



Deposited via The University of Sheffield.

White Rose Research Online URL for this paper:

<https://eprints.whiterose.ac.uk/id/eprint/118613/>

Version: Published Version

---

**Article:**

Linek, S., Teka Hadgu, A., Hoffmann, C.P et al. (2017) It's all about information? The Following Behaviour of Professors and PhD Students on Twitter. *The Journal of Web Science*, 3 (1). pp. 1-15. ISSN: 2332-4031

<https://doi.org/10.1561/106.00000008>

---

Article available under the terms of the CC BY-NC-ND 2.0 licence  
(<https://creativecommons.org/licenses/by-nc-nd/2.0/>)

**Reuse**

Items deposited in White Rose Research Online are protected by copyright, with all rights reserved unless indicated otherwise. They may be downloaded and/or printed for private study, or other acts as permitted by national copyright laws. The publisher or other rights holders may allow further reproduction and re-use of the full text version. This is indicated by the licence information on the White Rose Research Online record for the item.

**Takedown**

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing [eprints@whiterose.ac.uk](mailto:eprints@whiterose.ac.uk) including the URL of the record and the reason for the withdrawal request.

# It's All About Information? The Following Behaviour of Professors and PhD Students in Computer Science on Twitter

Stephanie Linek<sup>1</sup>, Asmelash Teka Hadgu<sup>2</sup>, Christian Pieter Hoffmann<sup>3</sup>, Robert Jäschke<sup>4,2</sup> and Cornelius Puschmann<sup>5</sup>

<sup>1</sup>Leibniz Information Centre for Economics (ZBW), Kiel, Germany

<sup>2</sup>L3S Research Center, Hannover, Germany

<sup>3</sup>Institute of Communication and Media Studies, University of Leipzig, Germany

<sup>4</sup>Information School, University of Sheffield, United Kingdom

<sup>5</sup>Hans-Bredow-Institut für Medienforschung, Hamburg

## ABSTRACT

In this paper we investigate the role of the academic status in the following behaviour of computer scientists on Twitter. Based on a uses and gratifications perspective, we focus on the activity of a Twitter account and the reciprocity of following relationships. We propose that the account activity addresses the users' information motive only, whereas the user's academic status relates to both the information motive and community development (as in peer networking or career planning).

Variables were extracted from Twitter user data. We applied a biographical approach to correctly identify the academic status (professor versus PhD student). We calculated a  $2 \times 2$  MANOVA on the influence of the activity of the account and the academic status (on different groups of followers) to differentiate the influence of the information motive versus the motive for community development.

Results suggest that for computer scientists Twitter is mainly an information network. However, we found significant effects in the sense of career planning, that is, the accounts of professors had even in the case of low activity a relatively high number of researcher followers – both PhD followers as well as professor followers. Additionally, there was also some weak evidence for community development gratifications in the sense of peer-networking of professors.

Overall, we conclude that the academic use of Twitter is not only about information, but also about career planning and networking.

*Keywords:* Twitter, netiquette, open science, unilateral versus reciprocal relationships, motives for following

ISSN 2332-4031; DOI 10.1561/106.00000008

© 2017 S. Linek, A. Teka Hadgu, C. P. Hoffmann, R. Jäschke, and C. Puschmann

## 1 Introduction

A plethora of terms, such as *open science*, *science 2.0*, *cyber-science*, or *networked science* are used to describe the effect of new information and communication technology (ICT) on the scientific process (Borgman, 2007; Nentwich and König, 2012; Nielsen, 2011). Social media, such as blogs, video and presentation sharing platforms, microblogging, and social networking sites, are increasingly being used by scientists. These new media facilitate more diverse and flexible forms of communication, community development, and networking than traditional outlets (Desai *et al.*, 2012; Eysenbach, 2011; Priem and Costello, 2010; Shema *et al.*, 2012). Since social media are frequently public (e.g., blogs or Twitter), they also render scholarly communication more accessible to diverse audiences (Mauranen, 2013; Mortensen and Walker, 2002). Informal discourse in social media has been shown to frame and complement traditional research output in journal papers and monographs (Bader *et*

*al.*, 2012; Pscheida *et al.*, 2013). Consequently, some scholars have expressed hope that new ICT will render scholarly cooperation and communication more open, more interconnected and less hierarchical. At the same time, it has been contended that social media may further a Matthew effect (Merton, 1968), that is, make highly visible scholars yet more visible. In an ironic twist, social media could thereby reinforce hierarchies.

In this paper we focus on the following behaviour of professors and PhD students in the field of computer science that are active on Twitter. In contrast to other social media, such as Facebook, connections facilitated by the microblogging service Twitter are not reciprocal by default, that is, a user can choose to follow an account without being followed back. As a result, some relationships maintained within communities of Twitter users are reciprocal in nature while others are not. Accordingly, the community emerging based on platform interactions can be more or less interconnected, and more or less characterised by hierarchies. To explore whether Twitter actually

contributes to more open, interconnected and less hierarchical interactions among academics, our study investigates motives and behaviour of academic Twitter users and their effect on the reciprocity of connections among computer scientists in relation to both their status within the academic hierarchy and their activity on Twitter. We argue that two motives are key determinants of the observed following behaviour, and thereby the emerging community structure:

- The *information motive*, because Twitter is often conceptualised as an information network (Kwak *et al.*, 2010).
- The *community development motive*, including both peer networking and strategic politeness. Peer networking relates to Twitter accounts of peers, that is, persons of the same academic status. Strategic politeness refers to building social connections with persons of a higher academic status.

