Use of Claim Graphing and Argumentation Schemes in Biomedical Literature: A Manual Approach to Analysis

Eli Moser
Department of Health Sciences
McMaster University
Hamilton, Ontario, Canada
mosere@mcmaster.ca

Robert E. Mercer
Department of Computer Science
The University of Western Ontario
London, Ontario, Canada
mercer@csd.uwo.ca

Abstract

Argumentation in an experimental life science paper consists of a main claim being supported with reasoned argumentative steps based on the data garnered from the experiments that were carried out. In this paper we report on an investigation of the large scale argumentation structure found when examining five biochemistry journal publications. One outcome of this investigation of biochemistry articles suggests that argumentation schemes originally designed for genetic research articles may transfer to experimental biomedical literature in general. Our use of these argumentation schemes shows that claims depend not only on experimental data but also on other claims. The tendency for claims to use other claims as their supporting evidence in addition to the experimental data led to two novel models that have provided a better understanding of the large scale argumentation structure of a complete biochemistry paper. First, the claim graph displays the claims within a paper, their interactions, and their evidence. Second, another aspect of this argumentation network is further illustrated by the Model of Informational Hierarchy (MIH) which visualizes at a meta-level the flow of reasoning provided by the authors of the paper and also connects the main claim to the paper's title. Together, these models, which have been produced by a manual examination of the biochemistry articles, would be likely candidates for a computational method that analyzes the large scale argumentation structure.

1 Introduction

The large and ever-growing quantity of biomedical literature is well known (Hunter and Cohen, 2006). Included in biomedicine are the foundational experimental life sciences, such as genetics and biochemistry. Despite the importance and abundance of this literature, few computational models have been proposed that address the argumentation that is used to support the claims made in the papers describing outcomes of experiments. Such models would allow the mechanization of argumentation analysis which could enable scientific claim validation, a task made difficult because of the huge number of claims being made. A claim is any statement made within a paper which presents a novel finding based on the conducted experiment (Leonelli, 2015). A claim requires evidence to verify it. The structure of individual claims and their evidence is well illustrated by the Toulmin model (Toulmin, 2003), and the logic underlying these relationships can be categorized with argumentation schemes and premise classes (Karbach, 1987; Green, 2014a; Al Qassas et al., 2015; Green, 2015; Mayer et al., 2018).

The Toulmin model of argumentation is adept at illustrating the components of an individual argument. When composing argumentation text, every claim that is made must be supported by evidence and a warrant connecting the claim and evidence (Karbach, 1987). In the experimental sciences, the warrant is very often implicit, given that the intended audience can easily fill that slot in the argument structure. These arguments with implicit premises (and sometimes conclusions) are called enthymemes. Although each individual argument can be modelled in this way, it fails to recognize the variations in how claims relate to their supporting evidence, and the means by which the warrant supports that relation. To account for this, the Toulmin model can be supplemented with argumentation schemes. Argumentation

Proceedings of the 7th Workshop on Argument Mining, pages 88–99
Barcelona, Spain (Online), December 13, 2020.

schemes vary by context, and none have been synthesized for biomedical literature specifically. In this research a set of 15 novel argumentation schemes developed for categorizing genetic research argumentation (Green, 2015) were used and were found to be fully transferable beyond genetics to biochemistry arguments. While categorizing all of the claims of a paper, it became evident that often the evidence for a claim was another claim, and that the majority of claims within the paper were interconnected in this way. As a result, it is possible to construct a visual representation of all of a paper's claims in a graph. This graph illustrates which claims are supported by what data and also which claims are the most representative of a paper's findings overall. The varying degrees of claim significance demonstrated by this graph can be organized into a hierarchical model, which accounts for all data within a paper while still allowing for nuance to be maintained. In addition, the analysis of individual claims is not lost, as each progression of information from one level of specificity to the next is facilitated by argumentation schemes and the Toulmin model structure.

The next step to understanding the argumentation set forth in a scientific paper is analyzing the complete argumentation structure of the full paper. We present here our preliminary work on the analysis of this larger scale argumentation structure of biochemistry papers. This study is done by examining the flow of the paper from its data used as premises to support certain claims, some of which are used as premises for other claims, ending finally with the main claim of the paper. This research follows a similar path to Lawrence and Reed (2017), who manually constructed large scale graphs, and adds to the argumentation structure research by examining complete biochemistry articles and linking the argument schemes. Two models of analysis will demonstrate the interactions of claims. The first model is a high-level argument diagram of the data and claims structure capturing the essence of Green (2014b) and in the spirit of other high-level diagramming models (Stab et al., 2014; Kirschner et al., 2015; Eger et al., 2017; Stab and Gurevych, 2017) and diagramming-assisting software (Janier et al., 2014). The network of data and claims allows one to investigate its properties. In particular, the data is found in figures, tables, and other comments by the paper's authors. This data initiates the argumentation flow which ends in the main claim of the paper. This investigation leads to the second model, the Model of Informational Hierarchy (MIH). It makes more precise the units of the argument structure and their position in that structure. This hierarchical description of data and claims differs from other argumentation structures such as online debates (Lawrence and Reed, 2017) and student essays (Eger et al., 2017). In addition to the main focus, our research also demonstrates the applicability of Green's (2014a; 2015) argumentation schemes not only to genetics, but also to other biomedical research, biochemistry, in particular. Also, as an application of these models, the MIH can be viewed as a precursor to biomedical literature summarization.

