Mapping correlates of adolescent emotion dysregulation: A text-mining based systematic review

Caspar J. van Lissa1

1 Utrecht University faculty of Social and Behavioral Sciences, department of Methodology & Statistics

Author note

Correspondence concerning this article should be addressed to Caspar J. van Lissa, Padualaan 14, 3584CH Utrecht, The Netherlands. E-mail: [c.j.vanlissa@uu.nl](mailto:c.j.vanlissa@uu.nl)

Abstract

Adolesence is a developmentally sensititve period for emotion dysregulation. Although substantial empirical research has addressed relevant risk factors, an integrating framework is lacking. This systematic review of 6305 papers used text mining to map correlates of adolescents’ (ages 10-25) emotion dysregulation. First, we established a baseline of relevant terms gleaned from theory and recent narrative reviews. Then, we conducted two text mining analyses to examine term/document frequency and co-occurrence in author-provided keywords and abstracts. Results reﬂected constructs commonly featured in theory and narrative reviews, but also identified underrepresented terms. This research constitutes a first step towards integrating the literature, and illustrates how text mining reviews may complement narrative reviews.

*Keywords:* emotion regulation, adolescence, systematic review, text mining, machine learning

*Word count:* 5400

Mapping correlates of adolescent emotion dysregulation: A text-mining based systematic review

A key developmental challenge in adolescence is acquiring mature emotion regulation skills (Crone & Dahl, 2012; Zimmermann & Iwanski, 2014). Emotion regulation is integral to mental health (Aldao, Nolen-Hoeksema, & Schweizer, 2010; Braet et al., 2014; Schäfer, Naumann, Holmes, Tuschen-Caffier, & Samson, 2017) and social functioning (Reindl, Gniewosz, & Reinders, 2016). For as many as one in five, adolescence marks the onset of emotion regulation-related mental illness, which can persist throughout the life course (see Lee et al., 2014). It is therefore crucial to identify which risk factors and environmental hazards render some adolescents more susceptible to emotional difficulties than others. The present review set out to address this question. Although there is an abundance of empirical work on adolescents’ emotion regulation, the literature is somewhat fragmented, because the topic has been investigated from different (sub)disciplines. As a result, there is a lack of unifying theory, and narrative reviews fall short of comprehensive coverage of the vast but diffuse literature (Buss, Cole, & Zhou, 2019).

The present systematic review seeks to overcome this limitation by adopting a novel method - text mining based bibliographic analysis - to map all factors currently considered to be relevant to adolescent emotion regulation. Traditional narrative literature reviews have been shown to be marred by reliance on (small) convenience samples of the literature, confirmation bias, and an undue emphasis on positive results (Littell, 2008). Text mining, by contrast, is uniquely suited to provide a comprehensive overview of the literature. It can cover vastly greater corpora than a human reader, and the process by which text mining gleans insights from the literature is more objective, transparant, and reproducible. Where human readers might be inclined to structure their reading of the literature around established ideas, the inherent “fairness” of text mining gives emerging themes in the literature a chance to come to the fore. Text mining identifies important topics in a body of literature and reveals which constructs are often studied together. This potentially reveals topics and connections missed by narrative reviewers, thereby laying the groundwork for future empirical and theoretical work.

## Why adolescence is an important life stage for emotion regulation

Adolescence is defined as a life stage ranging from pubertal onset to adult-like independence (Steinberg, 2014). It is thus demarcated by both biological and socio-cultural factors. It has been argued that, due to accellerated pubertal onset and delayed transitions to adult roles, adolescence now ranges from 10-24 years in the modern Western world (Sawyer, Azzopardi, Wickremarathne, & Patton, 2018).

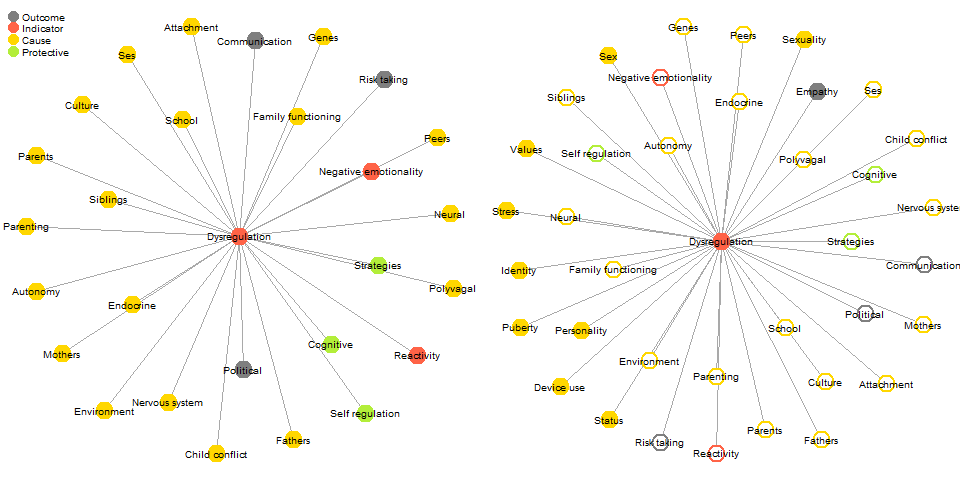
Adolescence is increasingly seen as a developmentally sensitive period for mature emotion regulation skills. During this period, children experience rapid biological, cognitive, and social changes that prompt new emotional experiences, and tax existing regulatory abilities (Steinberg, 2014). At the same time, the staggered development of motivational-emotional and regulatory brain circuits gives rise to a maturity gap, where youngsters pursue new experiences in life and love, without being fully prepared to cope with the emotional outcomes (Crone & Dahl, 2012). Due to these coalescing changes, adolescents display a restricted repertoire of emotion regulation strategies compared to younger or older individuals (Zimmermann & Iwanski, 2014). Consequently, adolescents experience more frequent and intense (negative) emotions than children or adults (see Silk, Steinberg, & Morris, 2003). The literature thus paints a picture of adolescence as a chrysalis for emotional development: Children enter this stage with emotion regulation skills adapted to the challenges of childhood. During adolescence, emotional systems are rearranged substantially. These changes render adolescents temporarily vulnerable to emotional dysregulation, but ultimately serve a functional role in acquiring mature emotion regulation skills. Eventually, most adolescents emerge ready to take on the challenges of early adulthood. Unfortunately, a sizeable proportion of adolescents instead manifests enduring emotional difficulties (Lee et al., 2014).

## The consequences of difficulties in emotion regulation

The consequences of difficulties in emotion regulation are extensive, well-documented, and persist throughout the life course. Meta-analyses indicate that such difficulties are associated with internalizing, and to a lesser extent, externalizing, psychopathology (Aldao et al., 2010; Schäfer et al., 2017). Individuals who engage in maladaptive emotion regulation strategies also experience more negative emotions, diminished well-being, and greater strain in interpersonal relationships (Bell & Calkins, 2000; Gross & John, 2003). Given the prevalence of emotional difficulties and their implications for individual mental health, wellbeing, and social functioning, and their downstream cost to society as a whole (Lee et al., 2014), it is essential to have a solid understanding of the risk factors associated with adolescents’ emotion regulation difficulties.

## Existing theoretical landscape

The results of our text mining analysis of the empirical literature are best understood against a backdrop of the existing theoretical landscape. We therefore provide a summary theoretical review. Other recent publications offer more detailed reviews of theories of emotional development (Buss et al., 2019) and emotion regulation in adolescence (Riediger & Klipker, 2014). The correlates of emotion regulation difficulties discussed in the theoretical review below are visually summarized in the left panel of Figure 1.



*Figure* *1:*. Document frequency of terms selected for Study 1.

