A Meta-Analysis of Theories and Topics in Social Media Research

Wietske Van Osch  
Michigan State University  
Vanosch@msu.edu

Constantinos K. Coursaris  
Michigan State University  
Coursari@msu.edu

Abstract

This paper offers an extensive meta-analysis of the theoretical evolution of social media as a domain of research. It examines the domain’s current intellectual core with respect to dominant topics and theories of study. Hereto a systematic examination of all 610 articles published between October 2004 and December 2011 was conducted for identifying popular research topics as well as dominant theories. Findings indicate an overall lack of theory as well as a dominance of socio-psychological topics and theories of inquiry. These findings establish a benchmark for tracking the state of the evolution of the social media domain, while focusing the reader’s attention on topics and theories requiring further inquiry.

1. Introduction

Recent anecdotal observations regarding social media as a field of inquiry have emphasized the dominance of studies of social media features at the expense of theoretical accounts and explanations, thereby highlighting the largely atheoretical nature of the field (c.f., [1; 2; 3]). Although alarming, without a systematic analysis of all articles published on social media, it is difficult to offer an evidence-based balance sheet of theory-related practices in the social media domain as a starting point for evoking an informed discussion vis-à-vis the future of social media research.

Therefore, the primary objective of the current study is to develop an overview of the identity and intellectual core of the social media domain with respect to dominant topics and theories through a systematic meta-analysis of all articles published in the domain between October 2004—during which the first scholarly reference was made to social network site [4]—until including December 2011.

Given the embryonic, yet exploding, nature of social media as a field of inquiry, this paper offers an important opportunity to pause and reflect upon both what has been analyzed and achieved thus far and what needs to be addressed and accomplished in the future so as to help social media researchers direct their efforts in the most relevant and productive manner. A lack of such an historical awareness is likely to result in the intellectual deterioration of a domain [5; 6].

In order to assess the theoretical core and evolution of social media research, this study examines the field’s core research topics and concepts as well as its dominant theoretical perspectives. Hence, four related research questions were established, namely:

1. What are the most popular (i.e., dominant) research topics?
2. What proportion of social media research studies is guided by theory?
3. What are the most frequently applied (i.e., dominant) theories?
4. How have the theoretical perspectives in social media research evolved over the course of eight years of inquiry?

Answering these four questions not only allows us to establish an evidence-based benchmark for assessing the domain’s current state but also helps us to identify under-researched topics and overlooked theories that could inform the field’s future advancement.

The remainder of this paper is organized as follows: we begin with a description of our research approach. Based on our in-depth literature review of the articles in the social media domain, we reflect on the topics and theories addressed in the field.

Our findings indicate an overall lack of theory as well as a strong bias toward socio-psychological research topics and theories. Finally, implications for future social media research are discussed.

2. Methodology

This paper aims to provide an evidence-based indication of the current state of the social media domain with respect to theory-related practices. Hence, in order to understand the body of knowledge that constitutes social media as a field of inquiry with respect to research themes and theories that have been addressed so far, an in-depth literature review was conducted in the Spring of 2012. We searched for scholarly articles in the ProQuest database from October 2004 until December 2011. Based on an initial
scan of all scholarly articles on social media related search terms, it was observed that the term “social media” was only used sparingly (i.e., 17 references between 1974 and 2004) prior to October 2004 and that its prior use had little bearing to our contemporary understanding of the term. In October 2004, [4] published their seminal paper on social network sites, marking the birth of a novel field of inquiry. Hence, October 2004 was established as the appropriate starting point for this meta-analysis.

In our search, we used all ProQuest databases and conducted an advanced search using the following eight keywords: “social medium”, “social media”, “social networking site”, “social networking sites”, “social network site”, “social network sites”, “online social network”, and “online social networks”. In addition to these eight social media related keywords, we limited our search to English, full text, peer reviewed articles published in three sources (i.e., conference papers, conference proceedings, and scholarly journals) and/or as three document types (i.e., articles, conference papers, and conference proceedings).