The paper is organized as follows: Some related work to provide context for the current work is presented next. This is followed by a short description of the the five biochemistry papers that were used in the study. Then, the two main contributions, the claim graph and the Model of Informational Hierarchy, are explained. We conclude with a summary and some proposed future research directions.

2 Related Work

Our interest in investigating the larger scale argumentation structure has a similar motivation to the works of Wachsmuth et al. (2017) and Lawrence and Reed (2017), who are interested in investigating various properties of large scale argument networks. This new dimension adds to the previous works they point to as examples that consider particular aspects of argument structure: distinguishing argumentative and non-argumentative sentences (Moens et al., 2007), classifying text spans as premises or conclusions (Mochales Palau and Moens, 2009), classifying relations between specific sets of premises and their conclusion (Feng and Hirst, 2011), or classifying the different types of premise that can support a given conclusion (Park and Cardie, 2014). In addition to these examples, various manual and computational studies have been done to analyze different argumentation aspects, including: the structure of valid arguments in legal documents using feature-based machine learning (Mochales Palau and Moens, 2011), opinion and didatic texts using Rhetorical Structure Theory (Saint-Dizier, 2012), debates using textual entailment (Cabrio and Villata, 2012), and deconstructing the argumentation into the premises (referred

to as evidence in the evidence-based medical text genre) and conclusions (also referred to as claims) in scientific articles manually, with feature-based machine learning, or with neural end-to-end machine learning (Blake, 2010; Teufel, 2010; Green et al., 2011; Liakata et al., 2012; Sándor and de Waard, 2012; Longo et al., 2012; Longo and Hederman, 2013; Graves et al., 2014; Green, 2014a; Green, 2015; Kirschner et al., 2015; Mayer et al., 2018; Mayer et al., 2020). Lippi and Torroni (2015) provide an excellent survey of research done until 2015 which is further updated to 2018 by Stede and Schneider (2018). Corpus creation and analysis has also been another aspect of argumentation mining studies (Stab and Gurevych, 2017). More recently, neural net machine learning has provided a new machine learning paradigm for doing cross-domain claim analysis (Daxenberger et al., 2017). Eger et al. (2017) and Mayer et al. (2020) provide neural end-to-end models for computational argumentation mining. They label student essays and randomized control trials, respectively, with a BIO encoding to indicate argumentative and non-argumentative text spans, component type, and the stance between the components.

The rich history of work in argumentation mining has tended to focus on non-scientific text, however work in scientific text argumentation mining does have a following. The research done on various aspects of argumentation in scientific text begins with Argumentation Zoning (AZ) (Teufel et al., 1999; Teufel and Moens, 2002), also being some of the earliest work in argumentation mining. AZ is based on rhetorical moves, an important precursor for mapping out certain aspects of argument structure. Rhetorical moves, captured as AZ or more generally, have been investigated in a few science genres: computational linguistics (Teufel et al., 1999; Teufel and Moens, 2002), biochemistry (Kanoksilapatham, 2005), molecular biology (Mizuta et al., 2006), and chemistry (Teufel, 2010). Argumentation schemes (Walton et al., 2008) have been an important aspect of argumentation and argumentation mining. Green (2014a; 2015) has provided an important addition to these argumentation schemes for experimental scientific writing, specifically for genetics articles. While Green's argumentation schemes deal with aspects of the experiment, its outcomes, and the analysis of those outcomes, other work (Teufel, 2014) focusses on a different aspect of argumentation (via rhetorical moves): placing a research paper in its scientific context. Al Qassas et al. (2015) propose argumentation schemes for clinical discussions and use these schemes in an argument graph to analyze a discussion.

Argument diagramming is a technique that is commonly used to describe argumentation. While a manual operation, in the digital age, some computer-supported argument visualization tools have been developed. Araucaria (Reed and Rowe, 2004) and OVA+ (Janier et al., 2014) are two examples. The first provides support for mapping argumentation schemes. The second was developed for assisting with diagramming the larger scale argumentation structures. Lawrence and Reed (2017) provides an argument diagram for large scale online discussions. Eger et al. (2017) while mainly focussed on a neural end-to-end model for computational argumentation mining also provides an almost tree-like argumentation structure of complete student essays.