Data about researchers using Twitter is not readily available. We base our study on a large dataset of Twitter accounts of computer scientists (Hadgu and Jäschke, 2014). Our study contributes to the understanding of the impact of new ICT on scholarly communication and cooperation. We introduce and apply a new approach for automatically identifying the academic status of computer scientists active on Twitter, allowing us to examine the relationship between the academic status and both information and community development motives. We contribute to current research on the use of social media in academia by highlighting the effect of the academic status on the gratifications provided by Twitter. More specifically, our analysis reveals that these gratifications may differ by academic status: while PhD students primarily seek information from reputable sources on Twitter, professors attract significantly more unilateral following relationships and show some signs of community development, that is, peer networking. Thereby, both uses and gratifications of Twitter become more varied with increasing academic status. Accordingly, Twitter may contribute to more open and accessible scientific *information*, but hierarchies do seem to remain influential and interconnectedness benefits appear limited to higher status users.

The following section provides an overview of previous findings on the subject matter and describes the theoretical background of our analysis. Subsequently, the research questions and the research design are outlined in Section 3, including the identification of the independent and dependent variables as well as the formulation of hypotheses. The methodology and results sections are structured by the two main empirical steps of this study: First, the identification of the academic status of Twitter users in the field of computer science is described in Section 4. Second, the hypothesis tests addressing following behaviour are reported in Section 5. The paper closes with a discussion and an outlook.

## 2 Theoretical Background

### 2.1 Social Media Usage in Scholarly Communication

Social media are increasingly being applied to scientific communication, though scepticism persists in many fields (Ander-

son, 1997). Research suggests that motives for social media usage vary among scientists. Gruzd and Goertzen (2013) conducted an online survey among the members of three professional social science organisations and identified distinct motives for social media usage such as “keeping up to date with topics”, “following other researchers’ work” and “discovering new ideas or publications”. Information usage was found to be more prevalent than communication and community building. Lupton (2014) conducted a study among an international interdisciplinary sample of 711 academics about their usage of social media. She found that key benefits perceived by users included connecting and establishing networks with other academics as well as non-academic audiences, promoting openness and sharing of information, publicising and development of research and giving and receiving support. Procter *et al.* (2010) found Web 2.0 adaptation among academics to be influenced by demographic characteristics, such as age (the younger the more frequent social media use), gender (males being more frequent users), but also by position, and discipline. The study found computer scientists and mathematicians to exhibit the highest percentage of frequent social media use (26%). The authors also showed that collaboration practices, support, skills, and attitudes play an important part in shaping scholars’ adoption of Web 2.0 applications. To summarise, the adoption of social media in academia has been driven by information seeking, the desire to stay up to date, and curiosity towards digital technology.

### 2.2 Twitter Usage by Researchers

From a user’s perspective, Twitter has been shown to provide multiple benefits: it serves both as an instrument to stay informed, as a tool for relationship management, and as a way of signalling affinity to particular issues and people (Marwick and Boyd, 2010). Ross *et al.* (2011) described Twitter as an ideal medium to establish a “more participatory conference culture” by expanding communication and participation. Consequently, Twitter usage during academic conferences is the focal point of several analyses (Ebner, 2009; Letierce *et al.*, 2010; Weller *et al.*, 2011; Wen *et al.*, 2014). Others analysed links to scholarly articles within tweets as a form of citation and new approaches to impact assessment (Priem and Costello, 2010; Eysenbach, 2011; Weller *et al.*, 2011). One challenge common to these approaches is the identification of scholarly content on Twitter. Weller *et al.* (2011) discussed different options: identifying tweets that contain scholarly content (which they considered very difficult) or links to scholarly content (cf. Priem and Hemminger, 2010), identifying academics (as Priem and Costello (2010) and Hadgu and Jäschke (2014) did), or identifying scholarly hashtags (as done by Letierce *et al.*, 2010 and Wen *et al.*, 2014). To date, no framework to cover all approaches exists and therefore, the sets of Twitter users and tweets analysed in the existing literature are very diverse and hardly comparable.

Priem and Costello (2010) studied whether and how scholars cite on Twitter by interviewing scholars and analysing their tweets. Their study showed the relevance of Twitter for researchers beyond conferences (as pointed out by Ross *et al.*, 2011) and therefore substantiated the need for more holistic analyses of how researchers use Twitter. Weller *et al.* (2011)

focused on tweets containing conference-related hashtags and extended the view of Twitter citations introduced by Priem and Costello (2010) to retweets and mentions. They argued that more work is required to understand why users cite on Twitter and which kinds of retweet behaviour exists among users. Eysenbach (2011) also pointed out that further research is necessary addressing how academics use Twitter and whether information about collaboration and information exchange among scholars can be used to better judge the academic merit of tweets. Using a machine learning approach, Hadgu and Jäschke (2014) identified a set of 9 191 computer scientists leveraging the fact that most conferences maintain Twitter accounts, whose followers they classified as researchers. The present study is based on the same dataset.

