How Hard Is It to Impact the Impact of Your Paper?

Yongjie Yang
Chair of Economic Theory, Saarland University, Saarbrücken, Germany
yyongjiecs@gmail.com

Abstract

Consolidation-disruption index (CD index) is a new metric for qualitatively measuring the contribution of a patent or a research paper. We embark on the study of the complexity of the CD index manipulation problems, which model scenarios where scholars seek to enhance the CD indices of their papers through the merging, addition, or deletion of papers. We show that these problems are generally computationally hard, even when restricted to very realistic special cases. Specifically, we analyze how various parameters influence the parameterized complexity of these problems.

1 Introduction

Quantitatively evaluating the achievements of scientific researchers plays a significant role in university recruiting, awarding, grant proposal determination, and more. Various measures have been proposed, including the h-index, the i10-index, and others. These measures consider a researcher's published papers and the number of citations these papers receive, assigning a numerical score where a higher score is deemed more favorable.

Another relatively recent but impactful bibliometric is the consolidation-disruption (CD) index, initially examined by Funk and Owen-Smith [2017] to assess the degrees of destabilization and consolidation of patents. Wu, Wang, and Evans [2019] first applied this index to scientific publications. According to this index, the contribution of a research paper can fall into the categories of consolidative, disruptive, or somewhere in between. Consolidative implies that a paper adds value to a field by integrating, improving, or expanding upon existing knowledge. Conversely, disruptive indicates that a paper delves into relatively novel directions that may deviate from preceding research, identifies noteworthy flaws in prior studies, or entirely challenges earlier works.

To compute the CD index of a paper p in a citation graph D (a directed graph whose vertices are papers and arcs represent citations between papers), three sets $N^D_{\mathrm{F}}(p)$, $N^D_{\mathrm{B}}(p)$, and $N^D_{\mathrm{R}}(p)$ are identified[1]:

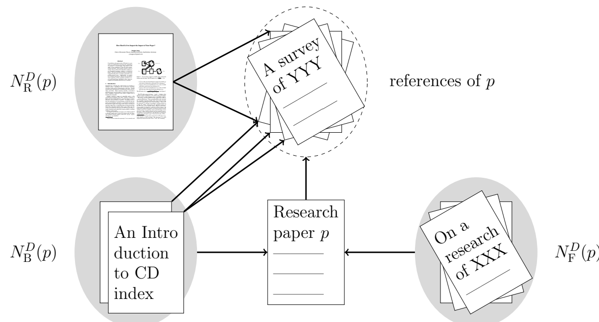

Figure 1: An arc from a paper to another paper means the former cites the latter. The CD index of p is $\frac{3-2}{3+2+1} = 1/6$.

- $N^D_{\mathrm{F}}(p)$ is the set of nonreferences (papers not cited by p) of p, each of which cites p but none of p's references.
- $N^D_{\mathrm{B}}(p)$ is the set of nonreferences of p, each of which cites both p and at least one reference of p.
- $N^D_{\mathrm{R}}(p)$ is the set of nonreferences of p, each of which does not cite p but cites at least one reference of p.

The CD index of p is $\frac{|N^D_{\mathrm{F}}(p)|-|N^D_{\mathrm{B}}(p)|}{|N^D_{\mathrm{F}}(p)|+|N^D_{\mathrm{B}}(p)|+|N^D_{\mathrm{R}}(p)|}$. See Figure 1.

The CD index ranges between -1 and 1. A higher value indicates greater disruption to the related area by the paper, whereas a lower value suggests a more consolidative impact.

The groundbreaking study by Wu, Wang, and Evans [2019] unveiled a somewhat surprising discovery: "large teams develop and small teams disrupt science and technology". Their work has amassed 777 citations (according to Google Scholar as of May 5, 2024), underscoring the significant impact of the CD index in the scientific realm. Based on this index, a more recent investigation conducted by Park, Leahey, and Funk [2023] suggests that papers are much less likely to break with the past in ways pushing science and technology in new directions. Other scholars pointed out that scientific innovation appears to be slowing down, and appealed to academia to "make science disruptive again" [Yanai and Lercher, 2023].

However, concerns regarding the reliability of the CD index have prompted public debate on the application of the

[1] The paper being analyzed for its CD index is typically referred to as the focal paper. The subscripts F, B, and R respectively stand for the words "focal", "both", and "remaining".

index. For example, as noted by Liu, Zhang, and Li [2023], even minor alterations in a paper's references could significantly impact its CD index. This has led several researchers to question the validity of the index and advocate for a more rigorous investigation of its theoretical underpinnings before its widespread adoption [Leibel and Bornmann, 2024]. In pursuit of a preliminary solution to this issue, we study the manipulability of this index. Specifically, we explore the complexity challenges associated with manipulating the CD index of papers using operations such as paper merging, addition, or deletion, through the lens of parameterized complexity. Our problem formulations encompass scenarios in which a scholar seeks to identify a set of representative papers and endeavors to achieve this objective through one of the aforementioned modification operations. The feasibility of these modification operations is evident on academic online platforms like Google Scholar, which allows scholars to manually add new manuscripts, delete articles, and merge their own papers. Our main contributions are as follows.

(1) We initiate the study of the CD index manipulation problems, and show that they are computationally hard even when restricted to very realistic special cases.

(2) In the problem of CD index manipulation by merging papers, a scholar seeks to ensure that at least ℓ papers in a citation graph attain a CD index of at least d (a threshold deemed satisfactory by the scholar), by performing at most β merging operations on papers from a given set W, such as papers authored by the scholar. We depict the similarity relation between papers through a compatibility graph H, where an edge signifies similarity, and only papers exhibiting similarity (forming a clique) are eligible for merging. We show that w.r.t. the parameters β and ℓ, this problem exhibits a W[1]-hardness or a para-NP-hardness lower bound, even when each connected component of H contains at most two papers and the citation graph is acyclic. However, we also derive XP-algorithms for these two parameters.

(3) In the problems of CD index manipulation by adding/deleting papers, a scholar aims to ensure that each paper in a given set J has a CD index of at least d by adding/deleting at most k papers. We explore the parameters k and $|J|$. The results for these problems exhibit a diverse range, including both fixed-parameter tractability and fixed-parameter intractability. It is noteworthy that the complexity varies in certain instances when considering cases where $d = 1$ and $d < 1$. Moreover, we identify several complexity dichotomies w.r.t. the parameter $|J|$. Another interesting finding is that when $d = 1$ and the parameter is k, the problem of adding papers is W[2]-hard, while the problem of deleting papers is fixed-parameter tractable (FPT).

(4) We also explore variants of the manipulation problems involving paper addition/deletion where the scholar does not concern themselves with the specific number of papers added or deleted. Our results indicate that, for the paper addition operation, the complexity of the original problem and the variant coincide. However, for the paper deletion problem, the variant is easier to solve than the original one.

