Academic Publishing as a Social Media Paradigm

Michael E. Payne and Linh B. Ngo and Amy W. Apon
School of Computing, Clemson University
Clemson, South Carolina
{mpayne3, lngo, aapon}@clemson.edu

Abstract—This work seeks to bridge areas of academic institutional research, social network analysis, and content analysis through the application of a social media paradigm to the academic research publishing environment. The concept is built upon the analysis of similarities and differences that exist in the structural and functional building blocks of academic publishing and social media. The potential impact in the form of new research directions that arise from this application are presented.

Keywords—social media; social networking; academic publications; higher education; big data;

I. INTRODUCTION

Social network analysis, dating as far back as the 1960s, has been a key instrument for studying the causes and effects of relationships between objects (people, groups, organizations) within a system [1]. While social networks focus on the links between the objects, advances in computing technologies have led to the rise of social media, which enable fast and highly scalable formations of these links as well as their user-generated content. A recent study shows that the services from social networking sites have been utilized to help understand and improve the performance of higher education institutions [2]. In a direct approach, social networking is used as a means for institutions to quickly communicate and disseminate knowledge and information to students as well as to strengthen the connections among an institution and its prospective students, current students, faculty, and alumni. In a more indirect approach, researchers also study social network activities to gain insight into community engagement and academic performance of students [3]. Network analysis, an important component of social networking analysis, has been used extensively to study the benefits of research relationships and collaborations between faculty [4, 5]. These studies typically concentrate on the social networking links at the individual level with less emphasis on the user-generated content that forms these links. As a result, these studies may not address concerns about science policies at a higher abstraction level by institutions and funding agencies.

This paper sets the foundation for the application of a social media paradigm to the study of the academic research publication environment in which the authors are viewed as the equivalent of users in social networking sites. The goal is to create academic authors’ scholarly profiles that are represented by the aggregated contents of all of their research publications in all publication venues over time. This is similar to the manner in which social network users build their social media profiles through their tweets and posts. We show that the academic research publication environment contains or has the potential to provide the same elements as those of social network and media sites as defined in [6, 7]. This will result in a comprehensive representation that:

- enables the investigation of social networks among all the entities within the academic research publication environment and the nature of causes, effects, contents, and timing components of these networks,
- illustrates the connections between entities at different organizational levels within the environment, an area that has not previously been well studied,
- discusses the similarities and differences between the characterizing elements of the traditional academic publishing environment and the modern social media environment and the potential of the differences, and
- identifies the advantages of having additional breadth statistics for institutions at this new abstraction level and adjusted time-scale to support longitudinal studies and the corroboration of external events.

The remainder of this paper is as follows: Section II presents the similarities and differences between social media and academic publishing, which justify the application of the social media paradigm to the ecosystem of academic research publication. We discuss the implications of this application toward new research directions that stem from this new model in Section III. Section IV discusses future work.

II. ACADEMIC PUBLISHING ENVIRONMENT AS A SOCIAL MEDIA ENVIRONMENT

Similarities and differences between the social media environment and the academic publishing environment can be described using structural, functional, and content comparisons.

A. Structural Comparison

Social media has been referred to as a “conversational, distributed mode of content generation, dissemination, and communication among communities” [8]. This definition can also be applied to the field of academic research. Figure 1 illustrates the structural similarities of the two fields.

1) Environment: The most significant difference between a social media-based and previous network-based analysis of academic publishing is the concept of an environment. In social media, companies like Facebook Inc. and Twitter Inc. create, develop, maintain, and promote the environment through which media content are posted and shared. They possess the capital and the infrastructure to promote social media and its uses. Similarly, organizations such as Springer Publishing Company or the IEEE Society are the creators...
The scholarly profile of an author is built from his past co-publication venue which belongs to a particular publisher. In a similar manner to social media, an article is published to that are published in journals and conference proceedings. In this way, they generate content in the form of research articles through their historical content and connections. Authors are form’s respective environment. Users gradually build profiles published through their respective platform within that plat- form user-generated topics. Content on social media can be categorized into different groups according to their topics. In addition, articles are typically accompanied by a set of keywords that are representative of the topics covered by the articles.

4) Catalysts for Content: Catalysts for content are the organized or unorganized “outside” structures that encourage and stimulate content. In social media, they can be communities, groups, or facets that exist online or offline. In the research publication sphere, authors’ works are, in part, influenced by strategic decisions from their academic departments and institutions. Research areas, on the other hand, are catalysts that expand to influence authors beyond the institutional boundaries.