A study by Wen *et al.* (2014) presented a longitudinal analysis of tweets and users from 16 computer science conferences over 5 years. Following the approach of Lin *et al.* (2014) and motivated by the findings of Ross *et al.* (2011), they constructed networks of users based on replies, mentions, and retweets. Although they found an increase in the information sharing activities of users from 2009 to 2013, actions indicating conversations between users are found to remain stable. The authors found more evidence of reciprocity in the reply- and mention-based conversation (between 17.2% and 23.7%) than in the retweet-based information network (between 5.3% and 7.0%). Wen *et al.*, 2014 argued that these results are to be expected, since social norms increase the likelihood of answering to replies and mentions, which directly address the users, in contrast to retweets. However, they neither investigated these differences further nor do they analyse motives for the following behaviour.

## 2.3 Researchers' Motives for Following on Twitter

### 2.3.1 Uses and Gratification Approach

Uses and gratification theory (U&G) provides a helpful approach to studying usage of ICT, such as Twitter. The uses and gratification model (Katz *et al.*, 1974) describes an active media user choosing from diverse media contents for a variety of purposes. It was originally applied to television viewing and the usage of traditional media (e.g., Conway and Rubin, 1991). Modified versions of the U&G approach have been successfully applied to new interactive media (e.g., Dimmick *et al.*, 2000; Ferguson and Perse, 2000; LaRose *et al.*, 2001; Ruggiero, 2000; Stafford *et al.*, 2004) including social media (e.g., Shao, 2009). Six motives have been identified as especially relevant to social media usage (Shao, 2009; Courtois *et al.*, 2009), namely information, entertainment, social interaction, community development, self-expression, and self-actualisation. These motives can be related to different activities in social media. According to Shao (2009) there are three separate but interdependent usages of social media: Consuming (for information and entertainment uses), participating (for social interaction and community development uses), and producing (for self-expression and self-actualisation uses). Other conceptions of motives for the use of Web 2.0 are very similar (Park *et al.*, 2009; Krishnamurthy and Dou, 2008).

From the perspective of the U&G approach, gratifications are not objective, general values but rather defined from the

subjective perspective of the individual user. Accordingly, different users can receive different gratifications from the same media and contents. This line of reasoning can also be applied to distinct social or professional roles.

Prior studies showed that motives for professional versus private use can differ. For example, an online survey with economists (Linek and Baessler, 2015) revealed that community development and information were important motives for private as well as for professional use. However, social interaction had a higher importance for private use. Additionally, the motivations entertainment, pastime, procrastination, and recreation mattered for private use only, but had no relevance for professional use.

Interviews with frequent Twitter users regarding professional use in the field of economics showed that they mainly follow other researchers to receive information on new articles, conferences and other scholarly activities (Linek and Baessler, 2015). Similarly, the perceived informativeness of a Twitter account has been shown to reduce the likelihood of unfollowing (Kwak *et al.*, 2012).

A study by Neier and Zayer (2015) on social media in higher education demonstrated that usage motives were strongly contingent upon the specific platform: For example, students were unwilling to use Facebook for educational purposes, because it was seen as a private medium. According to students, Twitter was not an effective educational tool for course management, but it could be effective for engaging students in personal reflections about the learning content. Additionally, Neier and Zayer (2015) found that the motives social interaction and information often go hand in hand when it comes to the educational use of social media.

A study with focus group interviews on researchers' use of Web 2.0 (Ostermaier-Grabow *et al.*, 2016) revealed that students were often overwhelmed by the variety of Web 2.0 services and were unable to identify the reliable and important information sources for their research. Students reported that they relied heavily on the recommendations of their lecturers and professors. This finding is in line with insights from persuasion research showing that social status positively affects the perceived competence and trustworthiness of a source (Hovland *et al.*, 1953). Formal titles, in this context, are an important indicator of status (Nawratil, 1997).

Applying these findings to the professional use of Twitter, the U&G approach suggests different possible motivations for PhD students versus professors: Given the distinct requirements of the respective professional roles, PhD candidates could more strongly be motivated by informational uses. Following the accounts of professors could help them reflect the ideas of their professors (cf. Neier and Zayer, 2015) and receive necessary input for their PhD work (cf. Ostermaier-Grabow *et al.*, 2016).

By contrast, professors might be motivated not only by the information motive but also by self-expression and self-actualisation (with respect to their reputation within the scientific community).

For both professors and PhD students, social motives might also matter. Neier and Zayer (2015) found that the motives of social interaction and information were often connected. Thus, in the case of professors and PhD students active on Twitter,

peer-networking and other more social activities might constitute relevant motives.

In this context it should also be noted that following a Twitter account provides other gratifications than following back. While following a Twitter account might be mainly motivated by information seeking, following back could be more directed by the motive of community development. Motivations of self-expression and self-actualisation, in turn, are less salient for users' following behaviour but rather use activity and the content of tweets (and replies to the tweets of others).