		β	ℓ
acyclic	$d \leq 1$	para-NP-h (Thm.4)	para-NP-h (Thms.1,4)
acyclic	$d = 1$	W[1]-h, XP (Thms.2,5)	W[1]-h, XP (Thms.2,3)
&small	$d < 1$	W[1]-h, XP (Thms.1,5)	para-NP-h (Thm.1)

Table 1: The complexity of CD index manipulation by merging papers. "Acyclic" (resp. small) means the corresponding results hold even when the citation graph is acyclic (resp. each connected component of the compatibility graph contains only a constant number of papers).

Our results are summarized in Table 1 and Table 2.

2 Related Works

Our study is closely related to the works of van Bevern *et al.* [2020; 2016], who are the first to explore the complexity of determining if a scholar could improve their h-index by adding, deleting, or merging papers.[2] In a subsequent paper, Pavlou and Elkind [2016] studied the complexity of similar problems for the g-index and the i10-index.

From a motivational standpoint, our work also bears relevance to election control problems. These problems model scenarios where an election controller endeavors to influence the election outcome by adding/deleting a(n) (un)limited number of voters/candidates [Bartholdi III *et al.*, 1992; Baumeister and Rothe, 2016; Erdélyi *et al.*, 2021; Erdélyi *et al.*, 2020; Faliszewski and Rothe, 2016; Yang, 2019].

The CD index can be mathematically regarded as an "centrality measure" of vertices in a network. Problems involving improving the influence of a vertex by adding/deleting vertices or edges (or arcs) have been thoroughly explored in the literature. Betzler *et al.* [2014] investigated the problems of determining whether a given vertex can achieve a maximum/minimum degree by adding/deleting vertices. Variants of the problems restricted to special graph classes have also received attention [Li *et al.*, 2023; Mishra *et al.*, 2015]. Analogous problems concerning other centrality measures have been studied as well [Bergamini *et al.*, 2018; D'Angelo *et al.*, 2019; D'Angelo *et al.*, 2015]. Furthermore, concepts for quantifying the centrality of groups of vertices and associated centrality-maximizing or -minimizing problems have also been examined [Waniek *et al.*, 2023; Waniek *et al.*, 2022]. For comprehensive summaries of many centrality measures, please refer to [Das *et al.*, 2018; Singh, 2022; Wan *et al.*, 2021].

Finally, we would like to point out that very recently, several variants of the CD index have been proposed [Leydesdorff *et al.*, 2021; Wang *et al.*, 2023; Yang *et al.*, 2023]. Particularly, Jiang and Liu [2023] proposed an analogous index for measuring the disruption of journals.

[2]The conference versions of the two papers first appeared in 2016 and 2015, respectively.

	CD-MANI-ADD		CD-MANI-ADD$^\infty$	CD-MANI-DEL		CD-MANI-DEL$^\infty$
	k	$\lvert J\rvert$	$\lvert J\rvert$	k	$\lvert J\rvert$	$\lvert J\rvert$
$d = 1$	W[2]-h (Thm.6) XP (trivial)	W[1]-h (Thm.7) XP (Thm.9)	W[1]-h (Cor.1) XP (Thm.9)	FPT (Thm.12)	$\lvert J\rvert = 1$: P (Thm.14) $\lvert J\rvert = 2$: NP-h (Thm.15)	W[1]-h (Thm.19) XP (Thm.20)
$d < 1$	W[2]-h (Thm.6) XP (trivial)	$\lvert J\rvert = 1$: P (Thm.8) $\lvert J\rvert = 2$: NP-h (Thm.10)	$\lvert J\rvert = 1$: P (Thm.8) $\lvert J\rvert = 2$: NP-h (Thm.11)	W[1]-h (Thm.13) XP (trivial)	$\lvert J\rvert = 1$: NP-h (Thm.16)	$\lvert J\rvert \leq 2$: P (Thm.18) $\lvert J\rvert = 3$: NP-h (Thm.17)

Table 2: The complexity of CD index manipulation by adding/deleting papers. All hardness results except the one for CD-MANI-DEL$^\infty$ with $d < 1$ hold for acyclic citation graphs. The tractability result for CD-MANI-DEL$^\infty$ with $d < 1$ and $|J| = 2$ holds for acyclic citation graphs, while other tractability results hold for general citation graphs.

3 Preliminaries

We assume the reader is familiar with the basics of graph theory [Bang-Jensen and Gutin, 2018; West, 2000].

3.1 Citation Graphs and the CD index

A *citation graph* D is a digraph, where each vertex represents a scientific paper, and an arc from v to u, denoted (v, u), means that v cites u. We use $V(D)$ to denote the vertex set of D, and use $A(G)$ to denote its arc set. Throughout the paper, we use the terms "vertex" and "paper" interchangeably. The set of *inneighbors* of $v \in V(D)$ in D is $\Gamma^-_D(v) = \{u \in V(D) \setminus \{v\} : (u, v) \in A(D)\}$. The set of *outneighbors* of v in D is $\Gamma^+_D(v) = \{u \in V(D) \setminus \{v\} : (v, u) \in A(D)\}$. We call papers in $\Gamma^+_D(v)$ the *references* of v, and call those in $\Gamma^-_D(v)$ the *citations* of v. Let $\Gamma_D(v) = \Gamma^-_D(v) \cup \Gamma^+_D(v)$. For $S \subseteq V(D)$, let $\Gamma^+_D(S) = (\bigcup_{v \in S} \Gamma^+_D(v)) \setminus S$ and let $\Gamma^-_D(S) = (\bigcup_{v \in S} \Gamma^-_D(v)) \setminus S$. For $v \in V(D)$, we define

- $N^D_{\mathrm{F}}(v) = \{u \in \Gamma^-_D(v) \setminus \Gamma^+_D(v) : \Gamma^+_D(u) \cap \Gamma^+_D(v) = \emptyset\}$.
- $N^D_{\mathrm{B}}(v) = \{u \in \Gamma^-_D(v) \setminus \Gamma^+_D(v) : \Gamma^+_D(u) \cap \Gamma^+_D(v) \neq \emptyset\}$.
- $N^D_{\mathrm{R}}(v) = \{u \in V(D) \setminus (\Gamma_D(v) \cup \{v\}) : \Gamma^+_D(u) \cap \Gamma^+_D(v) \neq \emptyset\}$.

For each $X \in \{\mathrm{F}, \mathrm{B}, \mathrm{R}\}$, let $n^D_X(v) = |N^D_X(v)|$.