Among the oldest relevant theories of adolescents’ emotional development is Hall’s notion of “storm and stress” (see Arnett, 1999). It describes how hormonal changes diminish self-control and increase reactivity, leading to difficulties in emotion regulation, conflict with parents, and risky behavior. In modern theories, this notion of diminished self-control and increased emotional reactivity has been retained - although it is now considered to be a normative change that facilitates emotional maturation (Crone & Dahl, 2012), at the risk of emotional disturbance (Arnett, 1999; Lee et al., 2014).

One example of a normative theory of emotional development in early childhood is presented by Sroufe (1995). Sroufe argues that, as children grow older, their increasing self-regulatory abilities drive a transition from external emotion regulation by primary caregivers towards autonomous emotion regulation. This theory focuses on two drivers of development: social and cognitive influences. Social influences mainly occur through parental co-regulation, parenting behaviors, and parent-child attachment. Cognitive influences occur through the development of the central nervous system (CNS), cognition, and self-regulation.

This emphasis on social influences and neurocognitive development is reflected in many theories. Some of these focus primarily on neurocognitive development. For example, polyvagal theory closely links emotional experience - and regulation - to autonomous nervous system functioning (Porges, 1995). A more recent neurocognitive theory relevant to adolescents’ emotion regulation development is Crone and Dahl’s model of social-affective engagement and goal flexibility (Crone & Dahl, 2012). Like other authors (e.g., Casey, Jones, & Somerville, 2011; Cracco, Goossens, & Braet, 2017), Crone and Dahl argue that the developmental asymmetry between motivational and inhibitory brain circuits gives rise to a “maturity gap” in middle adolescence. What distinguishes Crone and Dahl from related writings is the greater attention to adolescents’ diverging destinies: the question why some youngsters flourish while others languish. They argue that adolescents’ cognitive engagement is dynamically responsive to social and motivational goal salience. This flexibility, on the one hand, prepares adolescents to effectively engage cognitive systems in novel challenging situations in a way that facilitates developing mature regulatory abilities. On the other hand, it places them at risk to act impulsively in pursuit of peer approval. In terms of explanans, this theory focuses primarily on cognitive factors and the role of peers, with less attention to factors such as parenting. Although this theory focuses on adolescence specifically, an important limitation is that it only tangentially relates to emotion regulation.

Other theories emphasize socialization to a greater extent. One such theory is Morris’ tripartite model, which focuses specifically on parents’ role in emotion regulation development (2007). It describes three pathways through which parents shape emotion regulation development: Through observation and modeling, parenting practices, and the emotional family climate, which in turn involves attachment and marital relationship quality. Morris and colleagues also emphasize the relevance of fathers and siblings, and others have adapted the tripartite model to describe the role of peers in adolescents’ emotion regulation development (Reindl et al., 2016). A more abstract take on socialization is found in Holodynski and Friedlmeier (2006)’s internalization model of emotional development. Unique aspects emphasized in this theory are the interplay between emotion and communication, and the internalization of the cultural symbolic function of emotion. These theories explain socialization in great detail, but one potential limitation is that they are not embedded in a larger unifying framework.

Looking beyond the developmental literature, theories of the phenomenon of emotion regulation offer additional insight into intra-individual drivers of emotion regulation development. Perhaps the most influential theory is Gross’ (2013) process model, which describes the process of emotion regulation from eliciting cue to ultimate response. According to Gross, individuals use strategies to modulate the different stages of this process, consciously or otherwise. Culture plays a role too, as the effectiveness, desireability, and consequences of different strategies depend on the social context (see Bariola, Gullone, & Hughes, 2011). Similar to Gross’ theory, the social information processing theory of emotion also describe the role of cognitive processes and strategies in emotion experience and regulation (Lemerise & Arsenio, 2000).

Finally, given our interest in risk factors that render adolescents susceptible to emotional difficulties, we should mention two theories widely invoked to frame research on developmental influences: Bronfenbrenner’s bioecological model, and Sameroff’s transactional model. Bronfenbrenner’s model (Bronfenbrenner & Morris, 2007) describes how the environment shapes individual development. Each individual is imbued with certain biological predispositions, and develops over time, in interaction with contextual influences. These influences range from the microsystem, composed of those people close to the individual, to the macrosystem, consisting of indirect political and economic influences, to the exosystem, consisting of cultural norms and values. Sameroff’s model is somewhat compatible with the bioecological model (Sameroff, n.d.), but focuses more on the interplay between nature and nurture. It conceptualizes development as a product of reciprocal influences between the child and the environment. Sameroff distinguishes between proximal influences, for instance, through parents (the microsystem in Bronfenbrenner’s model), and distal influences; structural factors indirectly shaping development, like socio-economic status, schools, and the community (macro- and exosystem). With increasing age, distal influences gradually gain ground on proximal influences. These theories are broad enough to contextualize any developmental study, but limited by a lack of specificity, which hampers their utility in generating hypotheses.

### Shortcomings of existing theory.

Despite the abundance of *relevant* theory, the existing literature has several limitations when the goals is to identify which risk factors render some adolescents more susceptible to maladaptive emotional development than others. First, few theories have explicitly addressed the life stage of adolescence - a life stage that differs qualitatively from infancy, childhood, and adulthood (Bariola et al., 2011). It is therefore questionable whether more general theories, or those focused on different age groups, can be generalized to adolescents. Furthermore, few theories have comprehensively addressed important predictors of development in this life stage, and none directly guide contemporary research in the field (see Buss et al., 2019; Riediger & Klipker, 2014). Finally, available theories vary widely in scope: Some are broad but somewhat non-specific, and others describe a specific phenomenon in detail, but lack a broader perspective. It would be beneficial to bridge these levels of analysis. In sum, there is a need for more integrative theory formation, in order to provide a unified framework that could guide future empirical work. The present study lays the groundwork for such theory development, by providing a text-mining based inductive systematic review of the field.

## Bottom-up insights from narrative reviews

Whereas theory provides a top-down understanding of an area of research, literature reviews can provide a bottom-up, inductive understanding. Given the noted absence of a single overarching theory, reviews might provide additional insight into factors considered relevant in the etiology of adolescents’ emotion regulation competencies. Before we attempt to do so, using innovative text mining methods, we will review the ways in which existing narrative reviews complement the aforementioned theoretical literature. The right panel of Figure 1 visually summarizes unique concepts covered by these narrative reviews, relative to the issues addressed in the theoretical literature (left panel). Terms covered in the theoretical literature are displayed as transparent circles, to illustrates in what respects the narrative reviews complement the theoretical literature.

One highly cited review by Bariola et al. (2011) addresses individual psychological factors such as temperament, and biological factors like neurocognitive development and genetic predispositions. Furthermore, the authors address proximal social influences, including parents, teachers, and peers, and specific mechanisms through which these individuals exert their influence, including parenting and modeling. Finally, they address distal factors, including the role of culture and the media. Barriola and colleagues also call for future research to address parents’ role beyond early childhood. They further argued that the role of fathers in emotion socialization warrants further scrutiny - a topic that is now receiving more attention (e.g., see Van Lissa, Keizer, Van Lier, Meeus, & Branje, 2019; Van Lissa & Keizer, 2020).

A more recent review by Coe-Odess, Narr, and Allen (2019) offers a nuanced discussion of several issues that complement prior publications in this field. For instance, the authors cover the implications of physiological changes in detail, including neurocognitive development and pubertal maturation. As adolescents become increasingly individuated, conflict with parents peaks. This conflict, in turn, impacts emotion regulation, both in terms of day-to-day mood variability and dispositional emotion dysregulation (see also Van Lissa, Hawk, Koot, Branje, & Meeus, 2017). Compared to children, adolescents also become increasingly oriented towards peers. This includes increased sensitivity to social status and norms, along with the associated increase in peer pressure and risk taking. Pubertal development also initiates the emergence of sexual and romantic behavior, and the intensification of both biological sex differences and gender stereotyped behavior. With regard to the process of emotion regulation, Coe-Odess and colleagues discuss the importance of strategies, and point out that, in addition to negative emotionality, positive emotionality also peaks in adolescence. Relatedly, the authors describe how hormonal changes precipitate an increased stress response, which might help explain adolescents’ greater susceptibility to emotion dysregulation. Finally, going beyond the implications of cognitive development discussed in other publications, the authors discuss how an greater capacity for abstract thought relates to identity formation - a key challenge in adolescence (Meeus, 2011) - and to increased emotional understanding, and by extension, empathy (see also Van Lissa et al., 2014).