3 Dataset

Although there has been significant research on the deconstruction of individual claims and the methods of identifying claims within texts of various sorts (Mochales Palau and Moens, 2009; Blake, 2010; Feng and Hirst, 2011; Teufel, 2010; Green et al., 2011; Liakata et al., 2012; Sándor and de Waard, 2012; Longo et al., 2012; Longo and Hederman, 2013; Graves et al., 2014; Green, 2014a; Green, 2015; Kirschner et al., 2015; Mayer et al., 2018; Mayer et al., 2020), we are interested here in working with a new subset of scientific texts, biochemistry texts in particular, and linking claims into a larger argumentation structure. The dataset that we have used consists of five papers. Although the number of papers in the dataset used in this research is small, what sets it apart is that the annotation of the claims has been done by our domain expert, Dr. Derek McLachlin, one of the co-authors of the five papers. Our findings and the techniques and models presented are all based on the analysis of papers by our domain expert. The five papers all concern the dimerization interactions of the *b*-subunit of *Escherichia coli* ATP synthase. Having a co-author of the paper source the claims directly increased their reliability as representative of the papers' findings. Dr. McLachlin additionally provided detailed lists of claim interactions, indicating the direct source of evidence behind each claim.

Although the source material for the findings reported here is limited to one author, the literature used follows a structure ubiquitous to biomedical research, specifically, the progression of abstract, introduction, procedure/methods and materials, results, and discussion (Nair and Nair, 2014). This is standard for research of this nature and doesn't affect the applicability of the proposed models to other similar literature. Once a paper's claims and their sources have been determined it is possible to employ graphing techniques and information modeling.

4 Argumentation Structure

As stated earlier, the focus of this paper is the investigation of large scale argumentation structure in the spirit of Lawrence and Reed (2017) and Wachsmuth et al. (2017). The notion of large scale for our purposes is one complete biochemistry article. As discussed in Section 3, the claims and their spans have been provided by one of the co-authors of each paper. In addition, the interaction of the data and the claims has been provided by this co-author. Our research uses the Toulmin model (Toulmin, 2003). We have chosen the argumentation schemes provided by Green (2015) because they have an experimental science basis (cf. the clinical discussion schemes proposed by Al Qassas et al. (2015) and the randomized control trial evidence classes proposed by Mayer et al. (2018)). We refer to the model and scheme as the Toulmin-Green model. Green (2014a) has noted that many of the arguments in scientific writing are enthymemes, arguments that are missing premises or conclusions, and in most cases it is the warrants that are missing. What is missing from the information provided by our domain expert are the warrants. These have been provided by the first author of the current paper. An example of an argument:

> Premise (Claim 19) [grounds]: The highest level of disulfide formation was observed with the S60C + L65C and A61C + L65C combinations.
> Premise [missing warrant]: Proximity is necessary for disulfide binding between residues.
> Claim 21A (Green's Argumentation Scheme: Effect to Cause (5)): The result suggests that residue 65 of one subunit is close to residues 60 and 61 of the other

Given these elements, we are now able to analyze the argument structure of a biochemistry paper. We provide a categorization of the claims, a graphing technique that demonstrates the large scale argumentation structure, and an organization of the data and claims that we call the Informational Hierarchy.

4.1 Claim Categorization

In developing a claim based graphing technique the most fundamental component are the claims themselves. We were fortunate to have a co-author of the five biochemistry papers that we analyzed provide the claims and the reasons for making the claims. Having these claims and their support, they were individually categorized into three distinct groupings:

1. Figure-Claims (claims which were supported directly by experimental data)
2. Claim-Claims (claims supported by other claims, either claims based on figures or claims based on other claims)
3. Other (the majority of the claims in this category were experimental observations which could not fit into the displayed data)

In the five papers analysed over half of the claims fell into the "figure-claim" category, with a small minority falling into the "other" category.

The categorization step of claim analysis is important due to the informational distinction between claim-claims and figure-claims in the information hierarchy model, as well as their distinct visual treatment within the graph. Figure based claims are direct results of the data, while claim based claims encompass more findings and are generally more significant. See Figure 1 for an example of a claim graph for one of the analyzed papers. All claims used in the graph are represented by a number corresponding to the order in which they appear in the paper. Figure-claims and "other" claims are illustrated