Definition 1 (CD Index)**.** *The CD index of a paper v in a citation graph D is $\frac{n^D_F(v) - n^D_B(v)}{n^D_F(v) + n^D_B(v) + n^D_R(v)}$ if the denominator is nonzero, and is undefined otherwise.*

Notice that, by definition, for two papers v and v' citing each other,[3] when we analyze the CD index of v, v' is classified as a reference of v and, hence, is excluded from $N^D_{\mathrm{F}}(v) \cup N^D_{\mathrm{B}}(v) \cup N^D_{\mathrm{R}}(v)$. See Figure 2 for an illustration.

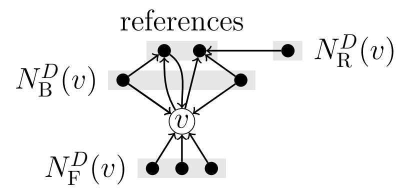

Figure 2: The CD index of v is $\frac{3-2}{3+2+1} = 1/6$.

3.2 The Merging Operation

Before formally introducing the manipulation problems, we need to clarify a few points.

First, in practice, scholars usually only merge papers with similar titles or topics. To formulate this assumption, we adopt the notion of *compatibility graph* used by van Bevern *et al.* [2016]. A compatibility graph is an undirected graph H whose vertices represent papers, and an edge between two papers represents that the two papers share some kind of similarity. A subset of papers can be merged into one paper only if they form a clique in the compatibility graph.

Second, it is essential to elucidate the CD index of a paper subsequent to certain merging operations. In the study of the h-index manipulation problems, van Bevern *et al.* [2016] examined three methodologies to ascertain the citation count a paper accrues post-merging operations. In our framework, not only the quantity but also the categories of these citations hold significance. Hence, the straightforward adoption of the three approaches is not viable. Nevertheless, the fundamental principle of an approach employed by van Bevern *et al.* [2016] posits that merging a set of papers entails considering this set as a singular entity. We find this approach to be the most intuitive and natural, thus opting to incorporate it into our study.

A lingering puzzle remains: how do we handle self-citations? In the majority of prior related research, self-citation is typically excluded or encountered infrequently in experimental works. However, the act of merging papers has the potential to transform a citation graph that initially lacks self-citation into one where self-citation is present. This transformation is exemplified when a paper is merged with some of its references. Such a scenario is not uncommon, particularly in computer science areas, where merging the journal version and its conference versions occurs, with the journal version often citing the conference versions. We emphasize that all our hardness reductions are meticulously designed to prevent the occurrence of such self-citations.

Having resolved potential confusion, let us now proceed to formalize the aforementioned discussion.

Definition 2 (Merging Operation)**.** *A merging operation on a subset $P \subseteq V(D)$ of papers in a citation graph D entails:*

(1) Create a new paper v_P and add arcs so that the set of outneighbors of v_P is exactly $\Gamma^+_D(P)$, and the set of in-

[3]This scenario might occur in cases where, for instance, two research groups independently discover the same result almost simultaneously, and the authors mutually agree to acknowledge each other's work for the sake of ensuring a secure publication, respectfulness, or addressing ethical considerations.

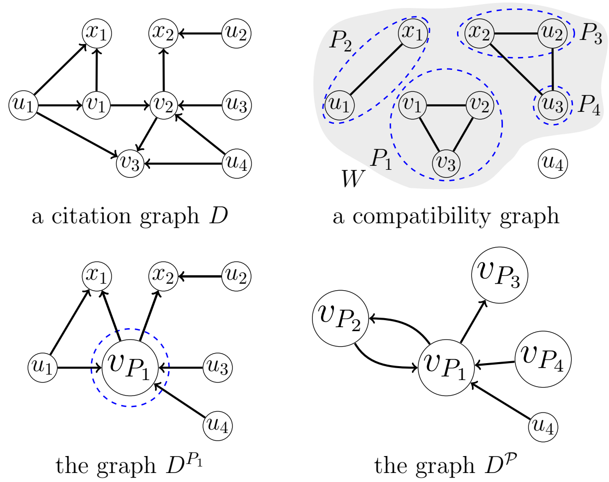

Figure 3: The CD indices of v_{P_1} in D^{P_1} and in $D^{\mathcal{P}}$ are respectively $\frac{2-1}{2+1+1} = 1/4$ and 1. Here, $\mathcal{P} = \{P_1, P_2, P_3, P_4\}$.

neighbors of v_P is exactly $\Gamma^-_D(P)$.

(2) Remove all papers in P from D.

We use D^P to denote the citation graph obtained from D by performing the merging operation on P.

For a partition $\mathcal{P}$ of a subset $W \subseteq V(D)$, we use $D^{\mathcal{P}}$ to denote the citation graph obtained from D by performing merging operations on all $P \in \mathcal{P}$, one after another. As $\mathcal{P}$ is a partition, the order of parts of $\mathcal{P}$ in which the operations are executed is immaterial to the resulting graph. For $S \subseteq W$, we define $S^{\mathcal{P}} = \{v_P : P \in \mathcal{P}, P \cap S \neq \emptyset\}$. Let $\mathcal{P}^{\geq 2} = \{P \in \mathcal{P} : |P| \geq 2\}$. We say that $\mathcal{P}$ *complies with* a compatibility graph H if every $P \in \mathcal{P}$ induces a clique of H. For an illustration of the above discussion, see Figure 3.

3.3 Parameterized Complexity

An instance of a *parameterized problem* is a tuple (I, κ), where κ is a numerical parameter. A parameterized problem is FPT (resp. XP) if it is solvable in time $f(\kappa) \cdot |I|^{O(1)}$ (resp. $|I|^{f(\kappa)}$), where f is a computable function and $|I|$ is the size of I. A well established hierarchy in the realm of parameterized complexity is $\mathsf{FPT} \subseteq W[1] \subseteq W[2] \subseteq \cdots \subseteq \mathsf{XP}$. A problem is W[$i$]-hard if all problems in W[i] are parameter reducible to it. Unless FPT $=$ W[i], W[i]-hard problems do not admit FPT-algorithms, but they (not necessarily all) may have XP-algorithms. A para-NP-hard problem is unlikely to even have an XP-algorithm: a problem is para-NP-hard if there is a constant c such that the problem is NP-hard even when the parameter is fixed to any constant no smaller than c. For further insights, refer to [Cygan *et al.*, 2015].

3.4 Remarks

All reductions presented in the paper are executable in polynomial time, and the problems utilized in the reductions are NP-hard. Hence, if a problem is asserted to be W[1]/W[2]-hard, it is also NP-hard. Moreover, for a para-NP-hardness result, we exclusively present the reduction for the minimum c where the NP-hardness holds when the parameter is set to c. These reductions can be slightly modified by incorporating dummy structures to demonstrate the hardness for all parameters greater than c. Furthermore, it's worth noting that W[1]/W[2]-hardness w.r.t. a combined parameter $\kappa + \kappa'$ implies W[1]/W[2]-hardness w.r.t. κ and κ' individually.