As can be seen in Figure 1, there are notable parallels between the relevant factors highlighted by these narrative reviews, and the aforementioned theoretical literature. Nevertheless, these literature reviews also touch upon issues that have received little attention in theoretical work. These include specific individual differences at the biological and psychological level, including genetic predisposition, hormones, pubertal onset, gender, sexuality, temperament, identity, empathy, and stress. Further introduced are proximal influences, including the role of media, norms, and social status. This illustrates that reviews of the empirical literature can contribute additional insight into key factors in adolescents’ emotion regulation development. However, an important remaining limitation is that all reviews in this field have been unstructured narrative reviews, which are known to be limited in scope and biased (Littell, 2008). We seek to complement preceding work by using text mining to conduct a more comprehensive empirical literature review, and map the factors associated with adolescent emotion regulation.

## The present paper

The present paper aims to develop a nomological network of the constructs related to adolescent emotion dysregulation, based on a systematic review of the literature. A nomological network is a diagrammatic representation of a phenomenon; a proto-theoretical device that describes relationships between constructs relevant to the theory (see Alavi, Archibald, McMaster, Lopez, & Cleary, 2018). It is a precursor to a more explicit theory of the phenomenon. The present paper uses an inductive approach to develop this nomological network. First, we conducted a systematic search to elicit a corpus of relevant literature. Second, we used text mining techniques to extract relevant constructs from the corpus. Third, we used a dictionary to pare down the extracted constructs to meaningful superordinate categories. Fourth, we mapped interrelations between these constructs. Fifth, we classified contstructs as potential predictors or outcomes. .

# Methods

## Search strategy

All searches were conducted on Web of Science. The search strategy was based on procedures described by Staaks (Staaks, n.d.). First, we manually constructed a reference set of 15 articles. Then we constructed a search string to retrieve these articles from Web of Science. The search was overly inclusive, returning 29 records. As these were all highly relevant, we updated the reference set to include all 29 results. Next, we tested our search string, which consisted of synonyms of emotion regulation and adolescence:

TS=("emotio\* regulation" OR "anger regulation" OR "sadness regulation" OR  
"emotion\* competence" OR "emotion\* adjustment" OR "emotio\* dysregulation" OR  
"anger dysregulation" OR "sadness dysregulation" OR "emotio\* problem\*" OR  
"emotion\* maladjustment") AND TS=(adolescent\* OR teen\* OR pubert\* OR youth)

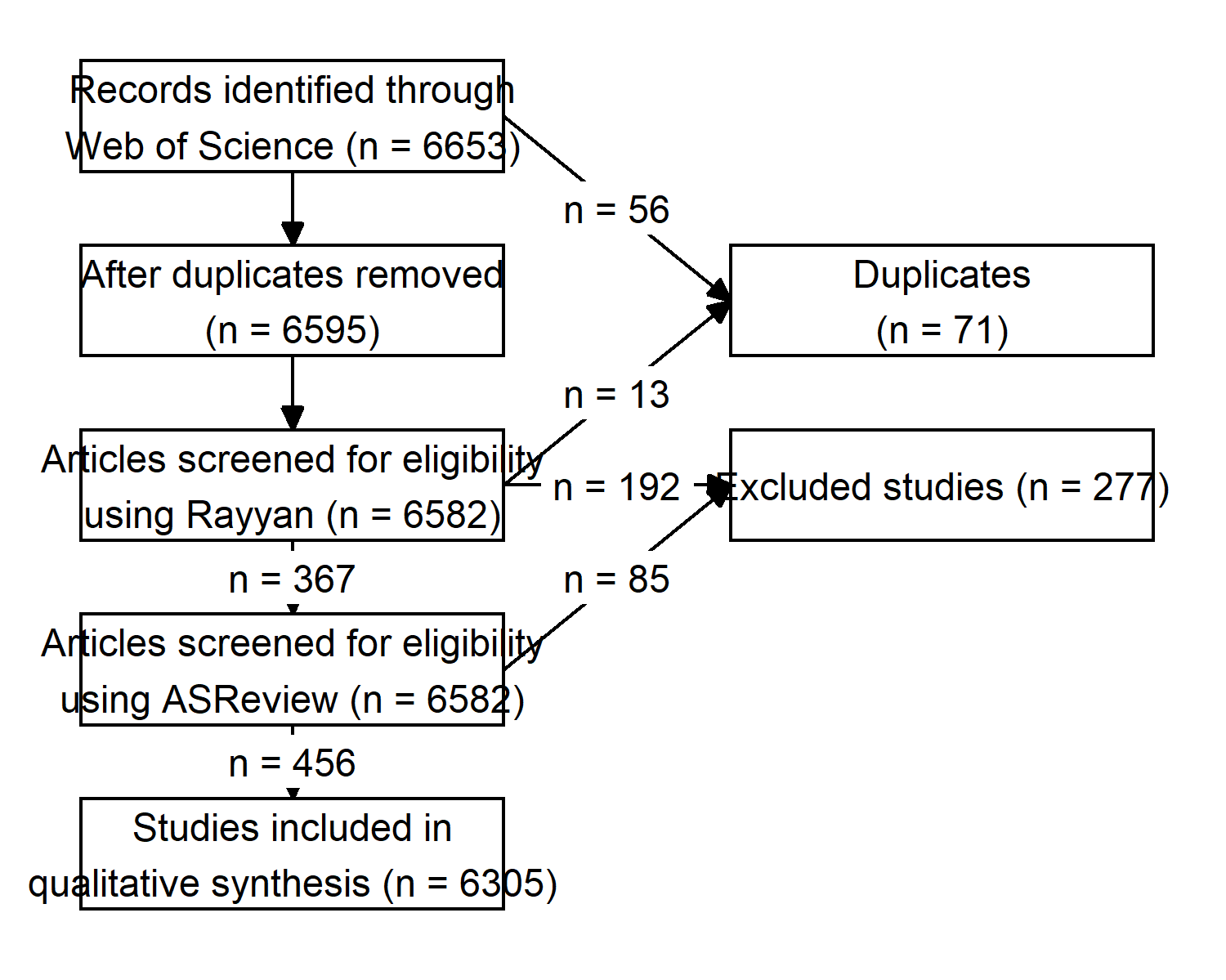
This string returned 6653 results, and matched 25 of the 29 records in the reference set. Three search terms could be added to match all 29 reference set items. The terms "emotio\* socialization" OR "emotio\* processes", as synonyms of emotion regulation, added 191 new hits. However, most of these did not meet the inclusion criteria used for screening, as explained below. The term "development\*", as synonym of adolescence, added 3628 new hits, many of which were not focused on the age range of adolescence. We thus deemed these terms to be overly inclusive, and proceeded with the original search string above.

## Deduplication

Duplicates were identified based on exact DOI matches (2 duplicates), and title similarity (54 duplicates). Manual screening in Rayyan resulted in the removal of an additional 13 duplicates.

