reception.

$$rank_\sigma = w_\sigma = \sum_{\rho=1}^{N} \sum_{\tau=0}^{T} \frac{w_\rho c_{\rho_\tau \sigma}}{c_{\rho_\tau}} \tag{2}$$

Here $c_{\rho_\tau \sigma}$ is 1 if σ receives award ρ in year τ and is 0 otherwise.

Since we treat each award as a dummy researcher and award delivery as a citation, Equation 2 can be used as a general model of ranking researchers through citations made by other researchers. So we rewrite Equation 2 by eliminating distinction between researchers and awards as follows:

$$rank_i = w_i = \sum_{j=1}^{N} \sum_{\tau=0}^{T} \frac{w_j c_{j_\tau i}}{c_{j_\tau}} \tag{3}$$

We have to update the interpretation of c_{j_τ} and $c_{j_\tau i}$ with this generalization. In Equation 3, c_{j_τ} is the total number of citations made by researcher j in year τ and $c_{j_\tau i}$ is the number of citations researcher i got from j among those in c_{j_τ}.

We recognize that the weight gained by a researcher i from another researcher j can come from two different routes. Researchers i and j can collaborate and jointly publish their work or j can cite i's publications. Equation 3 is limiting in the sense that it only captures the latter, while it is, in fact, the former that indicates greater trust between i and j. Hence we update the ranking formula as follows to include a term for research collaboration.

$$rank_i = w_i = \sum_{j=1}^{N}\sum_{\tau=0}^{T} w_j\left(\frac{c_{j_\tau i}}{c_{j_\tau}} + \frac{\kappa_{j_\tau i}}{\kappa_{j_\tau}}\right) \tag{4}$$

Here, κ_{j_τ} represents the number of publications that researcher j has in year τ and $\kappa_{j_\tau i}$ is the number of those publications where i is a coauthor of j.

Equation 4 distributes the entire weight gained by the researcher j to the other researchers j cites and collaborates with. This modeling is too simplistic. It is better to assume that citation and collaboration pass only a fraction of weights from one researcher to another. Consequently, we introduce two tuning parameters α and β with $0 < \alpha, \beta < 1$ as follows:

$$rank_i = w_i = \sum_{j=1}^{N}\sum_{\tau=0}^{T} w_j\left(\alpha\frac{c_{j_\tau i}}{c_{j_\tau}} + \beta\frac{\kappa_{j_\tau i}}{\kappa_{j_\tau}}\right) \tag{5}$$

Finally, note that Equation 5 is time-dependent, but we considered the weight of researcher j to be fixed for the whole time range. In the real world, quality of researchers' work grow or decline with time; so is their weight. Consequently, the worth of a citation made by a researcher j to another researcher i or a collaboration between the two should depend on the weight of j during the time of the said citation or collaboration. So we made the final update to the weight propagation model as follows:

$$rank_{i_t} = w_{i_t} = \sum_{\tau=0}^{t}\sum_{j=1}^{N} w_{j_\tau}\left(\alpha\frac{c_{j_\tau i}}{c_{j_\tau}} + \beta\frac{\kappa_{j_\tau i}}{\kappa_{j_\tau}}\right) \tag{6}$$

Here w_{i_t} is the weight or rank of the researcher i at time t. Note that, Equation 6 allows incremental updates of researchers' rank as new publications emerge every new year.

2.2 Ranking in the Research Citation Network

The idea of ranking researchers using the citation network is that award recipients are recognized due to the impact/significance of their work that should be reflected in their research publications. Therefore, instead of propagating weights from awards to researchers in a researcher collaboration network, we can propagate weights from awards to publications in the citation network.

Assuming ρ_τ representing an award given in year τ to a researcher σ among total R_{ρ_τ} researchers and j_σ^T representing a research publication of σ in year T then the weight propagation logic from ρ_τ to the publications is as follows:

$$a^{\rho}_{j^T_\sigma} = \begin{cases} \frac{w_{\rho_\tau}}{R_{\rho_\tau}} \frac{\log_\tau(T+1)}{\sum_{t=0}^{\tau-1} c^t_\sigma \log_\tau(t+1)} & T < \tau \\ 0 & T \geq \tau \end{cases} \tag{7}$$

Here, c_σ^t is the total number of publications of the researcher σ in year t. The idea of Equation 7 is that each joint winner gets an equal fraction ($\frac{1}{R_{\rho_\tau}}$) of the weight (w_{ρ_τ}) from the award recognition which is distributed to their previous publications (dated $1 \leq t < \tau$) on a logarithmic scale based on the year difference between the publication and award recognition dates. Notice that this scheme propagates weights to previous years' publication only. Publications in the year of award reception and later are ignored as research-related awards are given to individuals based on what they have done in the past.