4 The Manipulation Problems

Now we formally introduce the CD index manipulation problems. For an integer i, we use $[i]$ to denote the set of all positive integers not exceeding i. Let $[0, 1]$ be the set of all numbers ranging from 0 to 1.

For an undirected graph G, $V(G)$ and $E(G)$ represent the vertex set and the edge set of G, respectively. The first CD index manipulation problem is defined as follows.

CD Index Manipulation by Merging Papers
(CD-MANI-MERGE)

Input: A citation graph D, a subset $W \subseteq V(D)$, a compatibility graph H such that $V(H) = V(D)$, two integers ℓ, β, and a rational number $d \in [0, 1]$.

Question: $\exists$ a partition $\mathcal{P}$ of W that complies with H such that $|\mathcal{P}^{\geq 2}| \leq \beta$, and at least ℓ papers from $W^{\mathcal{P}}$ have a CD index of at least d in $D^{\mathcal{P}}$?

The subsequent manipulation problems model a scenario where a scholar has selected a set J of papers published in reputable venues, or the papers are regarded as robust papers w.r.t. other measures. This selection is made to strengthen her application for a research grant (or a scientific position, etc.). However, some of the papers exhibit relatively low CD indices. The grant committee, prioritizing projects with innovative ideas or those opening new directions, highly values such aspects. Assume that an online platform, endorsed by the grant entity, offers CD index calculation functions, and scholars are permitted to add new papers (included in a set denoted U) or remove existence ones (included in a set denoted V) from the system. To increase the chance of success, the scholar grapples with the question of whether she can elevate the CD indices of all these selected papers to a satisfactory level by adding or deleting at most k papers. The two problems are formally defined as follows.

For a citation graph D and a subset $S \subseteq V(D)$, $D[S]$ is the subgraph of D induced by S, and $D - S$ denotes the citation graph obtained from D by removing all papers in S.

CD Index Manipulation by Adding Papers
(CD-MANI-ADD)

Input: A citation graph D, a tuple (V, U) of two disjoint sets of papers such that $V \cup U = V(D)$, a subset $J \subseteq V$, an integer k, and a rational number $d \in [0, 1]$.

Question: $\exists U' \subseteq U$ such that $|U'| \leq k$ and each paper from J has a CD index of at least d in $D[V \cup U']$?

CD Index Manipulation By Deleting Papers
(CD-MANI-DEL)

Input: A citation graph D, a subset $J \subseteq V(D)$, an integer k, and a rational number $d \in [0, 1]$.

Question: $\exists V' \subseteq V(D) \setminus J$ such that $|V'| \leq k$ and each paper from J has a CD index of at least d in $D - V'$?

We also explore a natural variant of CD-MANI-ADD/CD-MANI-DEL, where the scholar aims to achieve their goal without considering the addition/deletion budget k. Precisely, the CD-MANI-ADD$^\infty$/CD-MANI-DEL$^\infty$ problem takes the same input as CD-MANI-ADD/CD-MANI-DEL without k, and determines the existence of a subset $U' \subseteq U$ (resp. $V' \subseteq V(D)$) such that each paper from J has a CD index of at least d in $D[V \cup U']$ (resp. $D - V'$).

5 Manipulation by Paper Merging

This section is dedicated to the complexity of CD-MANI-MERGE. We first show that the problem is hard to solve even when restricted to very special cases. Our reduction is from the CLIQUE problem, which takes as input an undirected graph G and an integer κ, and determines if G has a clique of κ vertices. The problem is W[1]-hard w.r.t. κ even when restricted to regular graphs [Mathieson and Szeider, 2008].

Theorem 1. *For any positive rational number $d < 1$, CD-MANI-MERGE is W[1]-hard w.r.t. β, even if $\ell = 1$, every connected component of the compatibility graph contains at most two papers, and the citation graph is acyclic.*

Proof for $d = 1/3$. Let $I = (G, \kappa)$ be an instance of CLIQUE, where G is t-regular with $t > 0$. Let n be the number of vertices of G, and let $m = \frac{n \cdot t}{2}$ be the number of edges of G. W.l.o.g., we assume $t > \kappa \geq 7$ (otherwise, I can be solved trivially). We create an instance $g(I) = (D, W, H, \ell, \beta, d)$, where $\ell = 1$, $\beta = \kappa$, and $d = 1/3$ as follows.

For every $v \in V(G)$, we create two vertices $v(1)$ and $v(2)$. For each $i \in [2]$, let $V(i) = \{v(i) : v \in V(G)\}$. For each edge e in G, we create a paper $v(e)$. Additionally, we create two papers $v^\star$ and y, two sets X and Z, each consisting of $4(n-\kappa)+t\cdot\kappa-\frac{\kappa\cdot(\kappa-1)}{2}$ papers, and a set S of $4(m+2n+1)$ papers. As $t > \kappa \geq 7$, X and Z are nonempty. The arcs of D are created so that exactly the following citations exist: (1) All papers in X cite $v^\star$. (2) The paper $v^\star$ cites all papers from $V(2)$. (3) All papers in $V(1)$ cite all papers in $V(2) \cup \{v^\star\}$. (4) For each edge $e = \{u, v\}$ in G, the paper $v(e)$ cites $u(1)$ and $v(1)$. (5) All papers in $V(1) \cup V(2)$ cite y. (6) All papers in Z cite all papers in $V(1) \cup V(2) \cup \{y\}$. (7) All papers from S cite y. Let $W = V(1) \cup V(2) \cup \{v^\star\}$. The edge set of the compatibility graph H is $\{\{v(1), v(2)\} : v \in V(G)\}$. See Figure 4 for an illustration of the reduction. We prove its correctness below.

($\Rightarrow$) Assume that G has a clique K of κ vertices. Let $\partial_G(K) = \{e \in E(G) : e \cap K \neq \emptyset\}$ be the set of edges in G covered by K. Obviously, $|\partial_G(K)| = t\cdot\kappa - \frac{\kappa\cdot(\kappa-1)}{2}$. Let $\mathcal{P}$ be a partition of W such that $\mathcal{P}^{\geq 2} = \{\{v(1), v(2)\} : v \in K\}$. Consider the citation graph $D^{\mathcal{P}}$, i.e., the graph obtained from D by merging, for each vertex $v \in K$, the corresponding two papers $v(1)$ and $v(2)$. It is fairly easy to verify that the CD index of $v^\star$ in $D^{\mathcal{P}}$ is $\frac{|X|-(|V(1)|-|K|)}{|X|+(|V(1)|-|K|)+|Z|+|\partial_G(K)|}$. Substituting the values of $|X|$, $|V(1)|$, $|K|$, $|Z|$, and $|\partial_G(K)|$ into the above expression yields the index threshold $1/3$.