This search resulted in a total of 1920 articles, out of which a total of 1050 unique scholarly articles1 were identified. Before proceeding to the actual literature review of topics and theories, the dataset first had to be sorted and filtered according to relevance of the articles. The main acceptance criterion for inclusion of an academic paper in this review was the use of the term social media or any of the abovementioned related search terms as either the core technology analyzed or as part of the core argument or theory developed in the paper. Hence, in order to assess whether the articles were truly relevant to this study, the two authors independently read the abstracts of all articles and judged the relevance of each article in the dataset. The initial agreement between the two coders had to be 92%. In a face-to-face meeting, disagreements were discussed and reconciled until the relevance of every article in the dataset was agreed upon. After this last step, the dataset included a total of 688 articles. For the final review, we analyzed 610 articles, since we were unable to retrieve the full text for 78 papers2 and after personally requesting a copy from the authors via email.

As aforementioned, the meta-analysis of the social media literature as presented in this paper was not restricted to journals; rather, we conducted a broad search for articles containing a social media related keyword, regardless of the primary outlet or disciplinary context. Given the embryonic and interdisciplinary nature of social media as a field of inquiry, focusing on journal articles or a specific disciplinary domain alone would provide an incomplete and myopic view of the field. Thus, when discussing social media research, we are referring to the entire body of knowledge from a broad set of outlets rather than a narrow subset of communication and/or IT journals.

2.1. Analysis and Procedure

The data analysis process was informed by a classification framework and approach that is visualized in Figure 1 and is consistent with taxonomies used in similar meta-analyses of other disciplines [5; 7; 8].

As can be seen from Figure 1, the first step in classifying articles was to decide on the use of theory or lack thereof in any given article. This was assessed through an in-depth review of the literature review and theory section of each article.

The second step was to decide if papers using theory were conceptual or empirical. In other words, was the paper based on logic or theory without collecting, analyzing, or presenting empirical data or was theory used to inform a framework guiding a data collection process or the interpretation of data. The results hereof will be presented in the third section of this paper.

The third step was to assess each paper in terms of its topical and theoretical focus. Hereto, the final data set was manually reviewed and coded by two independent coders (graduate research assistants) with respect to research questions/topics and theories. Hence, each article was systematically read and reviewed to classify important topics—based on author-identified keywords, concepts (theory section), and constructs (models and results section)—and theories—based on the literature review and/or theory section of the paper as well as key reference papers. Note, that multiple topics and theories could be assigned to each article.

With respect to topics, an open coding approach was used, where topics emerged from the paper—using the relevant sections as abovementioned. For theories, a combination of open and structured coding using a list of popular theories and allowing for additional theories to emerge from the data.

An initial interrater reliability coefficient of 0.75 (omnibus Cohen’s kappa3) and 94% (percentage

---

1 The initial search resulted in 1516 articles, of which 466 articles were duplicates. Therefore the 1050 unique scholarly articles is the total count after removing the duplicate articles from our data set.
2 We had access to the e-resources of three large University libraries
3 Cohen’s kappa coefficient is a statistical measure of interrater agreement and it is generally considered to be a more robust measure.
agreement) on a subset of the dataset provided a first assessment of the coding scheme validity and coding process reliability. Subsequently the coding scheme was assessed and disagreements were discussed and negotiated.

Following the completion of the coding of the full dataset and the assessment of adequate interrater reliability, the results of the various dimensions of the coding scheme were summarized into separate tables for further analysis of the intellectual core of the social media domain and an evidence-based assessment of prevailing topics and theories.

The final step was to decide on the popularity of each topic and theory; this was a three-step process. First, both topics and theories were clustered based on similarity through a process of axial coding. Second, for each cluster (set of topics or theories), sum scores and associated percentage scores were computed. Third, sum scores (and percentage scores) were rank-ordered based on popularity.

Figure 1. Taxonomy for the meta-analysis of theory-related practices in social media literature

3. Findings

In what follows, we report the results of our meta-analysis with respect to popular research topics, the proportion of theoretical to a-theoretical articles, dominant theories, and the evolution of theoretical perspectives overtime.

Before discussing the findings of this meta-analysis in detail, we want to highlight a few critical issues pertaining to the interpretation of these findings. First, in this project, we included all social media publications as retrieved through ProQuest for the analyzed period (October 2004 – December 2011). Although we tried to be as comprehensive in our search as possible—through the inclusion of full-text conference papers and proceedings in addition to journal articles—social media scholarship may appear in non-included outlets, including books, non-archival conference proceedings, or conference proceedings for which no full-text is available or may not have been indexed in the ProQuest database. Second, the identification of dominant research topics and theories; does not automatically correspond to research quality or impact. In order to fully understand such impact, we need to consider other factors, including scientometric and bibliometric insights (see [3]).