### 2.3.2 *Netiquette: Strategic Politeness and the Influence of the Academic Hierarchy*

In addition to the motivational perspective, scholarly communication on the internet is also influenced by so-called netiquette (e.g., Spinks *et al.*, 1999). Netiquette refers to general behavioural rules or discursive norms applied to ICT usage, such as the avoidance of defamation. Partly, these rules depend on the specific platform and social situation, for example the appropriate answering time or the usage of smiles.

Perceived social context is an important factor influencing netiquette. For example, students, even when using their university's Facebook profile for finding new friends, tend to consider the service a private context and do not wish to be contacted by university staff via Facebook (Karl and Peluchette, 2011). Similarly, occupational hierarchy can affect netiquette. Peluchette *et al.* (2013) found that Facebook friend requests by superiors are frequently perceived as problematic. Social context can play more or less of a complicating role depending on the dominant purpose of an online platform. Mainly business-related social networks, like Xing or LinkedIn, are less likely to lead to conflicts of netiquette as they are embedded in a relatively homogeneous social context. Similarly, Kwak *et al.* (2010) found that homophily affects networking, and that such effects are sensitive to characteristics such as geographical location and popularity. Accordingly, a core characteristic for homophily among researchers could be the academic status.

Another platform feature potentially affecting netiquette is the reciprocity of relationships. A connection between two users is reciprocal, when the connection is actively established by the mutual following of both users (i.e., the reciprocal connection is given and received by each of the two users). While connections on Xing or Facebook require affirmation from both parties, Twitter connections are unilateral by default: a user can follow another user without the need for the other user to follow back. In other words, connections on Twitter can be "unbalanced" (in the sense of Heider, 1958), potentially leading to netiquette uncertainty or conflict. Concerning reciprocal relationships on Twitter, Hopcroft *et al.* (2011) argued that social structures do affect online network structures in Twitter: They found that prominent, "elite" users are more likely to maintain reciprocal relationships among each other than with non-elite users. According to Kwak *et al.* (2010), the users with the most followers were often celebrities and media organisations that typically do not follow back. According to their study, Twitter showed a low level of reciprocity, since only about one fifth of the follower relationships were reciprocal, which is considerably less than, for example, on Flickr (Cha *et al.*, 2009)

and Yahoo! 360 (Kumar *et al.*, 2006). These findings suggest that in the case of Twitter, following back is not a general social rule, but rather based on characteristics of the account owner and the specific motivation of individual users.

General findings on the Twitter usage of non-academics raise the question of the role of the academic status in the following behaviour of professors and PhD students using Twitter. In the case of scholarly conversation on Twitter, netiquette might require following back. Additionally, reciprocity could also serve strategic considerations with respect to professional advancement. Two instances of such strategic considerations might be of particular importance, namely *peer networking* (in the sense of community development for scholarly social interaction) and *strategic politeness* towards persons of higher status in the academic hierarchy (for career planning).

## 3 Research Question and Research Design

### 3.1 *Outline of the Research Question*

This study analyses which attributes of a computer scientist's Twitter account influence the following behaviour directed at other computer scientists. Based on the U&G perspective, we propose that account attributes are important for following if they are in line with the motivations of potential followers. We focus on two key motives, namely information and community development.

*Information Motive:* The information motive could play a key role in academic Twitter users' following decisions for two reasons: First, previous studies have shown that for many users, Twitter has a predominantly informational character. Second, information exchange is of tremendous importance in the research process. The information motive can encompass both a quantitative and a qualitative dimension: On Twitter, a (high) quantity of information is reflected in the (high) activity of an account, that is, the number of tweets. The quality of information, instead, is related to the (assumed) expertise of the account owner. Thus, the academic status of an account owner can be conceptualised as an attribute that also signals (from a subjective point of view of the potential followers) the quality of the information provided. Thereby, the scope of our analysis is not the actual quality, but what other people think/assume about the quality. In other words: the higher the academic status of a user (professor versus PhD student), the higher the (assumed) quality of the information of his or her tweets is.

*Community Development Motive:* Despite the predominantly informational character of Twitter, community development and strategic considerations can also affect users' following decisions. Thereby, users possibly decide to follow (or follow back) even though an account offers only little information gratification. Community development in the form of peer networking (i.e., the connection with persons of the same academic status) is important for a successful academic career. On the other hand, connecting with persons of a higher status within the academic hierarchy can be helpful for career considerations. In

Table 1: Independent variables, related concepts, and addressed motives.

independent variables	related concepts	addressed motives of followers
IV1: activity of the account	quantity of information	information motive: following accounts with high activity (independent of the academic status)
IV2: academic status (Prof/PhD)	quality of information (assumed expertise/experience of the source/account owner) academic status: analogous academic status: higher	information motive: following Prof accounts due to the assumed higher quality of information (independent of the activity of the account) community development: peer networking: <ul style="list-style-type: none"> <li>• Prof – Prof (reciprocal)</li> <li>• PhD – PhD (reciprocal)</li> </ul> community development: strategic politeness/career planning by following the higher status: <ul style="list-style-type: none"> <li>• PhD → Prof (unilateral)</li> </ul>

the latter case, even an unilateral relationship (following without being followed back) can serve the community development motive. Peer networking, networking among researchers of the same academic status, can occur both among PhD students and among professors. In the case of accounts with high activity, this effect cannot be isolated from the effect of the information motive. However, in the case of accounts with low activity, peer networking alone should lead to a higher number of reciprocal relationships between researchers with the same academic status. While peer networking is important for both professors and PhD students, strategic considerations should be a more salient motive among PhD candidates who are at the very start of their academic career. In the case of professor accounts with high activity, a high number of PhD student followers can also (at least partly) be explained by the information motive. However, in the case of professor accounts with low activity, the effect of career planning alone should lead to a higher number of PhD student followers.