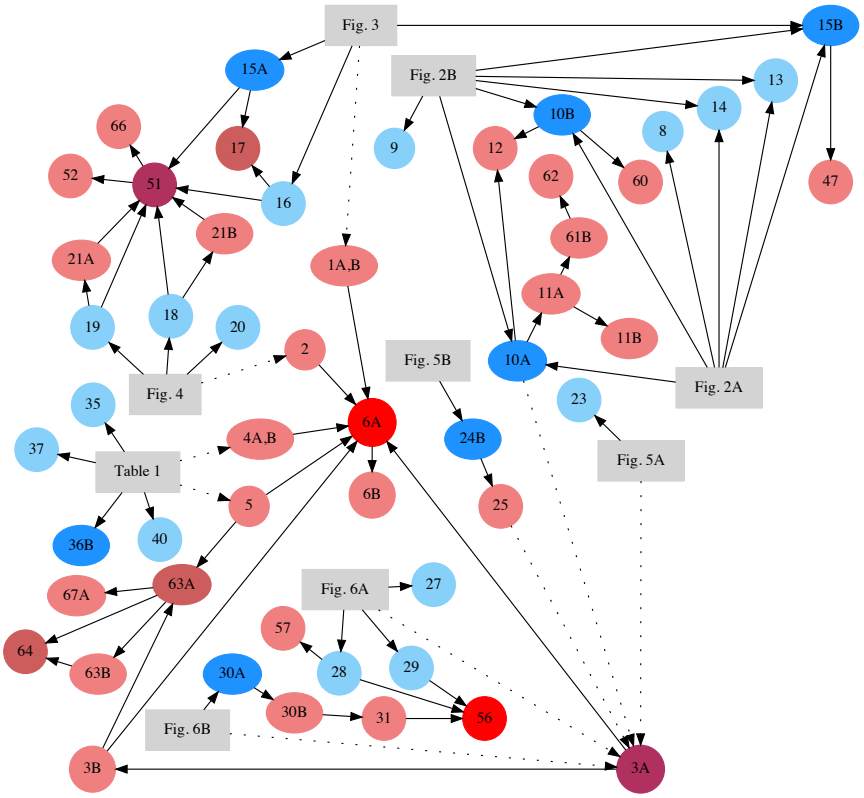

Figure 1: Claim graph for the paper "Dimerization Interactions of the *b* Subunit of the *Escherichia coli* F_1F_0-ATPase".

with a blue coloured circle labeled with their claim number. Claim-claims are illustrated with a red coloured circle labeled with their claim number. The intensity of the colour of a claim is dependent on the number of figures/claims which support it. As a result, the more intense the colour of the claim is, the more significant it is to the paper. This colour coding is done to make more central claims evident.

4.2 Claim Graphing

An example of a claim graph, created for the paper "Dimerization Interactions of the *b* Subunit of the *Escherichia coli* F_1F_0-ATPase" (McLachlin and Dunn, 1997) can be seen in Figure 1. The graph is composed of nodes representing the figures and tables of the paper, the grey rectangles labelled with their names in the paper; figure-claim nodes, the blue-toned nodes, which represent claims attributed to figures; and red-toned claim-claim nodes which represent claims that are supported by other claims and occasionally (but very rarely) supported by both another claim as well as a figure/table. Finally, colour intensity was adjusted to account for claim support, the more supporting figures/tables/claims a claim had, the darker the respective value. Directional edges indicate the premise-claim associations. The edges point from the support for a claim to the supported claim illustrating the directional flow of information. The dotted edges are meant to represent stylistic rather than argumentative moves. In the example given in Figure 1, nodes 1–5 are found in the abstract and are recapitulations of claims made in the discussion section of the paper (the nodes representing these discussion section claims are currently not shown in the graph). Nodes 6A and 6B assemble these claims ("Taken together the results are consistent with a model . . . "). The graph is presented with as few intersecting edges as is possible (none in the example given in Figure 1).

4.3 Discussion of the Graphing Technique

Once a graph has been constructed it is possible to understand a paper's components, their interactions, and their varying significance. The graph visualises the interactions between figures, claims, and most

importantly, which claims are the most illustrative of the paper's findings.

Claim significance is allocated based on the amount of support an individual claim receives, and not the amount of support it provides. This is because one of the future applications using the graph is to aid in the condensing or summarizing of a paper. As such, the concept of significance is dependent on how much information a claim encompasses. Figures and claims which have a noticeably high number of outgoing arrows are noteworthy on the grounds that they are important evidence. Their centrality is accounted for because figures and claims which provide a large amount of support are included in claims which receive a large amount of support. Significant claims represent a synthesis of smaller findings and data and provide a cumulative statement of the paper's results overall.

The density of support which classifies a claim as significant is relative to the total number of claims a paper makes. In applying this graphing technique universally the number of graphed claims will vary, and so will the density of support for individual claims. In a paper with twenty claims overall, the most densely supported claim may have only three supports while in a paper with sixty claims overall the main claim may be supported by eight. In general, a significant (secondary or main) claim will always have more than three supports and (in the papers that we analyzed) should be supported by 15% of all of a paper's claims. There is some variation in accordance with how many claims are not supported by results shown in the paper (the "other" category) however, it is usually unambiguous when looking at the graph which claims are the most supported.

Graphed claims can be separated into three categories according to the amount of support they have received, and the origin of that support. At this point in our research, this categorization is being subjectively determined by our sense of the claim's importance with respect to its role in summarizing the purpose of the paper. For this reason figure-based claims are never classified as main claims. They, like claim-based claims, have variation in the amount of support they receive, however since they are only statements about figures and not further analysis of figure statements, they do not have comparable summative value to claims based on claims based on figures. Regardless of how many figures a figure-claim has as direct supports, it is never a main claim, but will be accounted for as support for a claim-claim.