Hence, it is not our aim here to judge the quality or impact of particular authors or publications in the domain. Rather, we merely present the results based on an elaborate and reliable meta-analysis of the social media literature and leave it up to the reader how to interpret and utilize the findings that will be presented as follows.

3.1. Research Topics

Based on an indexing and meta-analysis of research questions as identified for the 610 social media related articles, we found 20 dominant topics in the social media domain (see Table 1). The categories in order of prevalence are:

<table>
<thead>
<tr>
<th>Popularity Rank</th>
<th>Topics</th>
<th>Count</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Education and Learning</td>
<td>85</td>
<td>13.9%</td>
</tr>
<tr>
<td>2</td>
<td>Culture/Community</td>
<td>70</td>
<td>11.5%</td>
</tr>
<tr>
<td>3</td>
<td>Healthcare</td>
<td>51</td>
<td>8.3%</td>
</tr>
<tr>
<td>4</td>
<td>Demographics</td>
<td>50</td>
<td>8.2%</td>
</tr>
<tr>
<td>5</td>
<td>Career</td>
<td>49</td>
<td>8.1%</td>
</tr>
<tr>
<td>6</td>
<td>Politics/Public Policy</td>
<td>46</td>
<td>7.6%</td>
</tr>
<tr>
<td>7</td>
<td>Psychological Well-being</td>
<td>38</td>
<td>6.2%</td>
</tr>
<tr>
<td>8</td>
<td>Post (Content)</td>
<td>35</td>
<td>5.8%</td>
</tr>
<tr>
<td>9</td>
<td>Platform Characteristics</td>
<td>32</td>
<td>5.3%</td>
</tr>
<tr>
<td>10</td>
<td>Self-Presentation/Identity</td>
<td>31</td>
<td>5.1%</td>
</tr>
<tr>
<td>11</td>
<td>Privacy/Security</td>
<td>28</td>
<td>4.6%</td>
</tr>
<tr>
<td>12</td>
<td>Social Capital</td>
<td>27</td>
<td>4.5%</td>
</tr>
<tr>
<td>13</td>
<td>Site Popularity</td>
<td>25</td>
<td>4.1%</td>
</tr>
<tr>
<td>14</td>
<td>Communication</td>
<td>25</td>
<td>4.1%</td>
</tr>
<tr>
<td>15</td>
<td>Knowledge/Information</td>
<td>24</td>
<td>4.0%</td>
</tr>
<tr>
<td>16</td>
<td>Consumerism/Branding</td>
<td>23</td>
<td>3.7%</td>
</tr>
<tr>
<td>17</td>
<td>Relationships</td>
<td>20</td>
<td>3.2%</td>
</tr>
<tr>
<td>18</td>
<td>Cyberbullying/Harassment</td>
<td>17</td>
<td>2.8%</td>
</tr>
<tr>
<td>19</td>
<td>Attitudes</td>
<td>13</td>
<td>2.1%</td>
</tr>
<tr>
<td>20</td>
<td>Disasters/Crises</td>
<td>10</td>
<td>1.7%</td>
</tr>
</tbody>
</table>
Cumulatively, the top five topics in social media research—education and learning, culture and sense of community, healthcare, demographic characteristics, career, as well as politics and public policy—account for half of the articles (50%) in the domain.

Furthermore, grouping topics together by similarity, we can observe that from the twenty popular topics, seven topics—namely demographic characteristics, psychological well-being, online identity and self-representation, social capital, relationships, cyber-bullying and harassment, as well as attitudes—pertain to characteristics or behaviors of the individual user and his or her connections.

Additionally, another six topics represent specific domains of application for social media technologies, such as education, healthcare, career development, politics, marketing (branding), and disasters (relief).

Finally, the remaining three topics represent social processes or phenomena that are facilitated and occur through the use of social media technologies, including the development of a shared culture or sense of community, communication, and knowledge or information sharing.