## Screening

Papers were screened based on two main criteria: First, papers had to address emotion regulation or a synonymous construct. Second, the target population must overlap with the life stage of adolescence, defined as the age range from 10-24. Thus, for example, an article reviewing the role of hormonal changes in emotional adjustment throughout the entire life course, including adolescence, would be deemed relevant. Screening was initially conducted in Rayyan QRCI (Ouzzani, Hammady, Fedorowicz, & Elmagarmid, 2016). Papers were sorted using Rayyan’s ranking algorithm. To train the algorithm on exclusion criteria, screening focused on abstracts most likely to be excluded. After coding 559 papers, screening continued in ASReview - a free, open source alternative to Rayyan QRCI (Van de Schoot et al., 2020). This program uses text mining and active learning to build a customizable machine learning model for article inclusion. We used a naive Bayes model, and focused on exclusion of irrelevant papers. In ASReview, an additional 541 papers were screened. Screening again focused on articles most likely to be excluded, and continued until, among the most recently screened 100 papers, only 6 were excluded. The screening procedure is detailed in Figure 2. In total, 6305 papers were retained for analysis.



*Figure* *2:*. Record screening flowchart

# Analysis 1: Author keyword mapping

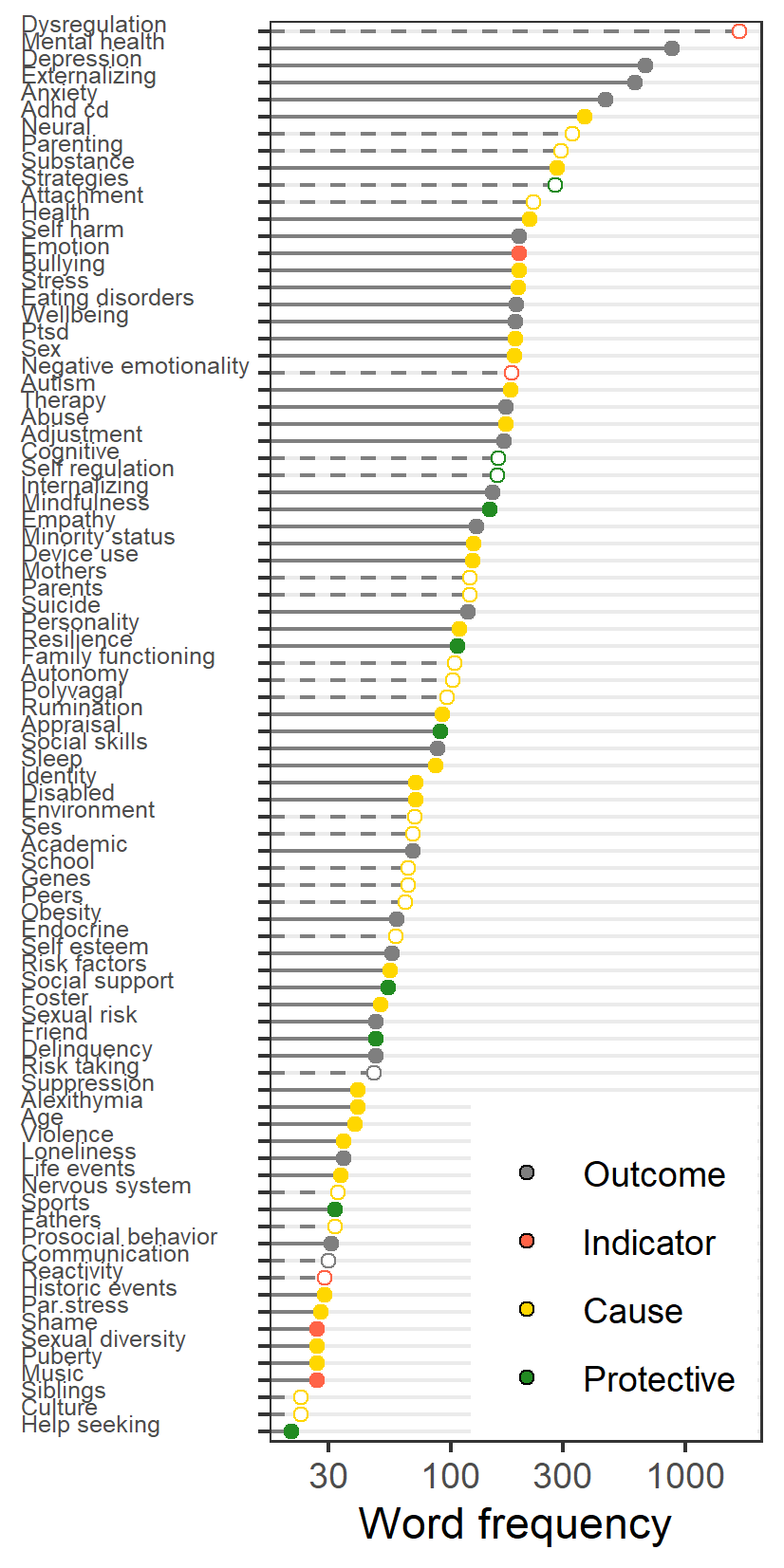
The corpus for this first analysis consisted of the author-provided keywords for the selected articles. We extracted keywords by document, and used a dictionary created specifically for this study to classify related terms. The dictionary describes 109 terms using 462 regular expression queries. The author composed a first draft of the dictionary. This draft was subsequently updated in two ways: First, key terms from the literature were added to the dictionary while writing the Introduction. Second, all frequently occurring uncategorized words were added to the dictionary, until none remained. An exclusion list of methodological terms and other non-substantive words was generated using the same procedure. After applying the dictionary and exclusion filter, and removing records with no (remaining) keywords, the resulting corpus consisted of 5031 documents with 8080 unique words.

## Structural topic modeling

To assess the homogeneity of the corpus, we conducted topic modeling using latent dirichlet allocation (Blei, Ng, & Jordan, 2003). For this analysis, we used the term frequency/inverse document frequency (TF-IDF) to select terms used frequently in a document, but not used frequently in the corpus, which could therefore be more diagnostic of subgroup membership. Selection terms with an TF-IDF greater than the median resulted in a corpus of 2362 documents and 3218 terms. We considered a range from 2-20 topics, evaluating fit based on the BIC, and interpretability based on the entropy of the posterior document/topic probabilities. The BICs followed a near-perfect linearly increasing trend, indicating that no subcorpora could be identified. Congruently, all entropies were near-zero, which means that the posterior document/topic probabilities were effectively uniformly distributed. Thus, no interpretable subcorpora could be identified, and we proceed with an analysis of the whole sample.