As an author of a research paper can receive multiple awards after its publication and there may be multiple award-winning authors in a single paper. The total weight received by a paper i published in year t due to award-related reasons can be expressed as follows:

$$a_i = \sum_{\sigma \in authors(i)} \sum_{\substack{\tau > t \\ \rho_\tau \in awards(\sigma)}} \frac{w_{\rho_\tau}}{R_{\rho_\tau}} \frac{\log_\tau(T+1)}{\sum_{t=0}^{\tau-1} c_\sigma^t \log_\tau(t+1)} \tag{8}$$

Note that in the research citation network, there is no convincing way of representing an award as a dummy publication. This is because the amount of weight being propagated to a paper of an award recipient from the award depends on his/her other publications; consequently, it is different for different receivers of the same award. Hence, the award weight a_i for a publication i must be considered as its initial weight.

Once we have the initial weights for all the papers, the logic of propagating weight from paper j to paper i due to a citation is quite simple, as given in the following equation:

$$w_i = \alpha \sum_{\forall j | j \ cites \ i} \frac{w_j}{c_j} \tag{9}$$

Here c_j is the total number of references in paper j and α is a tuning parameter.

Once we have the weights of the research papers, the weights of the researchers can be computed as the sum of the weights of their publications.

2.3 Kaykobad Sir's Original Writing

Dear Sir, I moved your original writing to the file old-writing.tex in the overleaf project because it is not needed at the moment.

3 Algorithmic Implementation

Different implementations with different practical merits and concerns are possible for the ranking models of Section 2. In this section, we discuss these implementations one by one.

3.1 Year-by-year Incremental Ranking Calculation in the Researcher Collaboration Network

This algorithmic implementation most closely aligns with the theory of the ranking model of 2 that accommodates the fact that a researcher's rank varies as years go by and adjust weights propagation from prize winning researchers to others accordingly on a yearly basis.

Assume that we represent the collaboration among researchers in a year † as a matrix $\mathcal{K}_\dagger$ where the rows and columns are researchers and entry (i, j) is ν if researcher i and j have ν joint publications in †. In addition, we have another matrix $\mathcal{C}_\dagger$ in a similar vein that represents the citation counts from i to j during †. Then we can have a PageRank-like [6] recursive algorithm to calculate the weights of the researchers. Assuming that the ranks of all scientists for year † is represented by the vector $\mathcal{R}_\dagger$ then the equation for ranking in the matrix format is as follows:

$$\mathcal{R}_\dagger = c(\alpha\mathcal{C}_\dagger + \beta\mathcal{K}_\dagger)\mathcal{R}_\dagger \tag{10}$$

Here c is a normalizing factor to keep the $\|\mathcal{R}_\dagger\|$ constant. That is, we are seeking the dominant eigenvector of the linear combination matrix $M_\dagger = \alpha\mathcal{C}_\dagger + \beta\mathcal{K}_\dagger$. As in PageRank, this can be achieved by repeatedly applying $M_\dagger$ on $\mathcal{R}_\dagger$ starting from a non-degenerate starting vector. **Question for Kaykobad Sir: how do ensure that $\mathcal{M}_\dagger$ is non-degenerate?** With this formalism, the recursive algorithm for calculating the ranks of researchers for the current year t is as follows.

Algorithm 1 Incremental Ranking algorithm

procedure RANK($\mathcal{R}_0, \mathcal{D}, t, \epsilon$) ▷ ϵ is the convergence threshold
  $\tau \leftarrow 1$
  while $\dagger \neq t$ **do** ▷ Iterate over the years
    $\mathcal{M}_\dagger \leftarrow \mathcal{D}.getMatrixForYear(\dagger)$
    $\mathcal{R}_\dagger \leftarrow \mathcal{R}_{\dagger-1}$
    $\delta \leftarrow 1$
    $i \leftarrow 1$
    while $\delta > \epsilon$ **do** ▷ Recursively update ranks for a year until convergence
      ${\mathcal{R}_\dagger}^{i+1} \leftarrow \mathcal{M}_\dagger {\mathcal{R}_\dagger}^{i}$
      $d \leftarrow \left\|{\mathcal{R}_\dagger}^{i+1}\right\| - \left\|{\mathcal{R}_\dagger}^{i}\right\|$
      ${\mathcal{R}_\dagger}^{i+1} \leftarrow {\mathcal{R}_\dagger}^{i+1} - d{\mathcal{R}_\dagger}^{i}$
      $\delta \leftarrow \left\|{\mathcal{R}_\dagger}^{i+1} - {\mathcal{R}_\dagger}^{i}\right\|$
      $i \leftarrow i + 1$
    end while
    $\dagger \leftarrow \dagger + 1$
  end while
  return $\mathcal{R}_t^i$ ▷ Return the final ranks of all researchers
end procedure