Within claim-claims there are additional levels of distinction, based on the number of support claims receive and the relevance of the claim content (i.e., what the claim is actually stating) to the paper itself. Relevance is also determined by how logically the claim content relates to the title of the paper, and therefore the thesis, of the paper. There are three levels of claim-claim relevance:

1. Tertiary Claim-Claims (1-2 supporting claims)
2. Secondary Claim-Claims (3 or more supporting claims, no direct relevance to the title/main concept)
3. Main Claims (3 or more supporting claims, direct relevance to the paper title/main concept)

When analysing the graph of a paper and the flow of information within it, there emerges a very clear progression of supports and detail. The importance of a claim is correlated to the amount of support it has, as is its summative ability. This makes sense, as the more information represented within a claim the more summative a claim will be. Once a main claim has been identified, it is possible to utilise it (in conjunction with the title) to summarize the entirety of the paper into a one or two sentence statement. Although convenient to distil information to that level, it is crucial to maintain the levels of complexity which precede it so that nuance can be sought at the observer's discretion. Luckily, this is preserved in the graph, as it is possible to trace the origins of the main claims content through the support arrows it is connected to, and the supports of those supports in turn. This progression from figure/table to main claim illustrated by the graphing process can itself be used as a means of understanding the inductive flow of information in research from the specifics of the data to the overall main claim, accounting for every component of the paper. The following is an example of this progression:

> Premise (Figures 2A and 2B observations): The figures show results of SDS-PAGE (protein gel electrophoresis) and Western blotting. Band intensities show molecular masses of proteins with A128C, R138C, S139C, and S146C mutations. A128C and S139C have higher masses.
> Premise [missing warrant]: band intensity indicates protein amount at a specific mass

Figure-Claims 10A and 10B (Green's Argumentation Scheme: Consistent with Predicted Effects (7)): The proteins containing the A128C and S139C mutations showed a strong tendency to dimerize, (Claim 10A) while cysteines at positions 138 and 146 did not tend to form disulphides under these conditions. (Claim 10B) [Claim 10A's slightly reworded restatement in the Discussion section is indicated by the dashed edge in the claim graph]

Premise (Figure 5B observation): The figures show results of SDS-PAGE (protein gel electrophoresis) and Western blotting. The band intensities indicate the molecular masses of the proteins containing the mutations at locations 124-132 and 138. The proteins with mutations at locations 124, 128, and 132 do not show dark bands at lower molecular masses.
Premise [missing warrant]: band intensity indicates protein amount at a specific mass
Figure-Claim 24B (Green's Argumentation Scheme: Effect to Cause (5)): . . . , the most complete disulfide bond formation was observed at positions 124, 128, and 132

Premise: Claim 24B
Premise [missing warrant]: Disulfide binding follows 4-residue periodicity in an α-helical protein structure.
Claim 25 (Green's Argumentation Scheme: Effect to Cause (5)): The 4-residue periodicity of cross-linking . . . suggests a parallel α-helical arrangement in this region. [Claim 25's reworded restatement in the Discussion section is indicated by the dashed edge in the claim graph]

Premise: the restated Claim 10A (indicated by the dashed edge)
Premise: the restated Claim 25 (indicated by the dashed edge)
Premise [missing warrant]: Proximity is necessary for disulfide binding
Premise [missing warrant]: Residues 128 and 128' and 139 and 139' are close together within the quaternary structure
[Information from Figs. 5A, 6A, and 6B (dashed edges) has not been included in this example.]
Claim 3A (Green's Argumentation Scheme: Effect to Cause (5)): Cysteines at positions 124, 128, 132, and 139 showed strong tendencies to form disulfides with their mates in the dimer, suggesting a parallel α-helical interaction between the subunits in this region.

4.4 The Model of Informational Hierarchy

The Model of Informational Hierarchy (MIH), shown in Figure 2, represents the development of information in experimental science literature by tracing the path of information through the graphing technique described above. The MIH is pyramidal in shape, in order to illustrate the inductive nature of the progression between levels. As the pyramid is traversed from the base to the top the specificity of the information decreases, and the summative ability of the information increases. The top levels of the MIH are built on those at the base, and thus account for the information contained within them. Unlike the graph, this model is able to account for levels of information even more specific than the data presented in figures and tables. Allocating data as the base of claim graphing is understandable as it is the results, and not the way that they were acquired, that are the main focus of claims made in scientific literature. In analysing the entirety of information in any given paper however, the raw data itself cannot be the base of the model. How the data was gathered informs the data, and even how the data can be interpreted. It is important to understand each level of the MIH and how it interacts with the levels surrounding it. The layers will be discussed in ascending order, beginning with the foundation.

Methods and Experimental Procedure Underlying the data of any scientific claim is the means by which that data was procured. Without an understanding of the methods used and the experimental procedure the veracity of the data presented is unknowable. The procedure informs how the data relates to the figure-claims. To know what a figure is showing, how it is showing it, and how that relates to figure-claims is all based on an understanding of the experimental methods and materials used, making the procedure the foundational level of the MIH.

Data (Figures and Tables) Following from the methods and experimental procedures of a biomedical paper comes the presentation of the experimental findings. The presentation of the data found in an experiment is as crucial as the means by which it was discovered, as it is the findings which the paper is

written to communicate. The data (displayed in figures or tables) directly support figure-claims. There is no argumentation scheme which relates the experimental procedures/methods to the data of a paper, as the relationship between the two is evidentiary rather than based on a logical progression. However, a Toulmin-Green argumentation scheme relationship does exist between the data presented and figure-claims made. In these schemes the premise is derived from interpretation of the data informed by the methods/experimental procedure, resulting in the conclusion that is the figure-claim.