## Identifying common terms

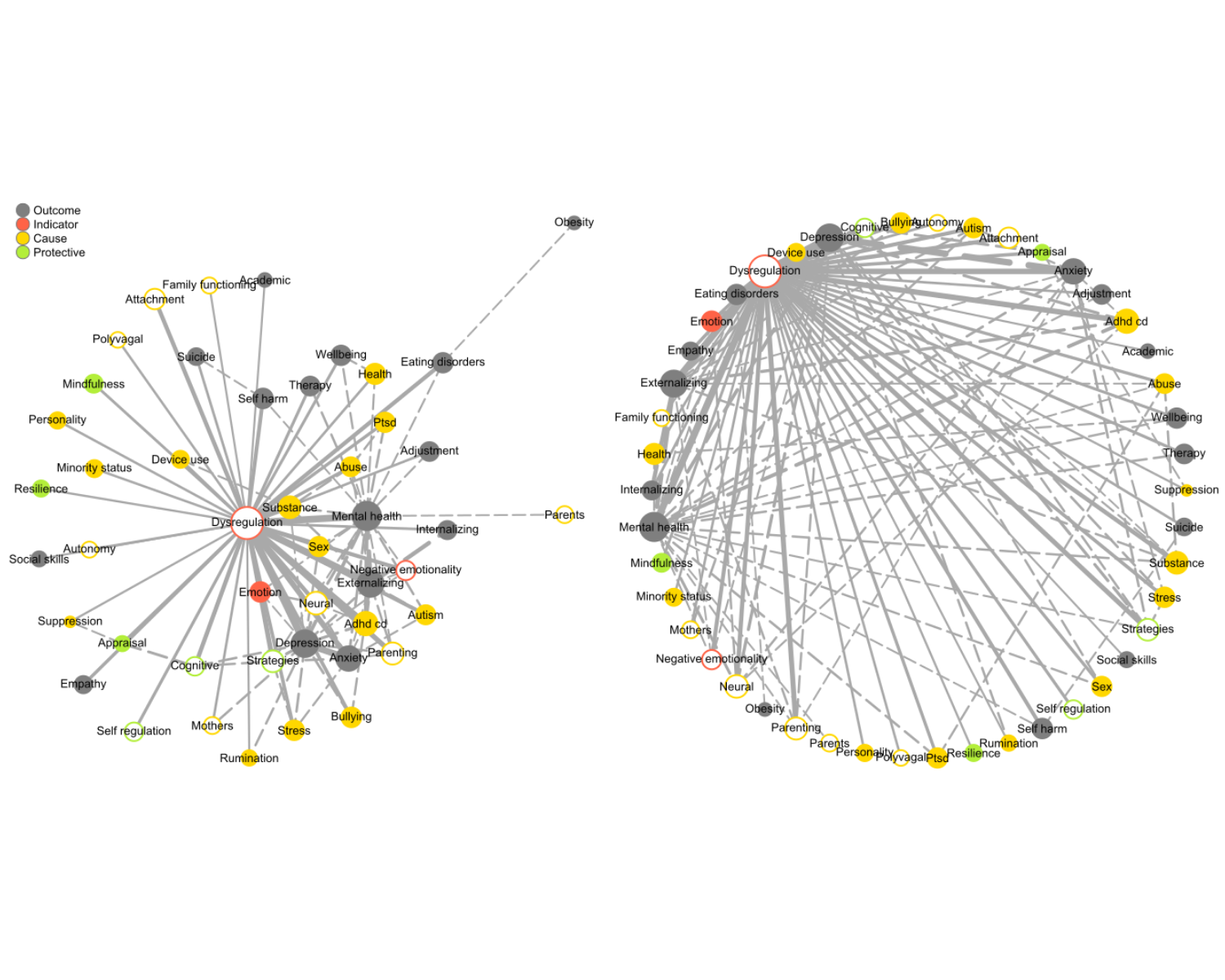
To identify what issues are covered in this corpus, we examined the terms occurring in the largest number of texts. We modeled word frequency using a negative binomial distribution, and retained those exceeding the percentile. This resulted in 83 terms, which occurred in 4826 documents. To illustrate what issues are covered in this body of literature, the selected terms are displayed in Figure 3, on a logarithmic scale to reduce the skew of the negative binomial distribution.



*Figure* *3:*. Document frequency of terms selected for Study 1.

## Mapping the literature

Based on our theoretical review, each of the selected terms was coded as being either a putative ‘Cause’, ‘Outcome’, ‘Protective factor’, or ‘Indicator’ (a closely related construct) of emotion dysregulation. We stress that this is a rudimentary and simplistic classification; many of these associations are likely to be bidirectional (e.g., emotion dysregulation is known to be both a cause and consequence of conflict with parents, Van Lissa et al., 2017). Next, we computed a term co-occurrence matrix, which represents how frequently words occurred within the same document. To aid interpretability, we again pruned small coefficients by modeling co-occurrence using a negative binomial distribution, and retaining co-occurrences exceeding the percentile. Figure 4 displays the pruned term cooccurrence matrix as a force directed graph.



*Figure* *4:*. Force directed graph for Study 1 co-occurrence analysis. Vertex size represents term frequency; dashed lines for co-occurrences not involving dysregulation. Two different layouts are presented to aid interpretation.

## Results

Based on term document frequency (Figure 3), it is clear that emotion dysregulation and its mental health-related outcomes are foremost among the common terms in the corpus. Other frequent terms reflect important themes discussed in the theoretical review of the literature; for instance, the term *neural*, *parenting*, and *stress* correspond to major themes discussed by Coe-Odess et al. (2019): Neurocognitive development, the role of the parents, and adolescents’ increased stress response. Importantly, the most common terms also include several concepts not featured prominently in the theoretical review. For example, *ADHD/CD* (cf. Braet et al., 2014), *substance* use (cf. Coe-Odess et al., 2019; Pierrehumbert et al., 2002), and *minority status* (cf. Myers, 2009, 2009) are common in the corpus, but featured less prominently in the theoretical review.

Based on the co-occurrence graph (Figure 4), it is evident that emotion dysregulation is a central construct, and that most of the other constructs are linked directly to emotion dysregulation, which suggests that our search was successful in identifying factors considered to be relevant for adolescents’ emotion dysregulation. Otherwise, the graph is relatively sparse, with few interconnections between terms.

## Discussion

This first analysis used author keywords to map common topics in the literature pertaining to adolescents’ emotion dysregulation. The results reflected some of the constructs commonly accepted as relevant in theoretical literature and empirical reviews. However, the analysis also revealed important areas less prominently featured in narrative reviews. This illustrates how inductive reviews may complement existing theory and narrative reviews, and can reveal blindspots.

Connections among topics were relatively sparse: Many constructs were only related to emotion regulation, and there were relatively few interconnections between elements. This observation substantiates the prior claim that this literature is somewhat fragmented (Buss et al., 2019).