($\Leftarrow$) Assume that there is a partition $\mathcal{P}$ of W that complies with the compatibility graph H such that $|\mathcal{P}^{\geq 2}| \leq \beta = \kappa$ and

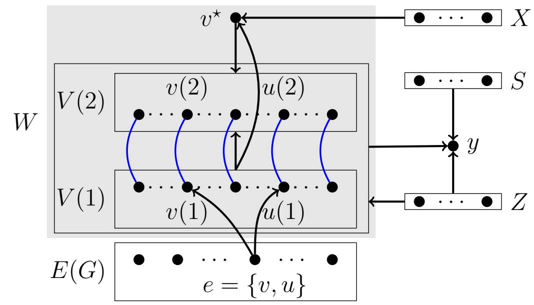

Figure 4: An illustration of the proof of Theorem 1.

at least one paper from $W^{\mathcal{P}}$ has a CD index of at least $1/3$ in the citation graph $D^{\mathcal{P}}$. The presence of y and S precludes any paper from $(V(1) \cup V(2))^{\mathcal{P}}$ from attaining a CD index of at least $1/3$. This leaves only the possibility that $v^\star$ has a CD index of at least $1/3$ in $D^{\mathcal{P}}$. Let $K = \{v \in V(G) : \{v(1), v(2)\} \in \mathcal{P}^{\geq 2}\}$. Merging two papers $v(1)$ and $v(2)$ for any $v \in V(G)$ changes the membership of $v(1)$ from $N_{\text{B}}^D(v^\star)$ to the set of references of $v^\star$, and simultaneously brings all edge-papers covered by v into $N_{\text{R}}^D(v^\star)$. Consequently, the CD index of $v^\star$ in $D^{\mathcal{P}}$ is

$$\frac{|X| - (|V(1)| - |\mathcal{P}^{\geq 2}|)}{|X| + (|V(1)| - |\mathcal{P}^{\geq 2}|) + |Z| + |\partial_G(K)|}. \tag{1}$$

Let m' be the number of edges in $G[K]$. As G is t-regular, it holds $|\partial_G(K)| = t \cdot |K| - m'$. It follows that (1) is at least $1/3$ if and only if $|K| = |\mathcal{P}^{\geq 2}| = \kappa$ and $m' = \frac{\kappa\cdot(\kappa-1)}{2}$, i.e., when K is a clique in G. □

The above hardness does not cover the case where $d = 1$. Recall that a paper possesses a CD index of 1 if and only if, among the three sets defining the CD index, exactly the set $N_{\text{F}}^D(v)$ is nonempty, significantly shrinking the search space. This might seem to simplify the problem. However, a difficulty arises in identifying which papers can have their indices increased to 1 through paper merging. This difficulty leads to the following hardness result.

Theorem 2. *For $d = 1$, CD-MANI-MERGE is W[1]-hard w.r.t. $\beta + \ell$, even if each connected component of the compatibility graph has at most two papers and the citation graph is acyclic.*

What if $d = 1$ and we target only a constant number ℓ of papers to have a CD index of 1? In the case where $\ell = 1$, we can hypothesize which paper v from W is the targeted paper, and then focus on making $N_{\text{F}}^D(v)$ nonempty while ensuring both $N_{\text{B}}^D(v)$ and $N_{\text{R}}^D(v)$ are empty. Would this simplicity turn the complexity of the problem? We show that the answer hinges on the structure of the compatibility graph, as elucidated in the ensuing two theorems.

Theorem 3. *For $d = 1$, CD-MANI-MERGE is in XP w.r.t. ℓ, when each connected component of the compatibility graph is of constant size.*

Theorem 4. *For any positive rational number $d \leq 1$, CD-MANI-MERGE is NP-hard, even if $\ell = 1$, $\beta = 3$, and the citation graph is acyclic.*

When considering β as the parameter, we obtain:

Theorem 5. CD-MANI-MERGE *is in* XP *w.r.t.* β *when each connected component of the compatibility graph contains a constant number of papers.*

6 Manipulation by Paper Addition

Now we examine the operation of paper addition. We observe that CD-MANI-ADD is in XP when parameterized by k: it can be solved by enumerating all possible sets of at most k papers to be added. Can we enhance the result to fixed-parameter tractability? The next theorem answers this question in the negative.

Theorem 6. *For any positive rational number* $d \leq 1$, CD-MANI-ADD *is* W[2]*-hard w.r.t.* k*, even if the citation graph is acyclic.*

Furthermore, we show that restricting the value of d to 1 does not render the complexity of the problem FPT, even when associated with a larger parameter. Our result relies on the MULTICOLORED CLIQUE problem, which takes as input an undirected graph G whose vertices are partitioned into κ sets $(V_i)_{i\in[\kappa]}$, and determines if G has a clique (termed κ-colored clique) which contains exactly one vertex from each V_i for all $i \in [\kappa]$. The problem is W[1]-hard w.r.t. κ [Fellows *et al.*, 2009].

Theorem 7. *For* $d = 1$, CD-MANI-ADD *is* W[1]*-hard w.r.t.* $|J| + k$*. This holds even if the citation graph is acyclic.*

Proof. Let $I = (G, \kappa)$ be an instance of MULTICOLORED CLIQUE, where $V(G)$ is partitioned into $(V_1, V_2, \ldots, V_\kappa)$ and $\kappa \geq 2$. We create an instance of CD-MANI-ADD as follows.

For each $v \in V(G)$, we create one paper denoted by the same symbol. For each edge $e = \{v, u\}$ in G, we create one paper $p(e)$. For distinct $i, j \in [\kappa]$, let $P_{\{i,j\}} = \{p(e) : e \in E(G), e \cap V_i \neq \emptyset, e \cap V_j \neq \emptyset\}$. Besides, we create a set $X = \{x_1, x_2, \ldots, x_\kappa\}$ of κ papers and a set $Y = \{y_{\{i,j\}} : i, j \in [\kappa], i \neq j\}$ of $\frac{\kappa\cdot(\kappa-1)}{2}$ papers. Let $V = X \cup Y$ be the set of registered papers, and let $U = V(G) \bigcup_{i,j\in[\kappa], i\neq j} P_{\{i,j\}}$ be the set of unregistered papers. Arcs in D are created so that exactly the following citations exist: (1) For every $i \in [\kappa]$, all papers from V_i cite x_i. (2) For distinct $i, j \in [\kappa]$, all papers from $P_{\{i,j\}}$ cite $y_{\{i,j\}}$. (3) For distinct $i, j \in [\kappa]$ and every edge $e = \{v, u\}$ between V_i and V_j in G, the paper $p(e)$ cites all papers from $(V_i \cup V_j) \setminus \{v, u\}$. (4) For distinct $i, j \in [\kappa]$, the paper $y_{\{i,j\}}$ cites all papers from $V_i \cup V_j$. Let $J = X \cup Y$ and let $k = \kappa + \frac{\kappa\cdot(\kappa-1)}{2}$. The instance of CD-MANI-ADD is $g(I) = (D, (V \cup U), J, k, d)$, where $d = 1$. See Figure 5 for an illustration.