To summarise, we assume that two attributes of Twitter accounts influence academic users' following behaviour: First, the *activity* of the account and second, the *academic status* of the account owner. The present study analyses the following behaviour (following and following back) among researchers in computer science on Twitter. The overall research question is:

*How does an account owner's status in the academic hierarchy influence the following behaviour of fellow academic Twitter users and the resulting reciprocity of relationships within the academic Twitter network?*

We focus on users' information and community development motives. With respect to the information motive, we analyse the influence of the activity of the account (tweets) on the following behaviour. With respect to the role of academic status, we compare professors (Prof) with PhD students (PhD) as two extremes within the academic hierarchy. The research question in its operational form is:

*How does the activity of the account and the academic status of the account owner influence the following behaviour and reciprocity of connections?*

A high *activity of the account* (IV1) signals a high quantity of information and thus should address the information motive (in the sense of the U&G approach). The *academic status of the account owner* (IV2) can affect following decisions both due to the information motive as well as for community development and strategic considerations:

- A *high academic status* signals a high quality of information (even though the total amount of tweets might be low) and thus, should address the *information motive*.
- If the account owner has the *same academic status* as the follower, the motive *community development* in the form of peer networking might be addressed.
- If the account owner is of a *higher academic status* than the follower, *strategic politeness* in the sense of career planning could drive following decisions.
- (If the account owner is of lower academic status, none of the mentioned potential reasons are addressed and thus, the academic status should not influence the number of followers and reciprocal relationships.)

Table 1 gives a systematic overview on the two independent variables, the related concepts and the addressed motives for following (back).

### 3.2 Variables: Independent, Dependent, and Control Variables

This part describes the set of variables employed in the statistical analysis. The extraction of the variables is described in Section 4.2.

### 3.2.1 Independent Variables (IV)

For the statistical analyses of our research question we defined two independent variables (as explained above):

- IV1: *activity of the account*
- IV2: *academic status* (Prof versus PhD)

### 3.2.2 Dependent Variables (DV)

In order to test the influence of the two independent variables on the information and community development motives, we analysed not only the total number of followers and reciprocal relationships but also considered the academic status of the follower. Thereby, we examined the following dependent variables (DV):

- DV1: *number of researcher followers*
- DV2: *number of Prof followers*
- DV3: *number of PhD followers*
- DV4: *number of reciprocal Prof followers*, that is, the account owner (Prof or PhD) is following back
- DV5: *number of reciprocal PhD followers*, that is, the account owner (Prof or PhD) is following back

### 3.2.3 Control Variables

We included the following control variables in our analyses: *number of publications* (of the account owner), *duration in days* of the existence of the Twitter account, *language* of the account (English versus not English), and *gender*.

## 3.3 Hypotheses

### 3.3.1 Hypothesis on the Influence of the Activity of the Account

*H1 – quantity of information:* Accounts with a high level of activity (compared to low activity) have a higher number of researcher followers (including Prof followers as well as PhD followers) due to the higher quantity of information.

⇒ MANOVA:

- *Main effect for the activity of the account (IV1) on the number of researcher followers (DV1)*

### 3.3.2 Hypotheses on the Influence of the Academic Status

*H2 – quality of information:* The accounts of professors (compared to PhD accounts) have a higher number of researcher followers (including Prof followers as well as PhD followers) due to the (assumed) higher quality of information of tweets from professors.

⇒ MANOVA:

- *Main effect for the academic status of the account owner (IV2) on the number of researcher followers (DV1)*

*H3 – peer networking of professors:* Prof accounts (compared to PhD accounts) have a higher number of reciprocal Prof followers (also in the case of low activity) due to peer networking  
⇒ MANOVA:

- *Main effect for academic status (IV2) on the number of reciprocal Prof followers (DV4)*
- *Interaction<sup>1</sup> between activity (IV1) and academic status (IV2) on the number of reciprocal Prof followers (DV4)*

*H4 – peer networking of PhDs:* PhD accounts (compared to Prof accounts) have a higher number of reciprocal PhD followers (also in the case of low activity) due to peer networking.