# Analysis 2: Abstract text mining

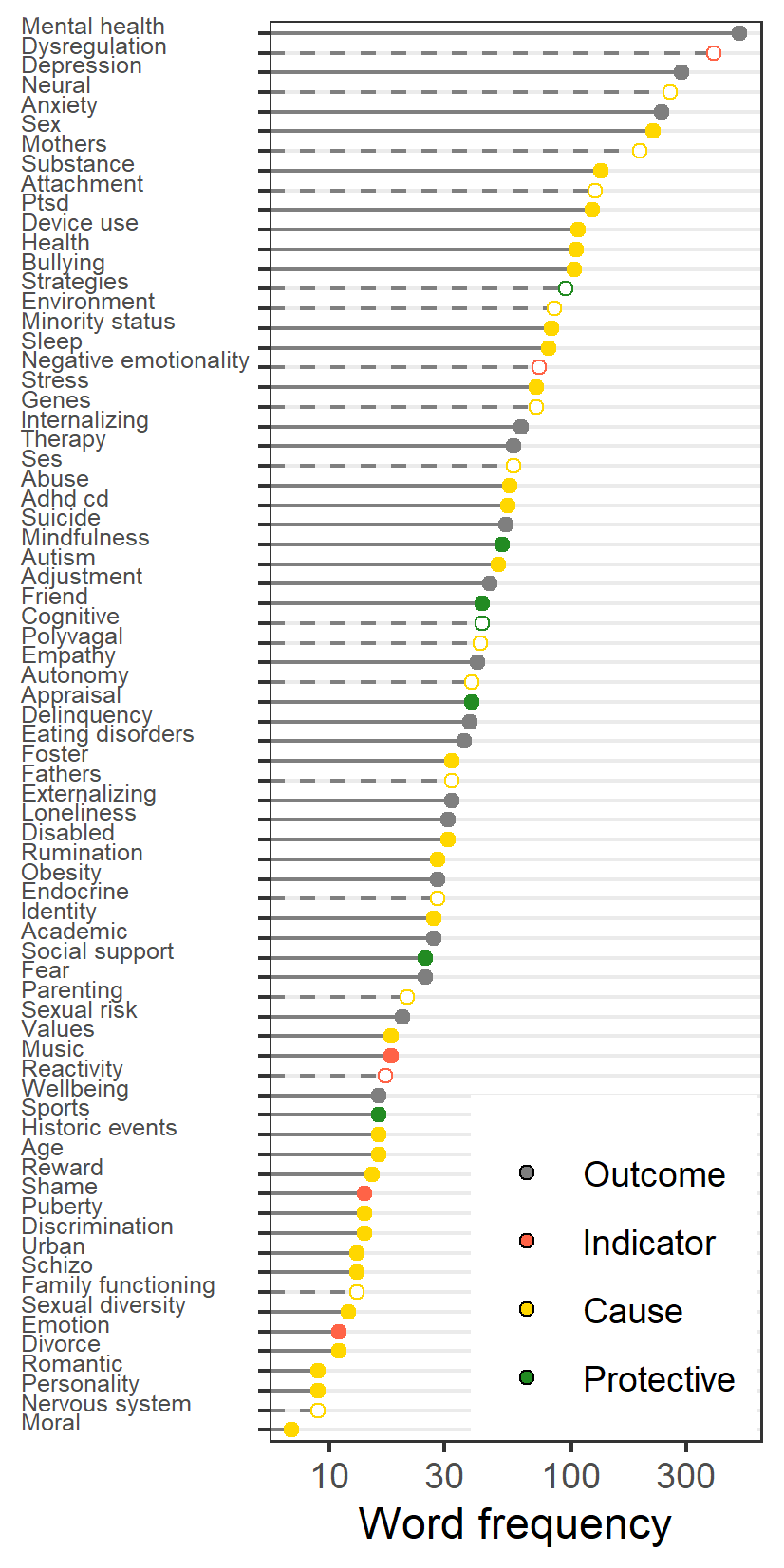
The corpus for this second analysis consisted of the abstracts of the selected articles. Keywords, as examined in Analysis 1, convey high-quality information because they are carefully chosen by authors to capture the essence of a study. However, as authors are typically limited to 5 keywords, some nuance may be lost. Abstracts, by contrast, offer greater freedom of expression, but present a greater challenge when it comes to extracting relevant information. To perform feature extraction, we first applied the natural language processing technique “part-of-speech tagging” (POS-tagging), which identifies a word’s grammatical function within the sentence context. Then, we used POS-tagging to retain only nouns and adjectives, which helps derive more interpretable text mining models (Martin & Johnson, 2015). Finally, we used stemming to reduce the retained terms to their root form.

## Structural topic modeling

To assess the homogeneity of the corpus of abstracts, we again conducted topic modeling (Blei et al., 2003), with features selected based on TF-IDF. This resulted in a dataset consisting of 6076 documents and 7800 terms. We considered a range from 2-20 topics, evaluating fit based on the BIC, and interpretability based on the entropy of the posterior document/topic probabilities. As before, the BICs followed a near-perfect linearly increasing trend, and all entropies were near-zero. Thus, no interpretable subcorpora could be identified, and we proceed with an analysis of the whole sample.

## Identifying keywords

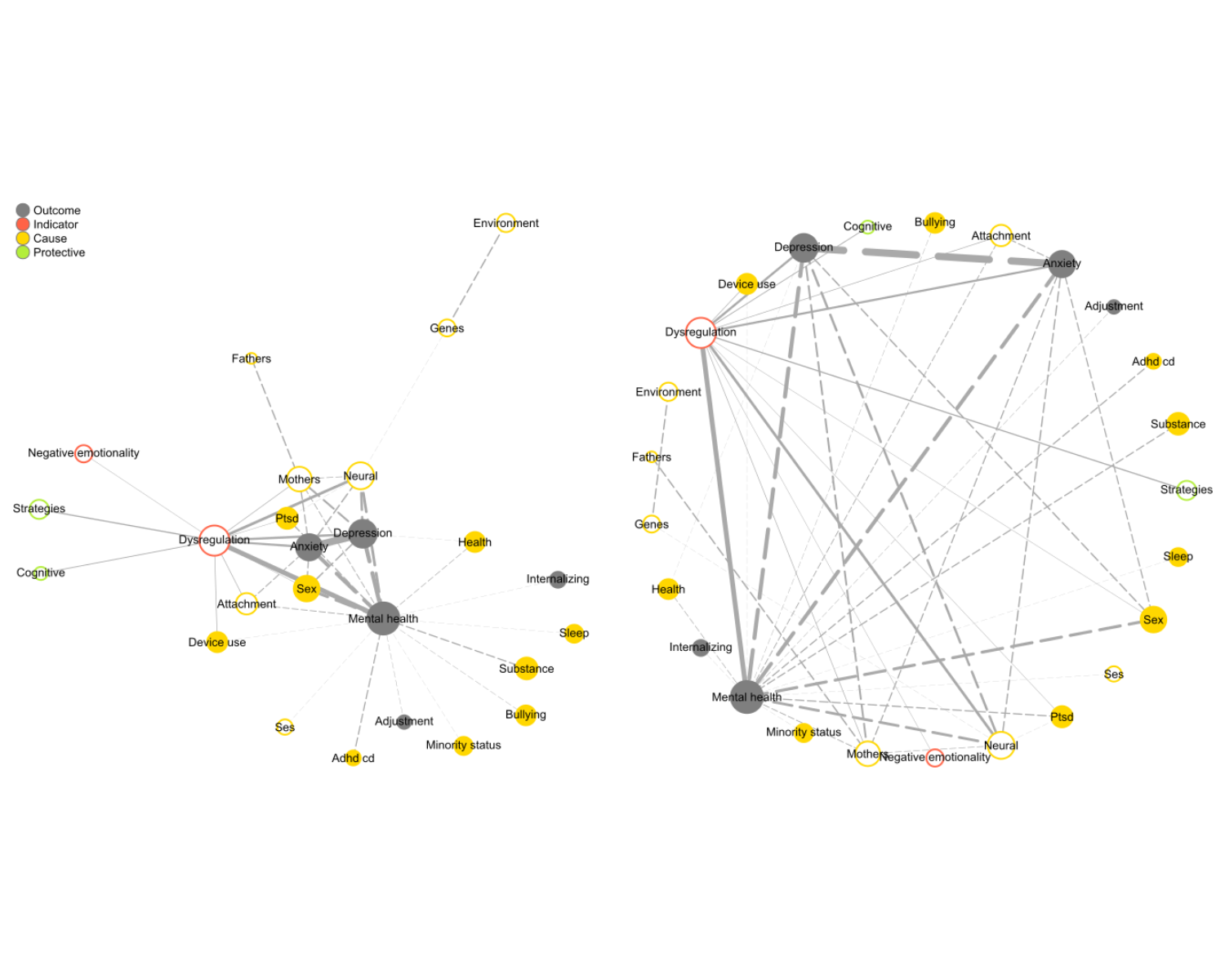
When conducting text mining analysis on unstructured text data (as opposed to author key words), focusing on individual words out of context can reduce interpretability. For instance, our core construct “emotion dysregulation” is already a bigram. To identify more meaningful units of analysis, we applied the textrank algorithm (Wijffels, 2019), an extention of Google’s PageRank (US7058628B1, 2006), to identify relevant -grams (with ). The resulting -grams were merged with the original data. After applying the dictionary and exclusion filter as explained in Analysis 1, the resulting corpus consisted of 5097 documents with 11435 unique words. Terms exceeding the percentile of a negative binomial distribution were retained for analysis. The identified important keywords are displayed in Figure 5.



*Figure* *5:*. Document frequency of terms selected for Study 2.

## Mapping the literature

A term co-occurrence matrix was constructed following the procedure described in Analysis 1. Figure 6 displays the resulting matrix as a force directed graph.



*Figure* *6:*. Force directed graph for Study 2 co-occurrence analysis. Vertex size represents term frequency; dashed lines for co-occurrences not involving dysregulation. Two different layouts are presented to aid interpretation.