Figure-Claims and Experimental Observations The category of figure-claims represents claims based on experimental data presented directly by either a figure or a table. Claims which fall under the category of "Experimental Observations" are statements of outcomes noticed by the researchers that could not be easily included in the tables and figures used, but are still based directly on the findings.

Tertiary Claim-Claims Tertiary claim claims, as discussed in the discussion of the graphing technique, are claim-claims which are not deemed to be significant. They are based on figure-claims and experimental observations, but lack summative power (only 1-2 supports). They are an intermediate step between figure-claims and secondary claims.

Secondary Claim-Claims Secondary claims have enough support to be significant, however lack relevancy. They are not main claims due to their lack of direct relation to the paper's title (and thesis).

Main Claim The main claim is the most representative statement of the paper's findings, and the highest level of summarization. These claims are claim-claims with the greatest amount of support and will often be supported by other highly supported claims. In conjunction with the title, a main claim can be used to summarize the entirety of a paper's findings in a single sentence.

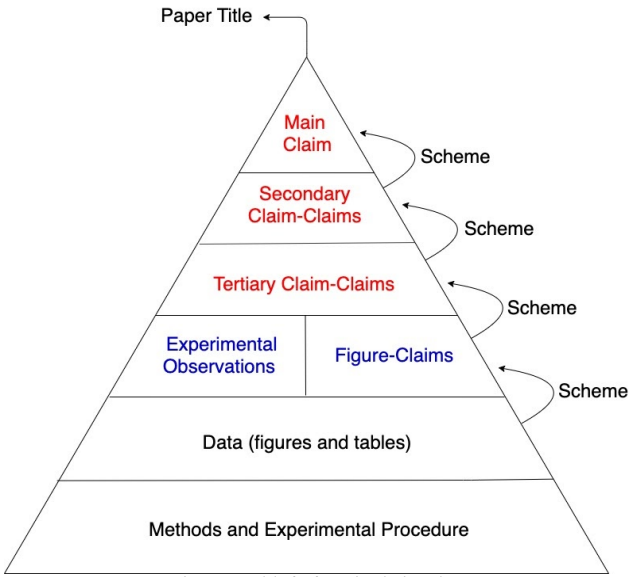

Figure 2: Model of Informational Hierarchy

4.5 Use of Argumentation Schemes

As seen in Figure 2, there are argumentation schemes included between each level of the information hierarchy. These schemes are from the Toulmin-Green model of claim argumentation and are what underlies all informational progression within the paper. In the claim graph as well, each arrow represents a logical progression which can be categorized with the Toulmin-Green model. For figure-claims the evidence comes from the data itself, the warrant is assumed knowledge or taken from the procedure. For claim-claims (at all levels of significance) the evidence comes from a previously made claim, and the warrant is often assumed knowledge. Although Dr. Green's argumentation schemes do not affect the progression of the information hierarchy, they are the connections which hold the graph and the information model together. Without them there would be no accountability for how statement significance progresses, and no legitimacy to the logic of those progressions.

5 Conclusions and Future Work

The purpose of analyzing the five papers with the Toulmin-Green model and subsequently generating the claim graph and the Model of Informational Hierarchy has been to illustrate a manual approach for generating a large scale argumentation structure for a complete biochemistry paper. The methods presented are novel for this science genre. One can compare this argumentation structure with those that have been established for debates (Lawrence and Reed, 2017), persuasive essays (Stab and Gurevych, 2017), and student essays (Eger et al., 2017). As is shown with the graph, claims in biochemistry articles are intrinsically interlinked, with claims using both other claims and data for their evidence. The MIH model lists argumentation schemes as intermediary steps between levels of information to illustrate the importance of claims relating to each other in a way that is categorizable and logically sound.

Beyond these results, there is potential for further research. The Toulmin-Green model, the claim graphing technique, and the MIH all have aspects which can be further developed.

Green's (Green, 2016; Green, 2018) innovative idea to use Prolog rules to generate argument schemes can be improved. As proposed, the body of each rule matches with the Prolog knowledge base extracted from the text. The body of each rule needs some discourse information to constrain what elements from these Prolog facts are allowed to be combined. We can use the information derived from the claim graph to add to these rule bodies, i.e., only those facts that are extracted from elements that are connected by an edge would be allowed to imply the argument scheme (the head of the rule). However, it still remains to be seen how to do this in an automatic way since the graphs have been constructed manually. Using discourse and rhetorical moves may be some directions to investigate. In addition to this, a method to produce implicit argument components for enthymemes is needed.

Biochemistry (and other experimental life sciences) papers are written following the IMRaD (Introduction, Methods, Results, and Discussion) structure. As indicated by dashed edges in the claim graph in Figure 1 and mentioned in the example in Section 4.3, the placement of a claim in this structure may have stylistic significance. This aspect of the writing style was not part of this initial study and needs to be further investigated and incorporated in the claim graph. Indeed, that a claim is restated and where this restatement is placed may provide further justification for the split into the three claim categories proposed in the MIH. As well, a characteristic of writing in the biochemistry genre is to explain the reasons for choosing the experiments discussed in the paper. This aspect is important to a full understanding of the argumentation structure. How to incorporate it in the claim graph needs to be investigated. Discussion of claims in the paper highlighted above commented on claims in previously published papers. This important inter-paper argumentation structure will be investigated.