An important concern for Algorithm 1 is that how do we get the initial vector $\mathcal{R}_0$ for the first year t_0. If we consider that only some award winners of year t_0 and the dummy researchers representing the awards having non-zero weight in $\mathcal{R}_0$ then we will have a restrictive model where citations and collaboration efforts in year t_0 of future award winners will be disregarded as they may yet not have worked with or cited by existing award receipients.

Common sense reasoning says that scientists receive major awards due to the significance of their research contributions until the year of award recognition. Some awards are for lifelong achievements and others for scientists' discoveries in recent years on a particular topic. In any case, typical award winners have the tendency to always focus on innovative work from their early career and grow more as a scientist with time with their achievements reaching the summit when the award is given. We model this pattern using a logarithmic curve with a threshold as follows for initiating the rank vector for the first year:

$$Rank_0^i = w_0^i = \sum_{\rho=1}^{N_{prize}} w_\rho \frac{c_{\rho 0 i}}{c_{\rho_0}} + \sum_{\rho=1}^{N_{prize}} \sum_{\tau=1}^{T} log(\frac{\gamma_\rho^{c_{\rho_\tau i}}}{\tau}) \tag{11}$$

That is, the weight of a prize winning researcher is the sum of the weights of the prizes he/she has won before the starting year for rank calculation added to a sum of logarithmic weight of future prizes he/she won in later years. Here, the logarithmic additive decreases with the year-wise distance of the future prize reception from the starting year. Notice that the exponent in $\gamma_\rho^{c_{\rho_\tau i}}$ ensures that no weight is added for researchers who never received an accolade.

Finally, Algorithm 1 has a problem with the rank of dummy researchers that represent various prestigious awards in $\mathcal{R}_\dagger$. Since none cites or collaborates with the dummy researchers – rather dummy researchers cite others only – weights of these dummy researchers depletes in the recursive weight update process. Consequently, as more and more years passes, the weight contribution of the dummy researchers to award winner real researchers keep diminishing also. To avoid this problem, we add another vector $\mathcal{R}_{award}$ of same dimensions with the entries for dummy researchers having weights equivalent to their underlying prizes and all other entries zero with $\mathcal{R}_\dagger$. Algorithm 2 present the pseudocode with this adjustment.

Practical Merits and Concerns The algorithm 2 has the advantage that it can be used to update the previous ranking of researchers using the new researcher collaboration matrix for the most recent year only. That is, this implementation should consume the least amount of memory (assuming that a sparse matrix representation such as CSR is used). However, a major concern is that the weight of dummy researchers representing various awards directly propagates only to the current year award recipients. Past winners, although retaining their previous weights, do not have citation/collaboration links from the awards as the matrix is changed. It is unclear how this loss of history will affect the rankings.

Algorithm 2 Incremental Ranking algorithm with Weight Adjustment

procedure RANK($\mathcal{R}_0, \mathcal{D}, t, \epsilon, \mathcal{R}_{award}$)
  $\tau \leftarrow 1$
  while $\dagger \neq t$ **do** ▷ Iterate over the years
    $\mathcal{M}_\dagger \leftarrow \mathcal{D}.getMatrixForYear(\dagger)$
    $\mathcal{R}_\dagger \leftarrow \mathcal{R}_{\dagger-1} + \mathcal{R}_{award}$
    $\mathcal{R}_\dagger \leftarrow \frac{\mathcal{R}_\dagger}{\|\mathcal{R}_\dagger\|}$
    $\delta \leftarrow 1$
    $i \leftarrow 1$
    while $\delta > \epsilon$ **do** ▷ Recursively update ranks for a year until convergence
      ${\mathcal{R}_\dagger}^{i+1} \leftarrow \mathcal{M}_\dagger {\mathcal{R}_\dagger}^{i}$
      $d \leftarrow \|{\mathcal{R}_\dagger}^{i+1}\| - \|{\mathcal{R}_\dagger}^{i}\|$
      ${\mathcal{R}_\dagger}^{i+1} \leftarrow {\mathcal{R}_\dagger}^{i+1} - d{\mathcal{R}_\dagger}^{i}$
      $\delta \leftarrow \|{\mathcal{R}_\dagger}^{i+1} - {\mathcal{R}_\dagger}^{i}\|$
      $i \leftarrow i + 1$
    end while
    $\dagger \leftarrow \dagger + 1$
  end while
  return $\mathcal{R}_t^i$ ▷ Return the final ranks of all researchers
end procedure