($\Rightarrow$) Assume that there is a κ-colored clique K in G. Let $P(K)$ be the set of the $\frac{\kappa\cdot(\kappa-1)}{2}$ papers corresponding to the edges within K. Consider the citation graph $D' = D[V \cup K \cup P(K)]$. For every $x_i \in X$, as $|K \cap V_i| = 1$, the construction of the citation graph D ensures that the CD index of x_i in D' is 1. Now consider a paper $y_{\{i,j\}}$, where $\{i, j\} \subseteq [\kappa]$. As K is a κ-colored clique of G, it holds that $|P(E) \cap P_{\{i,j\}}| = 1$. W.l.o.g., let $P(E) \cap P_{\{i,j\}} = \{p(\{v, u\})\}$, where $\{v, u\} \in E(G)$, $v \in V_i$, and $u \in V_j$.

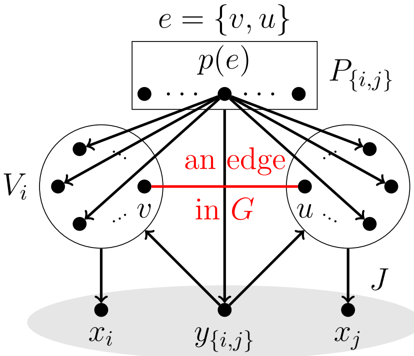

Figure 5: An illustration of the proof of Theorem 7.

Consequently, $K \cap V_i = \{v\}$ and $K \cap V_j = \{u\}$. According to the reduction, in D', v and u are the only references of $y_{\{i,j\}}$, $p(e)$ is the only paper citing $y_{\{i,j\}}$, and $p(e)$ cites neither v nor u. As a result, the CD index of $y_{\{i,j\}}$ in D' is 1. Since $|K \cup P(K)| = \kappa + \frac{\kappa\cdot(\kappa-1)}{2} = k$, $g(I)$ is a Yes-instance.

($\Leftarrow$) Assume that there is a subset $U' \subseteq U$ of at most k papers such that every paper in J has a CD index of 1 in the citation graph $D' = D[V \cup U']$. By the construction of D, a necessary condition for an $x_i \in X$ to have a CD index of 1 in D' is that U' contains at least one paper from V_i. A necessary condition for a paper $y_{\{i,j\}} \in Y$ to have a CD index of 1 in D' is that U' contains at least one paper from $P_{\{i,j\}}$. As $k = \kappa + \frac{\kappa\cdot(\kappa-1)}{2}$, we know that $|U' \cap V_i| = 1$ for all $i \in [\kappa]$, and $|U' \cap P_{\{i,j\}}| = 1$ for all distinct $i, j \in [\kappa]$. Let $K = U' \cap V(G)$. We prove that K is a clique in G by contradiction. Assume that K contains two vertices v and u respectively from two distinct V_i and V_j such that there exists no edge between v and u in G. Let $\{p(e)\} = U' \cap P_{\{i,j\}}$ where e is an edge in G crossing V_i and V_j. Clearly, $e \neq \{v, u\}$. Then, by the construction of D, $p(e)$ cites at least one of v and u. However, as $y_{\{i,j\}}$ cites both v and u, the CD index of $y_{\{i,j\}}$ in D' cannot be 1, a contradiction. So, K is a clique in G. From $|U' \cap V_i| = 1$ for all $i \in [\kappa]$, we know that K is a κ-colored clique in G. ☐

The above reduction also applies to CD-MANI-ADD$^\infty$.

Corollary 1. *For* $d = 1$, CD-MANI-ADD$^\infty$ *is* W[1]*-hard w.r.t.* $|J|$*. This holds even if the citation graph is acyclic.*

Theorem 7 dismisses the feasibility of achieving FPT-algorithms when parameterized by $|J|$, unless FPT=W[1]. Nevertheless, there still exists a potential for the problem to become tractable when $|J|$ is a constant. We first explore this prospect for the case where J is a singleton.

Theorem 8. CD-MANI-ADD *is polynomial-time solvable when* $|J| = 1$.

Proof. Let $(D, (V, U), J, k, d)$ be an instance of CD-MANI-ADD, where $J = \{v^\star\}$. Let $X = \{u \in U : V \cap \Gamma^+_D(v^\star) \cap \Gamma^+_D(u) \neq \emptyset\}$ be the set of all unregistered papers which cite at least one reference of $v^\star$ in V. Let $U' = \Gamma^-_{D[U]}(v^\star) \setminus (X \cup \Gamma^+_D(v^\star))$ be the set of all unregistered citations of $v^\star$ that do not cite any registered references of $v^\star$ and not cited by $v^\star$. Let $k' = \min\{k, |U'|\}$, and let U'' be any arbitrary k'-subset of U'. We conclude that I is a Yes-instance if and only if $v^\star$ has a CD index of at least d in $D[V \cup U'']$. □

As CD-MANI-ADD$^\infty$ is Turing reducible to CD-MANI-ADD, the former is also polynomial-time solvable when $|J| = 1$. We confirm next that for $d = 1$, tractability extends to all constant-bounded $|J|$.

Theorem 9. *For $d = 1$, CD-MANI-ADD and CD-MANI-ADD$^\infty$ are in XP w.r.t. $|J|$.*

However, for other positive values of d, the parameter $|J|$ complicates the problem, as stated in the following theorems.

Theorem 10. *For any positive rational number $d < 1$, CD-MANI-ADD is W[1]-hard w.r.t. k, even if $|J| = 2$ and the citation graph is acyclic.*

Theorem 11. *For any positive rational number $d < 1$, CD-MANI-ADD$^\infty$ is NP-hard, even if $|J| = 2$ and the citation graph is acyclic.*

7 Manipulation by Paper Deletion

We arrive at the problem of CD index manipulation by deleting papers. Clearly, the problem is in XP w.r.t. k. This poses the question if it is indeed FPT for the same parameter. In contrast to the negative answer for the problem of adding papers, we show that the values of d play a decisive role in this regard: CD-MANI-DEL is FPT w.r.t. k if and only if $d = 1$.