3.2. The Use of Theory in Social Media Research

Since [9]’s words “Nothing is quite so practical as a good theory,” (p. 169) many scholars have attempted to provide insights into why good theory is significant to the scientific enterprise. Good theory is practical and significant because it advances knowledge in a scientific discipline, guides research toward crucial questions, and enlightens the practice field associated with the domain’s core subjects [10; 11]. Hence, good theory not only helps to contribute knowledge and therewith advance a field of inquiry, but also to apply that knowledge for informing and improving practice [12].

Following this brief discussion of why theory is relevant to the scientific enterprise, we shift our focus to the results of our meta-review analysis of the social media domain, specifically pertaining to the presence or lack of theory in the social media literature.

Regarding the overall theoretical emphasis within the social media literature, our meta-analysis of 610 social media articles reveals that the majority of these papers are devoid of theory, with 455 papers (i.e., 74.6%) not referencing any theoretical foundation. These findings are in line with previous anecdotal observations regarding the dominance of studies of social media features at the expense of theoretical accounts and explanations (see [1; 2; 3]).

3.3. Dominant Theories

With respect to the particular theoretical perspective of influence to social media research, we delineated eight theoretical perspectives that each account for at least 1% (i.e., 6 papers) of the data set as follows. A summary of these eight theoretical perspectives and examples in the social media domain is provided in Table 2. This table also shows the frequency of use for each of these theoretical perspectives in raw count, percentage of the total data set (610 papers), and percentage of the 163 papers that reference theory.

Table 2. Most popular theories in social media research

<table>
<thead>
<tr>
<th>Theory</th>
<th>Disciplinary Origin</th>
<th>Count</th>
<th>% total</th>
<th>% (theory papers)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cooperation theory</td>
<td>Economics, political science, sociology</td>
<td>20</td>
<td>3.3%</td>
<td>12.3%</td>
</tr>
<tr>
<td>Network Theory</td>
<td>Computer science, graph theory, sociology</td>
<td>19</td>
<td>3.1%</td>
<td>11.7%</td>
</tr>
<tr>
<td>Social Exchange Theory</td>
<td>Social psychology, sociology</td>
<td>11</td>
<td>1.8%</td>
<td>6.7%</td>
</tr>
<tr>
<td>Social Capital Theory</td>
<td>Sociology, political science</td>
<td>10</td>
<td>1.6%</td>
<td>6.1%</td>
</tr>
<tr>
<td>Social Identity/Conformity/Influence/Comparison Theory</td>
<td>Social psychology</td>
<td>10</td>
<td>1.6%</td>
<td>6.1%</td>
</tr>
<tr>
<td>TPB/TRA</td>
<td>Social psychology</td>
<td>9</td>
<td>1.5%</td>
<td>5.5%</td>
</tr>
<tr>
<td>TAM</td>
<td>Information systems, Social psychology</td>
<td>8</td>
<td>1.3%</td>
<td>4.9%</td>
</tr>
<tr>
<td>Uses and Gratifications</td>
<td>Social psychology</td>
<td>8</td>
<td>1.3%</td>
<td>4.9%</td>
</tr>
</tbody>
</table>

The most dominant theory in the social media domain is cooperation theory. With its origin in economics, political science, sociology, and evolutionary biology, cooperation theory is a theory regarding the emergence and persistence of human cooperation [13]. Cooperation theory accounts for 3.2% of papers in the full data set and 12.3% of the 163 papers that reference theory. The second most
dominant theory is *network theory*, which originated in computer science, graph theory, and sociology. Network theory is concerned with examining the relationships between social entities [14] and in particular characteristics of the structure and nature of these relationships. Graph theory accounts for 2.9% of papers in the full data set and 11% of the 163 papers that reference theory.

Third, *Social Exchange theory*, originating from social psychology, examines how human relationships form, in particular as an outcome of a subjective cost-benefit analysis and utility function [15]. Social Exchange theory is used by 1.8% of papers in the full data set and 6.7% of the 163 papers that use theory.