5) Recognition of Content and Reputation: The social media and academic publishing environments are most similar in their need for quality content to be recognized. For example, on Facebook, content can be recognized with a “like”. A relevant tweet is recognized through a “retweet” on Twitter, which is the reposting of the original tweet by another user that includes the alias of the user who generated the tweet. High numbers of likes or retweets are indicators of popularity of content. Over time, the consistent generation of highly popular content contributes to users’ reputations in social media. These forms of recognition can be seen as equivalent to citations in academic publishing. Thus, a higher number of citations indicates a higher level of recognition for the content. Consequently, an author whose articles are consistently cited multiple times will acquire a high level of reputation and respect from fellow authors.

6) Categorization of Content: Content within social media can be categorized by topics or hashtags, which are a group of alphanumeric characters, preceded by the “#” character, to form user-generated topics. Content on social media can be searched by topics or hashtags. Likewise, articles in academic publishing can be categorized into different groups according to their topics. In addition, articles are typically accompanied by a set of keywords that are representative of the topics covered by the articles.

B. Functional Comparison

We use seven functional building blocks (identity, conversations, sharing, presence, relationships, reputation, and groups [7]) as a means to compare the functional components within academic research to a social media paradigm.

1) Identity: The identity of an author is constructed by aggregating characteristics such as the author’s name, email, address, and affiliations. The ability to identify an author in academic publishing is much higher than that of social media where users can distribute content under one or multiple aliases with little or no personal information required.

2) Conversations: Active fields of research in academic publishing equate to conversations in social media. When a topic is actively being researched, forms of positive, negative, and neutral citation of articles within the topic exist. The popularity of a conversation rises and falls following the number of participating users and activities. The activeness of a research field is illustrated through the number of researchers, published articles, and citations. Starting or joining a conversation in an unpopular or no longer popular topic is similar to publishing articles in an uninteresting or no longer active research field - it results in little or no likes, retweets, or citations.
3) Sharing: The sharing nature of social media is characterized by two questions: “What is being shared” and “How much of it can be shared.” The creators of social media content may retain the rights to own and share their content, even through the contents are published and distributed by social networking sites. On the other hand, in academic publishing, even though research results and scientific knowledge are being shared and owned by the authors through their research articles, the publishers often retain the rights to the articles. In recent years, publishers have permitted authors to provide public online access to early drafts of the articles or even the final versions of the articles.

4) Presence: Activity (publishing articles) in the general space of academia and within a particular field speaks to an author’s presence in the environment. Also taken into account are the platforms (journals and conferences) that authors use to publish their work. Authors who post their work to their personal website or publish to a non-indexed journal/conference will not generally have the same presence as another author who publishes in an indexed platform.

5) Relationships: Co-authorship remains one of the best ways to measure an author’s relationship. Following co-authorship, author’s citation relationship between other authors in the field can be used to measure relationships between authors. Colleagues within institutions and departments can have relationships, as well as being co-investigators for grants.

6) Reputation: The reputation of authors is determined by their level of participation and leadership within conversations (active research fields). This involves an author being knowledgeable of content produced by peers, and peers being knowledgeable of the content produced by that author. Attributes such as the number (or average number) of cited articles within an article, times cited and count of actively cited publications can be used to determine one’s reputation.

7) Groups: In academic publishing, a possible equivalent concept to social media’s group would be a conference series. A conference could allow any research articles (open), require refereed articles (closed - approval required), or only invite certain articles to be published. Authors who submit to the same conference are likely to share the same research interests.

C. Content Comparison

There are some distinct differences in content and culture of the two environments.

- Academic articles are orders of magnitude longer than social media posts. Also, sentence structures, spelling, and grammar in articles are more formal than those of social media.
- Content in academic publishing is mainly text based. In social media, there are platforms for publishing pictures and videos.
- The average time between publications for an active author can range from months to years as opposed to social media where the gap can literally be seconds. This also leads to a significantly higher number of published items in social media.

- Articles are subject to rigorous reviewing to determine whether they are accepted for publication. This self-filtering of content assures some reliability and accountability from the authors. Review in social media occurs post publishing of content and typically content is only removed when it is offensive or infringes on intellectual property. One of the active research areas in social media is the separation of high and low quality content.

Though there are some differences in the content of the social media and academic publishing, a social media paradigm can be applied to academic publishing while allowing it to preserve its context. This has the potential to allow researchers to ask new questions and draw new conclusions from a more comprehensive and unified perspective.