Theorem 12. *For $d = 1$, CD-MANI-DEL is FPT w.r.t. k.*

Theorem 13. *For any positive rational number $d < 1$, CD-MANI-DEL is W[1]-hard w.r.t. $|J| + k$. This holds even if the citation graph is acyclic.*

Theorem 13 implies that CD-MANI-DEL is W[1]-hard w.r.t. $|J|$. However, it does not rule out the possibility of polynomial-time solvability for $|J|$ being a constant. The following three theorems, collectively forming a complexity dichotomy, address this inquiry.

Theorem 14. *For $d = 1$, CD-MANI-DEL is polynomial-time solvable if $|J| = 1$.*

Proof. Let $I = (D, J, k, d)$ be an instance of CD-MANI-DEL, where $d = 1$. W.l.o.g., let $J = \{v^\star\} \subseteq V(D)$. We distinguish between two cases.

Case 1: $N_{\text{F}}^D(v^\star) \neq \emptyset$.

Let $V' = N_{\text{B}}^D(v^\star) \cup N_{\text{R}}^D(v^\star)$. We construct a bipartite graph G with the vertex partition $(\Gamma_D^+(v^\star), V')$. The edges of G correspond to arcs from V' to $\Gamma_D^+(v^\star)$ in D: if $v \in V'$ cites $u \in \Gamma_D^+(v^\star)$ in D, we create an edge between v and u in G. To make $v^\star$ have a CD index of 1 by deleting papers, for each $v \in V'$, either v needs to be deleted or all neighbors of v in G need to be deleted. This is equivalent to computing a minimum vertex cover in G, which can be solved in polynomial time, provided with G being bipartite [Lovász and Plummer, 1986; König, 1916; Hopcroft and Karp, 1973]. We conclude I is a Yes-instance if and only if G has a vertex cover of at most k vertices.

Case 2: $N_{\text{F}}^D(v^\star) = \emptyset$.

If $N_{\text{B}}^D(v^\star) = \emptyset$, we conclude I is a No-instance. Otherwise, to make $v^\star$ have a CD index of 1 by deleting papers, we need to delete certain papers from $\Gamma_D^+(v^\star)$ so that some paper v originally from $N_{\text{B}}^D(v^\star)$ becomes a member of the set $N_{\text{F}}^D(v^\star)$. We do not know the identity of v in advance, but there can be at most $|N_{\text{B}}^D(v^\star)|$ possibilities. In light of this, for every vertex $v \in N_{\text{B}}^D(v^\star)$, we do the following: (1) Decrease k by $|\Gamma_D^+(v) \setminus \{v^\star\}|$, and delete all outneighbors of v in D except $v^\star$. (2) Run the procedure described in Case 1, and if the output is "Yes", conclude that I is a Yes-instance, otherwise discard v. If all $v \in N_{\text{B}}^D(v^\star)$ are discarded, we conclude that I is a No-instance. □

Theorem 15. *For $d = 1$, CD-MANI-DEL is NP-hard, even if $|J| = 2$ and the citation graph is acyclic.*

Theorem 16. *For any positive rational number $d < 1$, CD-MANI-DEL is NP-hard, even if $|J| = 1$ and the citation graph is acyclic.*

Now we study the variant CD-MANI-DEL$^\infty$.

Theorem 17. *CD-MANI-DEL$^\infty$ is NP-hard even if $|J| = 3$.*

When J contains only one paper $v^\star$ with at least one citation, we can always achieve a CD index of 1 for $v^\star$ by deleting all papers except $v^\star$ and any arbitrary citation of $v^\star$. For $|J| = 2$, we can also obtain a polynomial-time algorithm.

Theorem 18. *CD-MANI-DEL$^\infty$ is polynomial-time solvable when $|J| \leq 2$ and when the citation graph is acyclic.*

For the case where $d = 1$, we obtain the following results.

Theorem 19. *For $d = 1$, CD-MANI-DEL$^\infty$ is W[1]-hard w.r.t. $|J|$, even when the citation graph is acyclic.*

Theorem 20. *For $d = 1$, CD-MANI-DEL$^\infty$ is in XP w.r.t. $|J|$.*

8 Concluding Remarks

We initiated the study of the CD index manipulation problems, offering a comprehensive understanding of the parameterized complexity of these problems for several significant parameters. See Table 1 and Table 2 for a summary. In general, we showed that these problems are hard to solve, with only a few exceptions. It is noteworthy that our results for CD-MANI-MERGE remain applicable even when the compatibility graph comprises a matching along with several isolated vertices, thereby ruling out fixed-parameter tractability of the problem concerning the treewidth of the compatibility graph. This conveys that, at least in theory, the CD index exhibits a high degree of reliability in the aspect of preventing manipulation. However, rather than concluding at this juncture, we aspire for our initiative to ignite further comprehensive investigations aimed at fully addressing questions surrounding the applicability of the CD index. To provide a more tangible direction, our study encourages a meticulous experimental exploration to verify whether these problems are difficult to solve in practice.

Acknowledgments

The author extends sincere appreciation to the three anonymous reviewers of IJCAI 2024 for their invaluable feedback and insightful comments.

References

[Bang-Jensen and Gutin, 2018] J. Bang-Jensen and G. Z. Gutin, editors. *Classes of Directed Graphs*. Springer, 2018.

[Bartholdi III *et al.*, 1992] J. J. Bartholdi III, C. A. Tovey, and M. A. Trick. How hard is it to control an election? *Math. Comput. Model.*, 16(8-9):27–40, 1992.

[Baumeister and Rothe, 2016] D. Baumeister and J. Rothe. Preference aggregation by voting. In J. Rothe, editor, *Economics and Computation: An Introduction to Algorithmic Game Theory, Computational Social Choice, and Fair Division*, chapter 4, pages 197–325. Springer, 2016.

[Bergamini *et al.*, 2018] E. Bergamini, P. Crescenzi, G. D'Angelo, H. Meyerhenke, L. Severini, and Y. Velaj. Improving the betweenness centrality of a node by adding links. *ACM J. Exp. Algorithmics*, 23:Nr. 1.5, 2018.

[Betzler *et al.*, 2014] N. Betzler, H. L. Bodlaender, R. Bredereck, R. Niedermeier, and J. Uhlmann. On making a distinguished vertex of minimum degree by vertex deletion. *Algorithmica*, 68(3):715–738, 2014.

[Cygan *et al.*, 2015] M. Cygan, F. V. Fomin, L. Kowalik, D. Lokshtanov, D. Marx, M. Pilipczuk, M. Pilipczuk, and S. Saurabh. *Parameterized Algorithms*. Springer, 2015.

[D'Angelo *et al.*, 2015] G. D'Angelo, L. Severini, and Y. Velaj. On the maximum betweenness improvement problem. In *ICTCS*, pages 153–168, 2015.