⇒ MANOVA:

- *Main effect for academic status (IV2) on the number of reciprocal PhD followers (DV5)*
- *Interaction<sup>1</sup> between activity (IV1) and academic status (IV2) on the number of reciprocal PhD followers (DV5)*

*H5 – career planning of PhDs:* Prof accounts (compared to PhD accounts) have (also) in the case of low activity a higher number of PhD followers due to career planning of PhDs

⇒ MANOVA:

- *Interaction<sup>1</sup> between activity (IV1) and academic status (IV2) on the number of PhD followers (DV3)*

## 3.4 Answering the Research Question: Procedure

We performed two main analyses to address the research question: First, analyses of the raw data of a selected sample of academic Twitter users in order to extract the relevant variables. This included the correct identification of the academic status of the account owner, the adequate selection of relevant Twitter variables (e.g., number of tweets) and the extraction of control variables (e.g., gender). The analyses are based on a previously published dataset and are described in more detail in the following section. Second, analyses of variance (MANOVA) with respect to the five hypotheses: The hypotheses on the differentiated influence of the activity of the account (high versus low) and the academic status of the account owner (Prof versus PhD) were addressed by a  $2 \times 2$  MANOVA for independent groups. Additionally, the influence of several control variables (e.g., gender) was analysed.

## 4 Analyses of the Raw Data of Computer Scientists on Twitter

To answer the research question and to address the hypotheses, it was necessary to correctly identify Twitter accounts of

---

<sup>1</sup>Interactions for the case of low activity (comparison Prof accounts versus PhD accounts):

Please note: In the case of low activity of the account, the information motive is absent or at least not dominant. The effect of quantity of information (activity) and quality of information (academic status) are reflected in the main effects of the IVs. The interaction and the pattern for low activity can be conceptualised as distinct indicator for peer networking or career planning.

Peer networking as well as career planning will also cause some effects in the case of high activity. However, in this case the effects are not distinct (and statistically independent) from the effects due to the information motive.

researchers. Subsequently, relevant variables needed to be extracted from the dataset. Many variables, for instance, number of tweets or retweets, could easily be extracted. However, the correct identification of the academic status (“role”) of an account owner was quite difficult, as Twitter profiles do not by default contain such information. Similarly, information about the gender and home country of users was not explicitly available. In this section we briefly introduce the dataset employed in the analysis and then explain the extraction process for the necessary variables.

#### 4.1 Twitter Accounts of Researchers

Our analysis is based on data on computer scientists using Twitter (Hadgu and Jäschke, 2014). The dataset was collected based on a *seed list* of 170 Twitter accounts of computer science conferences. All Twitter users that retweeted, followed, or were followed by one of those seed accounts were collected. The resulting list of 52 678 *candidates* was then matched with 73% accuracy against the computer science bibliography DBLP (Ley, 2009) by matching the real names from the Twitter accounts with the author names from DBLP. From the resulting 9 191 *matched candidates* we removed 586 *protected* accounts, that is, accounts whose tweet and network data is not publicly available, since the account owners have made their accounts private. The remaining 8 605 users are the target sample of this analysis. Using the sample of users matched to DBLP allows us to employ publication metadata from this database in our analysis.

#### 4.2 Extraction of Variables

The variables relevant for our analysis were derived from the raw data that we retrieved from the Twitter API and auxiliary information and required some processing which we describe here:

- For the *number of organic tweets in 2013* (which indicates the activity of the account owner) we counted the number of tweets that were neither a retweet nor sent in reply to another tweet in the year 2013. For a given tweet from the Twitter API, we checked the field “*in\_reply\_to\_status\_id\_str*” to determine whether a tweet was a reply and the field “*retweeted\_status*” to check whether it was a retweet.
- Identifying the *academic status* of researchers in the dataset is a complex task for which we developed two approaches that we describe in the next section. We assigned the roles “Prof”, “PhD”, or “none” to users by using the profile-based approach. Based on the academic status of the researchers and their list of followers as it was provided by Twitter, we derived the variables *number of researcher followers*, *number of PhD followers*, and *number of Prof followers* by counting the number of followers that were in our set of 8 605 researchers or were identified as PhDs or professors, respectively.
- Analogously, we counted the number of reciprocal follow relations the users had to different subsets of users.

The variables *number of reciprocal PhD followers* and *number of reciprocal Prof followers* indicate how many of the users’ follow relationships to PhDs and professors, respectively, are reciprocal.

- The matching of Twitter accounts to DBLP authors enabled us to extract information about the users’ publication activity. The *number of publications* indicates how many publications of a user we could find in DBLP. Furthermore, we used the publication year of the first publication as a proxy for their age in the year-based approach for the identification of the academic status.
- The *duration in days* is the number of days the account had been active before January 1st 2014. It is computed as the difference between the account creation date (as it is provided by the Twitter API) and January 1st 2014. Some few accounts were created on or after this date – their value was set to zero.
- The variable *language English* is derived from the language field in the user’s profile by checking whether the value was equal to “en” or “en-gb”.
- To estimate the *gender* of Twitter users (“female”, “male”, “none”), we used an approach similar to Mislove *et al.* (2011): we matched the real names of users to lists of common names. More specifically, we combined data from (i) the US social security administration<sup>2</sup> by considering the most popular 1 000 names for each year between 1960 and 2010, (ii) the US census bureau<sup>3</sup> by taking all frequent names from the 1990 census data, and (iii) all popular baby names in Germany.<sup>4</sup> After normalising all names to lowercase, we assigned a gender to users by performing exact string matching against the names on the lists. We did not assign a gender to 3 505 users whose name is ambiguous, that is, appeared in both the male and female lists.
- The variables *number of male researcher*, *female researcher*, *female Prof*, *male Prof*, *female PhD*, *male PhD*, *reciprocal female PhD*, *reciprocal male PhD*, *reciprocal female Prof*, and *reciprocal male Prof followers* additionally consider the gender of the target users.