## Results

Based on term document frequency, it is clear that emotion dysregulation and mental health-related outcomes of emotion dysregulation are among the most common terms in the corpus of abstracts. The term “neural”, again, occurred frequently, but other top terms were somewhat less predictable than those emerging in Analysis 1. The term “sex” is more frequent, and instead of “parenting”, we now see “mothers” on top of the list. We see that “minority status”, “substance use”, “attachment”,“ptsd”, and “bullying” are once more among the most frequent words.

Inspecting the co-occurrence graph for the keyword data, we see a much sparser network than in Analysis 1, with a different structure. We now see that dysregulation and mental health-related terms form a central axis, and that many related terms are connected to this axis, but not always directly to dysregulation. Only the terms “neural”, “mothers”, “attachment” and “ptsd” are connected directly to dysregulation.

There was substantial agreement between the findings of Analysis 1 and 2; of the most frequent terms identified in the author keywords (Analysis 1) were also present in the frequent terms extracted from the abstracts using textrank (Analysis 2), and conversely, of the most frequent terms from Analysis 2 were present in Analysis 1.

## Discussion

This second analysis focused on the abstracts of the reviewed articles. We observed a sparser network, with fewer selected terms. This might be explained by the unstructured format of abstracts, which likely introduces greater noise in the analysis. Thus, fewer terms will exceed the detection threshold. Nonetheless, the selected terms displayed substantial consistency with the analysis of author keywords. This supports the validity of the findings, suggesting that automatic keyword extraction from abstracts can identify relevant constructs, and may constitute a suitable alternative to author-provided keywords.

The emergence of a “central axis” of dysregulation and mental health-related outcomes suggest that these phenomena are consistently studied together. By contrast, the lack of interconnectivity among other related constructs again suggests that the literature is somewhat fragmented. This impression is reinforced by the absence of many frequently occurring terms from the term co-occurrence graph. Finally, the substitution of “parenting” with “mothers” amonst the top-ranking terms might suggest that parenting is commonly operationalized in terms of mothering (Pleck, 2004).

# General Discussion

The present study conducted a systematic review of the literature, and performed bibliographic analysis to map factors currently considered to be relevant to adolescent emotion regulation in a nomological network. The results of this inductive approach echoed many of the constructs considered relevant in the theoretical literature. It also revealed several constructs that have been underrepresented in theory, but nonetheless occur frequently in the empirical literature. The structure of the term co-occurrence networks revealed that most of the relevant constructs are connected directly with emotion dysregulation or its mental health correlates, with relatively few interconnections amongst constructs. This echoes the sentiment by Buss et al. (2019) that this literature is somewhat fragmented. The present systematic review takes a first step towards integrating this diffuse field.

Of course, the present study also has limitations. One such limitation is the fact that the text mining techniques used here are not able to extract *meaning* from the literature in the way a sentient reader would. This limitation is best addressed by using the output of our text mining analysis as a starting point for further inductive thought, or a more in-depth reading of a subset of the reviewed literature. Relatedly, our methods do not capture the nature of the relationship between co-occurring terms. Instead, we manually classified terms as potential causes, outcomes, protective factors, or indicators based on domain knowledge. It is worth noting that efforts are underway to develop unsupervised algorithms capable of distilling causal links from bodies of scientific abstracts (An, Xiao, Yuan, Yang, & Alterovitz, 2019). Applying such methods to the present corpus might substantially advance theory formation.

One final limitation of the present study is that our analysis was limited to keywords and abstracts. Analysis of full-text papers would undoubtedly yield additional insights. Regrettably, paywalls and the lack of a unified API currently prevent such analysis. This highlights the importance of open science: With comprehensive open access publishing, we may one day be able to mine the entire published literature for insights. A proof of concept for this approach was recently provided by the release of 50k+ full-text papers on topics related to COVID-19 for text mining analysis (Wang et al., 2020).

To conclude, this paper set out to map the factors associated with adolescents’ emotion dysregulation, based on a systematic review of the literature, and text mining analysis of author keywords and abstracts. We provided a map of the literature on adolescents’ emotion dysregulation, covering both familiar ground, and charting unknown territories. Our results suggested that several key constructs are currently underrepresented in theory, and that the empirical literature is somewhat fragmented by sub-topic. These inductive insights are relevant for the design of new deductive studies which, even though grounded in theory, might benefit from incorporating additional relevant factors to better elucidate the nature of their association with adolescents’ emotion dysregulation. Our analysis also identified several factors currently on the fringes of the literature, which might represent emerging themes that would benefit from further research (including, for example, the role of fathers Van Lissa & Keizer, 2020). Most importantly, given the lack of an overarching theory tailored specifically to adolescent emotion dysregulation, our nomological networks might serve as a starting point for theory development, with the aim of integrating the present inductive insights within the framework of existing relevant theory.

# References

Alavi, M., Archibald, M., McMaster, R., Lopez, V., & Cleary, M. (2018). Aligning theory and methodology in mixed methods research: Before Design Theoretical Placement. *International Journal of Social Research Methodology*, *21*(5), 527–540. <https://doi.org/10.1080/13645579.2018.1435016>

Aldao, A., Nolen-Hoeksema, S., & Schweizer, S. (2010). Emotion-regulation strategies across psychopathology: A meta-analytic review. *Clinical Psychology Review*, *30*(2), 217–237.

An, N., Xiao, Y., Yuan, J., Yang, J., & Alterovitz, G. (2019). Extracting causal relations from the literature with word vector mapping. *Computers in Biology and Medicine*, *115*, 103524. <https://doi.org/10.1016/j.compbiomed.2019.103524>

Arnett, J. J. (1999). Adolescent storm and stress, reconsidered. *American Psychologist*, *54*(5), 317–326. <https://doi.org/10.1037/0003-066X.54.5.317>

Bariola, E., Gullone, E., & Hughes, E. K. (2011). Child and adolescent emotion regulation: The role of parental emotion regulation and expression. *Clinical Child and Family Psychology Review*, *14*(2), 198. <https://doi.org/10.1007/s10567-011-0092-5>

Bell, K. L., & Calkins, S. D. (2000). Relationships as Inputs and Outputs of Emotion Regulation. *Psychological Inquiry*, *11*(3), 160–163.

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet Allocation. *Journal of Machine Learning Research*, *3*(Jan), 993–1022.

Braet, C., Theuwis, L., Durme, K. V., Vandewalle, J., Vandevivere, E., Wante, L., … Goossens, L. (2014). Emotion regulation in children with emotional problems. *Cognitive Therapy and Research*, *38*(5), 493–504. <https://doi.org/10.1007/s10608-014-9616-x>

Bronfenbrenner, U., & Morris, P. A. (2007). The bioecological model of human development. In *Handbook of Child Psychology*. John Wiley & Sons, Inc. <https://doi.org/10.1002/9780470147658.chpsy0114>

Buss, K. A., Cole, P. M., & Zhou, A. M. (2019). Theories of Emotional Development: Where Have We Been and Where Are We Now? In V. LoBue, K. Pérez-Edgar, & K. A. Buss (Eds.), *Handbook of Emotional Development* (pp. 7–25). Cham: Springer International Publishing. <https://doi.org/10.1007/978-3-030-17332-6_2>

Casey, B. J., Jones, R. M., & Somerville, L. H. (2011). Braking and accelerating of the adolescent brain. *Journal of Research on Adolescence*, *21*(1), 21–33. <https://doi.org/10.1111/j.1532-7795.2010.00712.x>

Coe-Odess, S. J., Narr, R. K., & Allen, J. P. (2019). Emergent Emotions in Adolescence. In V. LoBue, K. Pérez-Edgar, & K. A. Buss (Eds.), *Handbook of Emotional Development* (pp. 595–625). Cham: Springer International Publishing. <https://doi.org/10.1007/978-3-030-17332-6_23>

Cracco, E., Goossens, L., & Braet, C. (2017). Emotion regulation across childhood and adolescence: Evidence for a maladaptive shift in adolescence. *European Child & Adolescent Psychiatry*, *26*(8), 909–921. <https://doi.org/10.1007/s00787-017-0952-8>

Crone, E. A., & Dahl, R. E. (2012). Understanding adolescence as a period of socialAffective engagement and goal flexibility. *Nature Reviews Neuroscience*, *13*(9), 636–650. <https://doi.org/10.1038/nrn3313>

Gross, J. J. (2013). *Handbook of emotion regulation*. Guilford publications.