[D'Angelo *et al.*, 2019] G. D'Angelo, M. Olsen, and L. Severini. Coverage centrality maximization in undirected networks. In *AAAI*, pages 501–508, 2019.

[Das *et al.*, 2018] K. Das, S. Samanta, and M. Pal. Study on centrality measures in social networks: A survey. *Soc. Netw. Anal. Min.*, 8(1):Nr. 13, 2018.

[Erdélyi *et al.*, 2020] G. Erdélyi, C. Reger, and Y. Yang. The complexity of bribery and control in group identification. *Auton. Agent Multi-Ag.*, 34(1):Nr. 8, 2020.

[Erdélyi *et al.*, 2021] G. Erdélyi, M. Neveling, C. Reger, J. Rothe, Y. Yang, and R. Zorn. Towards completing the puzzle: Complexity of control by replacing, adding, and deleting candidates or voters. *Auton. Agent Multi-Ag.*, 35(2):Nr. 41, 2021.

[Faliszewski and Rothe, 2016] P. Faliszewski and J. Rothe. Control and bribery in voting. In F. Brandt, V. Conitzer, U. Endriss, J. Lang, and A. Procaccia, editors, *Handbook of Computational Social Choice*, chapter 7, pages 146–168. Cambridge University Press, 2016.

[Fellows *et al.*, 2009] M. R. Fellows, D. Hermelin, F. A. Rosamond, and S. Vialette. On the parameterized complexity of multiple-interval graph problems. *Theor. Comput. Sci.*, 410(1):53–61, 2009.

[Funk and Owen-Smith, 2017] R. J. Funk and J. Owen-Smith. A dynamic network measure of technological change. *Manag. Sci.*, 63(3):791–817, 2017.

[Hopcroft and Karp, 1973] J. E. Hopcroft and R. M. Karp. An $n^{5/2}$ algorithm for maximum matchings in bipartite graphs. *SIAM J. Comput.*, 2(4):225–231, 1973.

[Jiang and Liu, 2023] Y. Jiang and X. Liu. A construction and empirical research of the journal disruption index based on open citation data. *Scientometrics*, 128(7):3935–3958, 2023.

[König, 1916] D. König. Über Graphen und ihre Anwendung auf Determinantentheorie und Mengenlehre. *Math. Ann.*, 77(4):453–465, 1916.

[Leibel and Bornmann, 2024] C. Leibel and L. Bornmann. What do we know about the disruption index in scientometrics? An overview of the literature. *Scientometrics*, 129(1):601–639, 2024.

[Leydesdorff *et al.*, 2021] L. Leydesdorff, A. Tekles, and L. Bornmann. A proposal to revise the disruption index. *Profesional de la información*, 30:Nr. e300121, 2021.

[Li *et al.*, 2023] J. Li, W. Li, Y. Yang, and X. Yang. On the parameterized complexity of minimum/maximum degree vertex deletion on several special graphs. *Frontiers Comput. Sci.*, 17(4):Nr. 174405, 2023.

[Liu *et al.*, 2023] X. Liu, C. Zhang, and J. Li. Conceptual and technical work: Who will disrupt science? *J. Informetrics*, 17(3):Nr. 101432, 2023.

[Lovász and Plummer, 1986] L. Lovász and M. D. Plummer. *Matching Theory*. North Holland, 1986.

[Mathieson and Szeider, 2008] L. Mathieson and S. Szeider. The parameterized complexity of regular subgraph problems and generalizations. In *CATS*, pages 79–86, 2008.

[Mishra *et al.*, 2015] S. Mishra, A. Pananjady, and N. S. Devi. On the complexity of making a distinguished vertex minimum or maximum degree by vertex deletion. *J. Discrete Algorithms*, 33:71–80, 2015.

[Park *et al.*, 2023] M. Park, E. Leahey, and R. J. Funk. Papers and patents are becoming less disruptive over time. *Nature*, 613(7942):138–144, 2023.

[Pavlou and Elkind, 2016] C. Pavlou and E. Elkind. Manipulating citation indices in a social context. In *AAMAS*, pages 32–40, 2016.

[Singh, 2022] R. R. Singh. Centrality measures: A tool to identify key actors in social networks. In A. Biswas, R. Patgiri, and B. Biswas, editors, *Principles of Social Networking: The New Horizon and Emerging Challenges*, pages 1–27. Springer, 2022.

[van Bevern *et al.*, 2016] R. van Bevern, C. Komusiewicz, R. Niedermeier, M. Sorge, and T. Walsh. H-index manipulation by merging articles: Models, theory, and experiments. *Artif. Intell.*, 240:19–35, 2016.

[van Bevern *et al.*, 2020] R. van Bevern, C. Komusiewicz, H. Molter, R. Niedermeier, M. Sorge, and T. Walsh. h-index manipulation by undoing merges. *Quant. Sci. Stud.*, 1(4):1529–1552, 2020.

[Wan *et al.*, 2021] Z. Wan, Y. Mahajan, B. W. Kang, T. J. Moore, and J.-H. Cho. A survey on centrality metrics and their network resilience analysis. *IEEE Access*, 9:104773–104819, 2021.

[Wang *et al.*, 2023] R. Wang, Y. Zhou, and A. Zeng. Evaluating scientists by citation and disruption of their representative works. *Scientometrics*, 128(3):1689–1710, 2023.

[Waniek *et al.*, 2022] M. Waniek, T. P. Michalak, M. J. Wooldridge, and T. Rahwan. How members of covert networks conceal the identities of their leaders. *ACM Trans. Intell. Syst. Technol.*, 13(1):Nr. 12, 2022.

[Waniek *et al.*, 2023] M. Waniek, J. Woźnica, K. Zhou, Y. Vorobeychik, T. P. Michalak, and T. Rahwan. Hiding from centrality measures: A Stackelberg game perspective. *IEEE Trans. Knowl. Data Eng.*, 35(10):10058–10071, 2023.

[West, 2000] D. B. West. *Introduction to Graph Theory*. Prentice-Hall, 2000.

[Wu *et al.*, 2019] L. Wu, D. Wang, and J. A. Evans. Large teams develop and small teams disrupt science and technology. *Nature*, 566(7744):378–382, 2019.

[Yanai and Lercher, 2023] I. Yanai and M. J. Lercher. Make science disruptive again. *Nat. Biotechnol.*, 41(4):450–451, 2023.

[Yang *et al.*, 2023] A. J. Yang, H. Hu, Y. Zhao, H. Wang, and S. Deng. From consolidation to disruption: A novel way to measure the impact of scientists and identify laureates. *Inf. Process. Manag.*, 60(5):103420, 2023.

[Yang, 2019] Y. Yang. Complexity of manipulating and controlling approval-based multiwinner voting. In *IJCAI*, pages 637–643, 2019.