3.2 Direct Current-year Ranking Calculation in the Researcher Collaboration Network

In this approach, we have a single collaboration network adjacency matrix $\mathcal{M}$ that contains information about researchers' collaboration through citation and co-authorship from the beginning year to the current and the recipients of various awards in different years being assigned an award-based initial weights in a single weight vector $\mathcal{R}$. There is no dummy researcher entries in $\mathcal{R}$ to represent the awards anymore and a researcher's initial weight is assigned based on Equation 2. The revised recursive formula for computing ranks is now as follows:

$$\mathcal{R} = c(\alpha\mathcal{C} + \beta\mathcal{K})\mathcal{R} = c\mathcal{M}\mathcal{R} \tag{12}$$

Here the (i, j)th entry of M is computed, in light of Equation 6, using the following formula:

$$\mathcal{M}_{i,j} = \alpha \frac{\sum_{\tau=0}^{t} c_{j_\tau i}}{\sum_{\tau=0}^{t} c_{j_\tau}} + \beta \frac{\sum_{\tau=0}^{t} \kappa_{j_\tau i}}{\sum_{\tau=0}^{t} \kappa_{j_\tau}} \tag{13}$$

The recursive algorithm for this implementation of ranking is quite simple and given in Algorithm 3.

Practical Merits and Concerns Notice that since $\mathcal{M}$ contains information of multiple years, the concern that the significance of a citation or collaboration from researcher i to researcher j should depend on i's current rank is absent in

Algorithm 3 Direct Ranking Algorithm

procedure RANK($\mathcal{R}, \mathcal{M}, \epsilon$) ▷ ϵ is the convergence threshold
  $\mathcal{R}^0 \leftarrow \mathcal{R}$
  $\delta \leftarrow 1$
  $i \leftarrow 1$
  while $\delta > \epsilon$ **do** ▷ Recursively update ranks until convergence
    $\mathcal{R}^{i+1} \leftarrow \mathcal{M}\mathcal{R}^i$
    $d \leftarrow \left\|\mathcal{R}^{i+1}\right\| - \left\|\mathcal{R}^i\right\|$
    $\mathcal{R}^{i+1} \leftarrow \mathcal{R}^{i+1} - d\mathcal{R}^i$
    $\delta \leftarrow \left\|\mathcal{R}^{i+1} - \mathcal{R}^i\right\|$
    $i \leftarrow i + 1$
  end while
  return $\mathcal{R}^i$ ▷ Return the final ranks of all researchers
end procedure

this approach. That is all collaboration/citations are treated equal no matter how old or recent they may be. Furthermore, the memory requirement is higher as $\mathcal{M}$ should be significantly more dense than each $\mathcal{M}_\dagger$ of the previous case. The only advantage is that all awards winners' information is considered in the initial vector $\mathcal{R}$ – no history is lost.

3.3 Ranking Calculation using the Research Citation Network

Assuming $\mathcal{P}$ representing the vector of all research publications with $\mathcal{P}$(i) is the weight of the paper i, the recursive equation for updating weights of papers from other papers citing them is as follows:

$$\mathcal{P} = \kappa(\alpha \mathcal{M}_c \mathcal{P}) \tag{14}$$

Here $\mathcal{M}_c$ is the weighted adjacency matrix of the citation network with $\mathcal{M}_c(i,j) = \frac{1}{c_j}$ if paper i is one of the c_j papers j cited, otherwise it is zero. κ is the normalizing factor that keeps $\|\mathcal{P}\|$ constant. Finally, the initial value of $\mathcal{P}$ is computed using the award weights propagation rule of Equation 8.

The matrix format equation for computing the researchers' rank from their publications' weight is follows:

$$\mathcal{R} = \mathcal{M}_r \mathcal{P} \tag{15}$$

Here, $\mathcal{M}_r$ is the authorship matrix where $\mathcal{M}_r(i,j)$ is 1 if researcher i is an author of publication j, otherwise it is 0.

Note that unlike in the case of ranking researchers in the collaboration network, it is not possible to update weights of papers on a yearly basis using only recent year's publications in the citation network. This is because of the manner award weights are computed for prize-winning authors (see Equation 8). Consequently, there is only one algorithm for ranking update in the citation network that computes the rank of all researchers using all publication and award information. Algorithm 4 provides the corresponding pseudocode.