An important next step is to produce the claim graph automatically. Methods to identify claims in biomedical text (albeit not always directly applicable to this type of experimental biochemistry article) ranging from rule-based (Blake, 2010) to neural end-to-end (Mayer et al., 2020) have been previously investigated and would comprise a first step toward this goal.

A future application of the Model of Informational Hierarchy could be summarization of the paper, a noted motivation for some of the earliest argumentation research (Teufel and Moens, 2002). Moving from the top of the hierarchy (the paper's title and main claim) downward would provide more and more detail which is not contained in an abstract. Real-time summarization at a user-specified level of detail seems possible. And a summarization focussed on a particular aspect of a paper's research claims combining the MIH and the claim graph could also be a possibility.

Acknowledgements

This research was funded by The Natural Sciences and Engineering Research Council of Canada (NSERC) through an Undergraduate Summer Research Award to Eli Moser and a Discovery Grant to Robert E. Mercer. We thank Dr. Derek McLachlin of the Department of Biochemistry, The University of Western Ontario for providing his analysis of the five papers used in this study. We also acknowledge the helpful comments provided by the reviewers.

References

Malik Al Qassas, Daniela Fogli, Massimiliano Giacomin, and Giovanni Guida. 2015. Analysis of clinical discussions based on argumentation schemes. *Proceedia Computer Science*, 64:282–289.

Catherine Blake. 2010. Beyond genes, proteins, and abstracts: Identifying scientific claims from full-text biomedical articles. *Journal of Biomedical Informatics*, 43:173–189.

Elena Cabrio and Serena Villata. 2012. Natural language arguments: A combined approach. In *Proceedings of the 20th European Conference on Artificial Intelligence (ECAI 2012)*, pages 205–210.

Johannes Daxenberger, Steffen Eger, Ivan Habernal, Christian Stab, and Iryna Gurevych. 2017. What is the essence of a claim? Cross-domain claim identification. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 2055–2066.

Steffen Eger, Johannes Daxenberger, and Iryna Gurevych. 2017. Neural end-to-end learning for computational argumentation mining. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 11–22.

Vanessa Wei Feng and Graeme Hirst. 2011. Classifying arguments by scheme. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, pages 987–996.

Heather Graves, Roger Graves, Robert E. Mercer, and Mahzereen Akter. 2014. Titles that announce argumentative claims in biomedical research articles. In *Proceedings of the First Workshop on Argumentation Mining*, pages 98–99.

Nancy Green, Rachael Dwight, Kanyamas Navoraphan, and Brian Stadler. 2011. Natural language generation of biomedical argumentation for lay audiences. *Argument and Computation*, 2(1):23–50.

Nancy Green. 2014a. Towards creation of a corpus for argumentation mining the biomedical genetics research literature. In *Proceedings of the First Workshop on Argumentation Mining*, pages 11–18.

Nancy L. Green. 2014b. Argumentation for scientific claims in a biomedical research article. In *Proceedings of the Workshop on Frontiers and Connections between Argumentation Theory and Natural Language Processing (ArgNLP2014)*, volume 1341 of *CEUR Workshop Proceedings*.

Nancy Green. 2015. Identifying argumentation schemes in genetics research articles. In *Proceedings of the 2nd Workshop on Argumentation Mining*, pages 12–21.

Nancy L. Green. 2016. Implementing argumentation schemes as logic programs. In *Proceedings of the 16th Workshop on Computational Models of Natural Argument*, volume 1876 of *CEUR Workshop Proceedings*, pages 1–7.

Nancy L. Green. 2018. Towards mining scientific discourse using argumentation schemes. *Argument & Computation*, 9:121–135.

Lawrence Hunter and K. Bretonnel Cohen. 2006. Biomedical language processing: What's beyond PubMed? *Molecular Cell*, 21(5):589–594.

Mathilde Janier, John Lawrence, and Chris Reed. 2014. Ova+: an argument analysis interface. In *Proceedings of COMMA 2014*, pages 463–464.

Budsaba Kanoksilapatham. 2005. Rhetorical structure of biochemistry research articles. *English for Specific Purposes*, 24(3):269–292.

Joan Karbach. 1987. Using Toulmin's model of argumentation. *Journal of Teaching Writing*, 6(1):81–92.

Christian Kirschner, Judith Eckle-Kohler, and Iryna Gurevych. 2015. Linking the thoughts: Analysis of argumentation structures in scientific publications. In *Proceedings of the 2nd Workshop on Argumentation Mining*, pages 1–11.

John Lawrence and Chris Reed. 2017. Using complex argumentative interactions to reconstruct the argumentative structure of large-scale debates. In *Proceedings of the 4th Workshop on Argument Mining*, pages 108–117.

Sabina Leonelli. 2015. What counts as scientific data? A relational framework. *Philosophy of Science*, 82(5):810–821.

Maria Liakata, Shyamasree Saha, Simon Dobnik, Colin Batchelor, and Dietrich Rebholz-Schuhmann. 2012. Automatic recognition of conceptualization zones in scientific articles and two life science applications. *Bioinformatics*, 28(7):991–1000.