In addition, we used the *total number of followers* and the *total number of followees* as they were provided by the Twitter API to provide some basic statistics in Table 2.

#### 4.3 Identifying the Status of Researchers within the Academic Hierarchy

To analyse the following behaviour and the reciprocity of Twitter relationships among PhD students and professors in the field of computer science, it is necessary to correctly identify

<sup>2</sup><http://www.ssa.gov/oact/babynames/background.html>

<sup>3</sup>[http://www.census.gov/topics/population/genealogy/data/1990\\_census/1990\\_census\\_namefiles.html](http://www.census.gov/topics/population/genealogy/data/1990_census/1990_census_namefiles.html)

<sup>4</sup><http://www.beliebte-vornamen.de/>

Twitter users and ascribe them to one of the two professional roles. User profiles on Twitter do not contain information about the academic status of users by default. However, approaches exist that leverage the network structure among Twitter users to deduce their academic status. For example, Wang *et al.* (2010) focused on relationships between researchers by proposing a method to infer advisee-advisor relationships based on the coauthor network of scholarly publications. Their time-constrained probabilistic factor graph model achieves an accuracy of more than 80%. Similarly, Tang *et al.* (2012) studied the problem of inferring the type of academic relationship (e.g., advisor-advisee) by learning across heterogeneous networks. They propose a transfer-based factor graph model that learns a predictive function on a source network and infers the type of relationship on a target network. They achieved this by abstracting domain-specific features inferred from social theories (such as structural balance, structural hole, and social status). The sparseness of the coauthor network in the dataset by Hadgu and Jäschke (2014) applied in this study indicates that it contains only few scholars in an advisor-advisee relationship. Therefore, both approaches are of limited use. In this section we propose and compare two methods to extract the academic status from the Twitter profiles of users.

#### 4.3.1 The Year-Based Approach

We assume that the academic status can be identified by the progress that researchers make in their academic career over time. Typically, one progresses from PhD student to post-doc/senior researcher, assistant/associate professor to full professor. This implies that young researchers are likely PhD students. The older (still publishing) researchers are, the more likely it is that they are professors. Based on these assumptions, we can use the age of researchers and their publication activity to identify their academic status.

Since we do not have explicit information about the age of researchers, we devised an approach based on the years of their first and last publication in the DBLP dataset as a proxy for their real age. If the first paper was published after 2009 and the last paper after 2012, we considered them to be *PhD students*, whereas if the first publication appeared before 2004 and the last publication after 2012, we assigned the role *Prof.* The reasoning behind these rules is that PhD students started to publish only recently and are still publishing whereas professors started to publish some time ago and are still publishing. We also investigated other roles such as *industry* or *senior researcher*, however, it was difficult to disambiguate them. Therefore, we considered only the two roles *PhD* and *Prof.*

#### 4.3.2 The Profile-Based Approach

We identified Twitter users that have certain terms in the “bio” field of their Twitter profiles indicating their status (cf. Figure 1). Therefore, we matched a list of keywords for each role against the description in the users’ bio field. We identified professors by using the keywords “prof”, “professor”, and “profesor” for matching. In the remaining set of users we identified PhD students by searching for the keywords “phd”, “ph.d”, “ph d”, “graduate student”, “grad. student”, “doctoral student”, and

**Jure Leskovec**  
 @jure  
 Professor of #computerscience  
 @Stanford. Thinking about #datamining  
 massive social and information  
 #networks, #bigdata, #web and  
 #socialmedia.

Figure 1: A sample Twitter profile as shown on <https://twitter.com/jure>. The content of the “bio” field is shown below the user name “@jure”.



Figure 2: Tag clouds of terms from the Twitter profiles for professors (a) and PhD students (b) that were identified by the year-based approach.

“doctoral candidate”. All matches were performed by ignoring the case of words.

#### 4.4 Analysis of the Status Identification

The year-based approach allowed us to assign the role “PhD” (2283) or “Prof” (1851) to 4138 (48%) of the 8605 identified researchers. To qualitatively evaluate this approach, we extracted terms from the “bio” field of the researchers’ Twitter profiles (cf. Figure 1) and visualised them in a word cloud (Figure 2). The font size of each word reflects the number of researchers that have the corresponding word in their self-description.

Although each word cloud clearly shows that the corresponding roles appear very frequently in the profile, it also shows the fuzziness of this method: Both the terms “phd” and “student” also appear in the professors’ cloud and the term “professor” in the PhD students’ cloud. Based on this observation, we developed the more robust profile-based approach for the identification of the academic status.

For comparison, we analysed the distribution of authors that have published their first paper in the same year for the different groups identified by the profile-based approach. Figure 3 shows for each year, how many authors have published their first paper in that year for all users, users from the “Prof” group, and users from the “PhD” group. We can see that most PhD students have published their first paper recently, in contrast to professors.