Fourth, *Social Capital Theory*—which originated from sociology and political science—focuses on the value of social networks and relationships. Social capital theory explains how personal objectives can be achieved and how particular norms and behaviors can be enforced through leveraging social capital [16; 17]. It is employed by 1.6% of papers in the full data set and 6.1% of the 163 papers that reference theory. Fifth, *Social Identity, Social Conformity, Social Influence, and Social Comparison Theory* are a relatively homogeneous set of social psychological theories that aim to understand attitude change. Specifically, these theories focus on the desire for a sense of belonging, the benefits of belonging—such as status or legitimacy gains [18]—as well as the social-conformist pressures that emerge as a result of belonging to a particular social group [19; 20]. These four related theories jointly account for the theoretical foundation of 1.6% of papers in the full data set and 6.1% of the 163 papers that use theory.

Sixth, the *Theory of Planned Behavior* (TPB) and the *Theory of Reasoned Action* (TRA) are social psychological theories for understanding the relationship between attitudes and human behavior. Both theories argue that attitudes and subjective norms jointly shape an individual's behavioral intentions and subsequent behavior. TPB was suggested by [21] in order to improve the predictive power of TRA [22; 23] through the inclusion of perceived behavioral control as an additional antecedent of behavioral intention. Jointly, TPB and TRA are employed by 1.5% of all papers in the data set and 5.5% of all 163 papers that reference theory.

Seventh, the Technology Acceptance Model (TAM) [24] is an information systems theory that emerged as an extension of TRA and TPB for explaining the underlying motivations for users to accept and subsequently use a new technology. The two main motivations underpinning TAM are perceived usefulness and perceived ease-of-use. TAM is used by 1.3% of all papers in the data set and by 4.9% of all 163 papers that reference theory.

Finally, *Uses and Gratifications theory* [25] is a social psychological theory that explains why people select specific media tools and media content in order to meet one or more of several gratifications, including the need for social interactions, diversion, or knowledge enhancement. Uses and Gratifications accounts for the theoretical foundation of 1.3% of all papers in the data set and 4.9% of all 163 papers that use theory.

To sum up, our meta-analysis revealed eight dominant theoretical perspectives in the social media literature, namely Cooperation theory, Network theory, Social Exchange theory, Social Capital theory, theories of Social Identity and Social Influence, TPB and TRA, TAM, Uses and Gratifications theory.

Further reflection on these eight theoretical perspectives reveals that four of them, namely Social Exchange theory, theories of Social Identity and Social Influence, TPB and TRA, as well as Uses and Gratifications theory have their origin in social psychology, while TAM—although being an information systems theory—is build on social psychological constructs. The remaining three perspectives, Cooperation theory, Network theory, and Social Capital theory share their sociological origin.

3.4. Evolution of Theoretical Perspectives

The next step in our analysis involved identifying how the use of these eight dominant theories has evolved over time. Figure 2 shows their evolution between October 2004 and December 2011.

![Figure 2. Dominant theoretical perspectives in social media research](image)

Given the relative explosion of social media research from 2008 onwards, the trend-lines from 2008 on are most relevant and meaningful for interpretation. Hence, as Figure 2 shows, there is an evident upward trend in the use of socio-psychological theories, with all socio-psychological theories—Social Exchange
theory, Social Identity theory, TPB/TRA, TAM, and Uses and Gratifications—showing a strong growth in their application since 2008.

The remaining three theories—Cooperation, Network, and Social Capital theory—all show a growth in their application in the social media literature until 2009 and a steady decline since then.

4. Discussion

Through an extensive meta-analysis of all social media articles published between October 2004 and December 2011, we have been able to provide an evidence-based indication of the current state of the social media domain, its intellectual core, and its evolution over this eight-year period with respect to dominant research topics and theories.

Our first set of findings pertains to the dominant research topics as addressed by the 610 analyzed social media articles. We found that the top twenty topics could be classified into three broad areas. The first area included seven topics describing characteristics, behaviors, or activities of the individual user, including: demographic characteristics, psychological well-being, identity and self-presentation, social capital, relationships, cyber-bullying, and attitudes. The second area included six application domains for social media technologies, namely education, health, career, politics, marketing, as well as disasters and crises. The third and final area included three topics pertaining to social processes or phenomena that can be enabled through the use of social media technologies, including the building of culture and a sense of community, communication, as well as knowledge and information sharing.