Marco Lippi and Paolo Torroni. 2015. Context-independent claim detection for argument mining. In *Proceedings of the Twenty-Fourth International Joint Conference on Artificial Intelligence (IJCAI 2015)*, pages 185–191.

Luca Longo and Lucy Hederman. 2013. Argumentation theory for decision support in health-care: A comparison with machine learning. In *Proceedings of the International Conference on Brain and Health Informatics*, pages 168–180.

Luca Longo, Bridget Kane, and Lucy Hederman. 2012. Argumentation theory in health-care. In *Proceedings of the 25th IEEE Symposium on Computer-Based Medical Systems*.

Tobias Mayer, Elena Cabrio, and Serena Villata. 2018. Evidence type classification in randomized controlled trials. In *Proceedings of the 5th Workshop on Argument Mining*, pages 29–34.

Tobias Mayer, Elena Cabrio, and Serena Villata. 2020. Transformer-based argument mining for healthcare applications. In *Proceedings of the 24th European Conference on Artificial Intelligence (ECAI 2020)*, pages 2108–2115.

Derek T. McLachlin and Stanley D. Dunn. 1997. Dimerization interactions of the *b* subunit of the *Escherichia coli* F_1F_0-ATPase. *Journal of Biological Chemistry*, 272(34):21233–21239.

Yoko Mizuta, Anna Korhonen, Tony Mullen, and Nigel Collier. 2006. Zone analysis in biology articles as a basis for information extraction. *International Journal of Medical Informatics*, 75(6):468–487.

Raquel Mochales Palau and Marie-Francine Moens. 2009. Argumentation mining: The detection, classification and structure of arguments in text. In *Proceedings of the 12th International Conference on Artificial Intelligence and Law*, page 98–107.

Raquel Mochales Palau and Marie-Francine Moens. 2011. Argumentation mining. *Artificial Intelligence and Law*, 19(1):1–22.

Marie-Francine Moens, Erik Boiy, Raquel Mochales Palau, and Chris Reed. 2007. Automatic detection of arguments in legal texts. In *Proceedings of the 11th International Conference on Artificial Intelligence and Law*, page 225–230.

P. K. Ramachandran Nair and Vimala D. Nair, 2014. *Organization of a Research Paper: The IMRAD Format*, chapter 2, pages 13–25. Springer.

Joonsuk Park and Claire Cardie. 2014. Identifying appropriate support for propositions in online user comments. In *Proceedings of the First Workshop on Argumentation Mining*, pages 29–38.

Chris Reed and Glenn Rowe. 2004. Araucaria: Software for argument analysis, diagramming and representation. *International Journal on Artificial Intelligence Tools*, 13(4):961–979.

Patrick Saint-Dizier. 2012. Processing natural language arguments with the <textcoop> platform. *Argument and Computation*, 3(1):49–82.

Ágnes Sándor and Anita de Waard. 2012. Identifying claimed knowledge updates in biomedical research articles. In *Proceedings of the Workshop on Detecting Structure in Scholarly Discourse*, pages 10–17.

Christian Stab and Iryna Gurevych. 2017. Parsing argumentation structures in persuasive essays. *Computational Linguistics*, 43(3):619–659.

Christian Stab, Christian Kirschner, Judith Eckle-Kohler, and Iryna Gurevych. 2014. Argumentation mining in persuasive essays and scientific articles from the discourse structure perspective. In *Proceedings of the Workshop on Frontiers and Connections between Argumentation Theory and Natural Language Processing (ArgNLP2014)*, volume 1341 of *CEUR Workshop Proceedings*.

Manfred Stede and Jodi Schneider. 2018. *Argumentation Mining*. Synthesis Lectures on Human Language Technologies. Morgan & Claypool Publishers.

Simone Teufel and Marc Moens. 2002. Summarizing scientific articles: Experiments with relevance and rhetorical status. *Computational Linguistics*, 28(4):409–445.

Simone Teufel, Jean Carletta, and Marc Moens. 1999. An annotation scheme for discourse-level argumentation in research articles. In *Proceedings of the Ninth Conference of the European Chapter of the Association for Computational Linguistics*, pages 110–117.

Simone Teufel. 2010. *The Structure of Scientific Articles: Applications to Citation Indexing and Summarization*. CSLI Studies in Computational Linguistics. Center for the Study of Language and Information.

Simone Teufel. 2014. Scientific argumentation detection as limited-domain intention recognition. In *Proceedings of the Workshop on Frontiers and Connections between Argumentation Theory and Natural Language Processing (ArgNLP2014)*, volume 1341 of *CEUR Workshop Proceedings*.

Stephen E. Toulmin. 2003. *The Uses of Argument*. Cambridge University Press.

Henning Wachsmuth, Benno Stein, and Yamen Ajjour. 2017. "PageRank" for argument relevance. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 1, Long Papers*, pages 1117–1127.

Douglas Walton, Chris Reed, and Fabrizio Macagno. 2008. *Argumentation Schemes*. Cambridge University Press.