Gross, J. J., & John, O. P. (2003). Individual differences in two emotion regulation processes: Implications for affect, relationships, and well-being. *Journal of Personality and Social Psychology*, *85*(2), 348–362. <https://doi.org/10.1037/0022-3514.85.2.348>

Holodynski, M., & Friedlmeier, W. (2006). *Development of Emotions and Emotion Regulation*. Springer Science & Business Media.

Lee, F. S., Heimer, H., Giedd, J. N., Lein, E. S., Šestan, N., Weinberger, D. R., & Casey, B. J. (2014). Adolescent mental healthOpportunity and obligation. *Science*, *346*(6209), 547–549. <https://doi.org/10.1126/science.1260497>

Lemerise, E. A., & Arsenio, W. F. (2000). An Integrated Model of Emotion Processes and Cognition in Social Information Processing. *Child Development*, *71*(1), 107–118. <https://doi.org/10.1111/1467-8624.00124>

Littell, J. H. (2008). Evidence-based or biased? The quality of published reviews of evidence-based practices. *Children and Youth Services Review*, *30*(11), 1299–1317. <https://doi.org/10.1016/j.childyouth.2008.04.001>

Martin, F., & Johnson, M. (2015). More Efficient Topic Modelling Through a Noun Only Approach. In *Proceedings of Australasian Language Technology Association Workshop* (pp. 111–115).

Meeus, W. (2011). The Study of Adolescent Identity Formation 2000-2010: A Review of Longitudinal Research. *Journal of Research on Adolescence*, *21*(1), 75–94. <https://doi.org/10.1111/j.1532-7795.2010.00716.x>

Morris, A. S., Silk, J. S., Steinberg, L., Myers, S. S., & Robinson, L. R. (2007). The role of the family context in the development of emotion regulation. *Social Development*, *16*(2), 361–388. <https://doi.org/10.1111/j.1467-9507.2007.00389.x>

Myers, H. F. (2009). Ethnicity- and socio-economic status-related stresses in context: An integrative review and conceptual model. *Journal of Behavioral Medicine*, *32*(1), 9–19. <https://doi.org/10.1007/s10865-008-9181-4>

Ouzzani, M., Hammady, H., Fedorowicz, Z., & Elmagarmid, A. (2016). Rayyana web and mobile app for systematic reviews. *Systematic Reviews*, *5*(1), 210. <https://doi.org/10.1186/s13643-016-0384-4>

Page, L. (2006). *US7058628B1*.

Pierrehumbert, B., Bader, M., Miljkovitch, R., Mazet, P., Amar, M., & Halfon, O. (2002). Strategies of emotion regulation in adolescents and young adults with substance dependence or eating disorders. *Clinical Psychology & Psychotherapy*, *9*(6), 384–394. <https://doi.org/10.1002/cpp.339>

Pleck, J. (2004). Paternal involvement: Revised conceptualization and theoretical linkages with child outcomes. In M. E. Lamb (Ed.), *The role of the father in child development* (pp. 58–93). Hoboken, New Jersey: John Wiley & Sons.

Porges, S. W. (1995). Orienting in a defensive world: Mammalian modifications of our evolutionary heritage. A Polyvagal Theory. *Psychophysiology*, *32*(4), 301–318. <https://doi.org/10.1111/j.1469-8986.1995.tb01213.x>

Reindl, M., Gniewosz, B., & Reinders, H. (2016). Socialization of emotion regulation strategies through friends. *Journal of Adolescence*, *49*, 146–157. <https://doi.org/10.1016/j.adolescence.2016.03.008>

Riediger, M., & Klipker, K. (2014). Emotion regulation in adolescence. In *Handbook of emotion regulation* (pp. 187–202). Guilford Press.

Sameroff, A. (n.d.). A unified theory of development: A dialectic integration of nature and nurture. *Child Development*, *81*(1), 6–22. <https://doi.org/10.1111/j.1467-8624.2009.01378.x>

Sawyer, S. M., Azzopardi, P. S., Wickremarathne, D., & Patton, G. C. (2018). The age of adolescence. *The Lancet. Child & Adolescent Health*, *2*(3), 223–228. <https://doi.org/10.1016/S2352-4642(18)30022-1>

Schäfer, J. Ö., Naumann, E., Holmes, E. A., Tuschen-Caffier, B., & Samson, A. C. (2017). Emotion Regulation Strategies in Depressive and Anxiety Symptoms in Youth: A Meta-Analytic Review. *Journal of Youth and Adolescence*, *46*(2), 261–276. <https://doi.org/10.1007/s10964-016-0585-0>

Silk, J. S., Steinberg, L., & Morris, A. S. (2003). Adolescents’ emotion regulation in daily life: Links to depressive symptoms and problem behavior. *Child Development*, *74*(6), 1869–1880.

Sroufe, L. A. (1995). *Emotional Development: The Organization of Emotional Life in the Early Years*. Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9780511527661>

Staaks, J. (n.d.). Systematic Review Search Support. https://osf.io/49t8x/.

Steinberg, L. (2014). *Age of Opportunity: Lessons from the New Science of Adolescence*. Houghton Mifflin Harcourt.

Van de Schoot, R., De Bruin, J., Schram, R., Zahedi, P., De Boer, J., Weijdema, F., … Oberski, D. (2020). ASReview: Active learning for systematic reviews (Version v0.9.5). Zenodo.

Van Lissa, C. J., Hawk, S. T., de Wied, M., Koot, H. M., van Lier, P., & Meeus, W. (2014). The longitudinal interplay of affective and cognitive empathy within and between adolescents and mothers. *Developmental Psychology*, *50*(4), 1219–1225. <https://doi.org/10.1037/a0035050>

Van Lissa, C. J., Hawk, S. T., Koot, H. M., Branje, S., & Meeus, W. H. J. (2017). The cost of empathy: Parent-adolescent conflict predicts emotion dysregulation for highly empathic youth. *Developmental Psychology*, *ePub*(ePub), ePub. <https://doi.org/10.1037/dev0000361>

Van Lissa, C. J., & Keizer, R. (2020). Mothers’ and fathers’ quantitative and qualitative parenting in relation to children’s emotional adjustment: A between- and within-family investigation. *Developmental Psychology*, *in press*.

Van Lissa, C. J., Keizer, R., Van Lier, P. A. C., Meeus, W. H. J., & Branje, S. (2019). The role of fathers’ versus mothers’ parenting in emotion-regulation development from midLate adolescence: Disentangling between-family differences from within-family effects. *Developmental Psychology*, *55*(2), 377–389. <https://doi.org/10.1037/dev0000612>

Wang, L. L., Lo, K., Chandrasekhar, Y., Reas, R., Yang, J., Eide, D., … Kohlmeier, S. (2020). CORD-19: The Covid-19 Open Research Dataset. *ArXiv*.

Wijffels, J. (2019). Textrank: Summarize text by ranking sentences and finding keywords.

Zimmermann, P., & Iwanski, A. (2014). Emotion regulation from early adolescence to emerging adulthood and middle adulthood: Age differences, gender differences, and emotion-specific developmental variations. *International Journal of Behavioral Development*, *38*(2), 182–194.