III. Opportunities and Challenges

Previous work in higher education institutional research identified influences among academic publications and other entities such as institutions, fields of research, and authors [9]. What has not been evaluated is the comprehensive relationship among all entities of the research and academic publishing ecosystem of higher education institutions. The application of a social media paradigm will create a new infrastructure for content in academic publishing. This infrastructure promotes new research approaches in higher education institutional research through the application of social media research techniques such as sentiment analysis and content-based trend analysis and predictions. On the other hand, the academic publishing environment augments its social media paradigm with data that spans decades. Additionally, data can be included from other entities and interactions in the higher education ecosystem such as research funding, research expenditures, student enrollment, and research equipment and infrastructure. This comprehensive data set, in combination with the dynamic and robust structure of a social media paradigm, is a valuable resource from which answers to new research questions become feasible.

An example topic of interest is the study of the impact of research investment, for example, investment in high performance computing resources (HPC), to research productivity in higher education institutions. From this topic, we can identify a set of related questions that has not been addressed previously. The first question is, “What research trends and topics motivate the use of HPC?” A related question is, “If we observe the development in research topics and collaborations from researchers at an institution that does not have HPC resources, can we predict whether it would be beneficial for the institution to acquire HPC?” Another question is, “Does an institution’s HPC investment lead to an influx of leading researchers to the institution?” At a broader perspective, if a leading researcher is publishing and contributing to a reputable journal, if the acquisition of HPC changes the nature of research and the publication venue of this researcher, how does that impact the reputation of the previous and new journals? These types of questions require interdisciplinary and intercommunal investigations in institutional research, social network, and social media content. They cannot be answered by just looking at the social network, the content, and quantitative institutional and publications metrics such as bibliometrics [10], scientometrics [11], webometrics [12], or altmetrics [13], independently of...
each other. We need to extract the corroborating content from the publication articles that illustrate changes in the nature of research, to look at the network structure between the researchers, and to identify changes in network structures that correspond to changes in content. The aggregation and quantification of these changes, the essence of the social media paradigm, can be combined with the other metrics as inputs to statistical methodologies in order to find the answers.

The lengthy, text-rich nature of content generated from higher education research has the potential to provide a vast repository for researchers in topic modeling, semantics, sentiment analysis, text categorization, and information retrieval. Econometrics and hypothesis testing can be applied to the validation of potential causal observations. Similar to its online counterpart, examination of growth and nature of the academic research environment must be done in context, taking an in-depth look at institutions/researchers finances and expenditures. In other words, a social media paradigm provides a domain for researchers across various fields to work in tandem to understand the complex environment of research and academic publishing in higher education, in which they are already participants.

A challenge of doing research is the collection, curation, and aggregation of data. There exist many repositories of data on publishers, institutions, authors, and full-text articles from sources such as Thompson Reuters, the National Science Foundation, and the National Center for Educational Statistics. This leads to the issue of complexity in data format and size of data. As mentioned in [14], publication data for just computer and information science can be as high as 1.7 terabytes of data. Every year, this is expected to grow at the rate of roughly 14% [15]. It is reasonable to believe that the size of publications for the other scientific fields are on a similar scale. As publication data are aggregated with institutional data to form the inputs to the social media and institutional analysis, the underlying hardware and software frameworks have to be able to support big data storage and big data analytics. Potential contributions lie in the development of data overlay and presentation tools that can be used for concurrent statistical and visual analysis. There is a need for standardization for data representation and labeling.

IV. Conclusion and Future Work

This paper describes the foundation for the application of a social media paradigm to the study of the academic research publishing environment in higher education institutions. The structural and functional similarities between academic publishing and social media promote interdisciplinary and data-intensive approaches to new research topics in higher education institutional research. Our future work builds on this paradigm. Future work includes study of the scholarly profile of the authors of academic publications and observation of changes in research direction and co-authorship. A statistical analysis on the aggregation of these observations will quantify the relationship between these changes. This discussion has been focused on the area of academic publications and higher education. The social media paradigm can be applied to other aspects of higher education, for example, the impact of investment in research instruments and the impact of grants and contracts from federal research agencies. Future research will examine NSF awards data from a social media perspective, including reports of publications produced by the awards.

Acknowledgments

We gratefully acknowledge the research support from National Science Foundation grants NSF INSPIRE #123436 and NSF #1212680.

References