Our second set of findings concern the prevalence of theory in social media research, or, as based on our findings, the lack thereof. Nearly three-quarters of all 610 social media related articles did not reference any theory. This further underscores concerns about the seeming lack of theory in social media research emphasized by other communication and information systems scholars (see [1; 2; 3]). Primarily in an embryonic, interdisciplinary domain like social media, this lack of theoretical grounding is alarming. Perhaps the applied nature of the field attracts practical researchers rather than those interested in theoretical pursuits.

Our third set of findings relate to the dominant theoretical perspectives in the social media domain. Amongst the 163 that were grounded in or referenced theory, eight dominant theoretical perspectives were identified, namely cooperation theory, network theory, social exchange theory, social capital theory, social identity theory, theory of planned behavior and the theory of reasoned action, the technology acceptance model, and uses and gratifications theory.

Evidently, the majority of these theories have their foundation in social psychology. Hence, despite the interdisciplinary nature of the social media domain and its central research topics, the field displays an overreliance on a single reference domain, namely social psychology.

The most commonly applied theories are those dealing with issues of interpersonal relationships and interactions. This is understandable since these are the basic elements that make up the social media domain, hence, provide the basic building blocks for further advancing the social media research domain into other, more promising directions.

Although this dominance would be appropriate for a domain dealing solely with the individual as a unit of analysis, the applied nature of the domain and the increasing awareness for social media as a tool for supporting various group and organizational level phenomena and processes, reveals that a dominance of sociopsychological theories is one-sided and myopic.

The absence of theoretical perspectives grounded in the organizational and managerial sciences and limited inter-disciplinarity is disconcerting and poses a considerable threat to our ability as social media scholars to provide useful insights, actionable guidance, and suitable recommendations to practitioners.

Furthermore, the dominance of a limited number of socio-psychological theories is evidence of a further lack of sociotechnical or “sociomaterial” perspectives and a general neglect of material considerations and theories that could inform and improve the design and utilization of social media across a variety of social contexts. Fortunately, recent special issues focusing on organizational uses of social media [26; 27] and some recent publications on social media affordances (c.f. [28; 29]), are positive indicators for increased interest and coverage of these organizational, managerial, and sociotechnical or sociomaterial topics.

Our final set of findings reveals the evolution of these dominant theories over the course of the eight years of scholarship (2004-2011) analyzed.

Although with the explosion of the social media domain, all eight theoretical perspectives display some growth over the eight years analyzed, the number of papers applying socio-psychological theories displayed the most ongoing upward trends. Although the remaining three theories—Cooperation, Network, and Social Capital theory—showed strong initial growth, more recent years showed a declining trend in their utilization.

Hence, not only are theories with their foundation in socio-psychology dominating the social media
literature, their prevalence seems to be growing as well.

Although developing detailed explanations for these trends would require further analysis, a common and frequently observed phenomenon across scientific disciplines is the Matthew Effect. Traditionally used to refer to the sociological phenomenon where the “rich get richer” [30], in the context of science it is used metaphorically to refer to the continued success and dominance of scholars, research teams, and/or institutions that were foundational to the initial shaping of the research domain [31]. Hence, the strong leverage of socio-psychological topics and theories by these early shapers of social media research leveraged seems to significantly restrict the domain’s focus today.

While moving toward prescriptions in terms of future theoretical directions for the social media domain would require further theorization, there are two areas of theorizing that seem particularly well-suited for social media research, namely multi-level theories—various organizational and network theories—as well as theories of system design and use—including theories from design science and human-computer interaction.

5. Concluding Remarks

Building on a meta-analysis of 610 social media related articles published between October 2004 and December 2011, this study provided an overview of the identity and intellectual core of the social media domain with respect to its prevailing theory-related practices. As such, this study provides an important opportunity to pause and reflect on what has been accomplished by social media scholars to date.

Moving forward from diagnosis—which reveals a very limited set of topics and theories that have hitherto dominated social media scholarship—to prescription, requires further theorizing and strategizing that is beyond the scope of this paper.

Yet, our hope is that this evidence-based reflection will encourage considerable intellectual discussion vis-à-vis the future of social media research. Most importantly, caution and critical judgment is required in order to direct the efforts of social media researchers in the most productive and relevant manner with respect scholarly significance, practical impact, and future progression.

6. References