Algorithm 4 Ranking through Publication Algorithm

procedure RANK($\mathcal{P}, \mathcal{M}_c, \mathcal{M}_r, \epsilon$) ▷ ϵ is the convergence threshold
    $\delta \leftarrow 1$
    $i \leftarrow 1$
    while $\delta > \epsilon$ **do** ▷ Recursively update paper weights until convergence
        $\mathcal{P}^{i+1} \leftarrow \mathcal{M}_c \mathcal{P}^i$
        $d \leftarrow \left\|\mathcal{P}^{i+1}\right\| - \left\|\mathcal{P}^i\right\|$
        $\mathcal{P}^{i+1} \leftarrow \mathcal{P}^{i+1} - d\mathcal{P}^i$
        $\delta \leftarrow \left\|\mathcal{P}^{i+1} - \mathcal{P}^i\right\|$
        $i \leftarrow i + 1$
    end while
    $R \leftarrow \mathcal{M}_r \mathcal{P}^i$ ▷ Compute researchers' rank from papers' weights
    return $\mathcal{R}$
end procedure

Practical Merits and Concerns The memory and runtime requirement for this version of the ranking algorithm is the highest, as the citation network is supposed to be several times larger and denser than the researcher collaboration network. However, arguably not giving award weights to publications after an award has been given is a fairer approach that was missing in algorithms using the collaboration network.

4 Limitations and Future Work

A big issue with citation count-based research/researcher ranking is that the approach ignores why a publication has been cited in the first place. If we refer to the general structure of a conference paper or journal article, we observe that the citations in the introduction and related work sections are mainly there for three reasons:

- Show that the problem that the current paper addresses is of broader interest.
- Enable cross-checking earlier publications on the same/similar problem for validation of current paper's novel contributions.
- To bolster the claim of the authors that their work is better than those preceding it.

None of the above suggests that the paper that cited these other works is indebted to them in anyway. Consequently, the rewards for these types of citation should be questioned. On the other hand, citations that appear in the later parts of a paper discussing methodology, algorithms, evaluation results, etc. indicates practical use of earlier work on the work of the current paper's authors. So, these latter citations should get significant weights.

Another important metric for evaluating the merits of a citation is to consider the origin of the citation. For example, if a research work is cited from a book, as opposed to some conference paper or journal article, then we can typically infer that the cited work has a lasting impact in the underlying field. A similar useful quality metric for a citation's worth can be the publication year gap between the cited publication and the work citing it. A larger year gap typically means the contribution of the former publication is still relevant despite its age. Furthermore, if data are available, we can check if a publication from one area has been cited in other areas, which is an indication that the cited publication has much broader impact than a publication that has been cited in its own narrow field.

Finally, when weighing citations, we should consider the fact that certain areas of study attract much less citations than some others. For example, applied and cross-disciplinary research works typically get cited several times more than core theoretical, infrastructure, methodology, and algorithm/mathematical discoveries. Therefore, if a mechanism is present to tag publications with their field/sub-fields then we can measure the relative density of citations in different fields. Afterwards, we can normalize the weights of citations differently based on where the cited work belongs.

Some Concerns

1. There are so many awards. How can we obtain information on all the important prize winners?
2. The prize winner lists available on various websites have a naming format that differs from the same winners' author information in their papers. How to reconcile this difference?
3. Some papers of award winners have empty citation entries. How can we correct them?
4. If we cannot consider all important prizes, then what conclusion can we derive about the other researchers in the citation network?
5. PageRank-like ranking algorithms have been proposed for ranking authors in citation networks several times in the past. We need to read the cited work of the following paper to discover the novelty of our proposal [7], if exists.

References

1. Jichao Li, Yian Yin, Santo Fortunato, and Dashun Wang. A dataset of publication records for nobel laureates. *Scientific Data*, 6(1):33, Apr 2019.
2. Nobel prize laureates dataset, May 2017.
3. Fields medal winners — wolfram data repository, Jul 2017.
4. Eduardo Casais. The complete list of acm turing award winners.
5. Abel prize laureates — the abel prize.
6. Sergey Brin and Lawrence Page. The anatomy of a large-scale hypertextual web search engine. *Comput. Netw. ISDN Syst.*, 30(1–7):107–117, April 1998.
7. Ying Ding, Erjia Yan, Arthur Frazho, and James Caverlee. Pagerank for ranking authors in co-citation networks. *J. Am. Soc. Inf. Sci. Technol.*, 60(11):2229–2243, November 2009.