The figure explains the low discriminatory power of the year-based approach for professors: their first publication year has a high variance. To test the quality of the role assignment using the profile-based approach, we manually checked a sample of 100 users. The overall accuracy is 94%. With respect to each role, it is 100% (40/40) for the “Prof” group and 90%

Table 2: Basic statistics about the Twitter users for that we could identify their academic status.

user group	Prof	PhD	Prof + PhD	all researchers
number of identified users	570	911	1 481	8 605
mean / median total number of followers	756 / 208	283 / 87	465 / 120	3 842 / 84
mean / median total number of followees	292 / 163	278 / 147	284 / 155	556 / 136

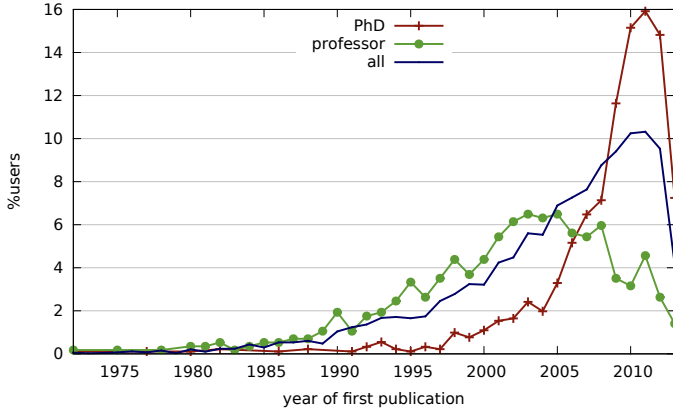


Figure 3: The distribution of the users from the “PhD” and “Prof” group that were identified by the profile-based approach over the different years of first publications of authors.

(54/60) for the PhD group. Interestingly, only one of the miss-classified users was actually a professor who happened to refer to his PhD on his Twitter profile. The other miss-classified users were not professors but also could not clearly be classified as PhD students. Mostly, these were individuals who after completing their PhD left academia. Since we preferred a good precision over a high recall, we finally decided to use the results from the profile-based method for the remainder of this analysis.

The profile-based approach could identify 1 481 (17%) professors and PhD students, most of which (62%) are PhD students (cf. Table 2). This fraction is similar to the 55% in the year-based approach. This slightly unequal distribution is not surprising, given that there are more PhD students than professors. Based on age demographics, one could even expect to have more professors, since according to Jackson (2010) 54% of the Twitter users are older than 26.

#### 4.5 Basic Statistics About the Identified Researchers

On average, professors had considerably more followers (756) than PhD students (283), even the median number of followers was more than twice as high (208 versus 87). For both groups combined, the median number of followers (120) was higher than for all researchers (84). The same did not apply to the number of followees, though: professors and PhD students had similar mean numbers of followees (292 and 278, respectively) and also their medians were more similar – to each other (163 and 147, resp.) and to all researchers (136).

Table 3 shows the number (and percentage) of unilateral and reciprocal follow relationships between users with different

Table 3: Follow relationships between the groups “PhD” and “Prof” and for all researchers.

	unilateral	reciprocal
all researchers	75 684 (73%)	27 358 (27%)
Prof → PhD	298 (22%)	1 088 (78%)
PhD → Prof	2 585 (70%)	1 088 (30%)
PhD → PhD	1 095 (57%)	817 (43%)
Prof → Prof	1 478 (59%)	1 016 (41%)

academic status. The row “Prof → PhD” shows that only a small portion (22%) of the following relationships from professors to PhD students were unilateral (i.e., the students were not following the professors). On the other hand, in the case of PhD students following professors, most of these relationships (70%) were unilateral. Unfortunately, the Twitter API does not provide information about the time a relationship was established. Therefore, it was difficult to estimate which of the users started the relationship in the reciprocal cases, rendering further analyses necessary (cf. Meeder *et al.*, 2011; Zhang *et al.*, 2014). The following behaviour within the groups was very similar: around 60% of the relationships were unilateral which was slightly lower than the 73% for all 8 605 researchers. Overall, the share of reciprocal relationships among professors and PhD students was higher than the 27% for all researchers, which in turn was close to the 22.1% reported by Kwak *et al.* (2010). A similar analysis of the *retweet* and *mention* activities can be found in Appendix A.

Overall, this analysis shows a large deviation of the fraction of reciprocal relationships between professors and PhD students compared to relations among all researchers and also among PhD students or professors only. The descriptive results on the followers and followees suggest an imbalance of following behaviour of professors versus PhD students. However, the descriptive results were not distinctive with respect to the research question, that is, the underlying motives and the related hypotheses. It remains unclear if the imbalance is due to the (assumed) higher qualitative informational value of professor accounts or due to community development in the sense of strategic politeness/career planning. This emphasises the need for an in-depth analysis that can uncover the possible reasons underlying these observations.

## 5 Analyses of the Variance on the Five Hypotheses: Influence of the Activity and the Academic Hierarchy

In the following the main results with respect to the hypotheses are reported. We start with the description of the data

sample and the categorisation of low versus high activity. Subsequently, we present the findings of the analysis of variance on the five hypotheses. The section closes with a short summary of the findings.

### 5.1 Description of the Data Sample

For the statistical tests of the hypotheses we defined the activity of the account (“high” versus “low”) by the number of organic tweets in 2013. Organic tweets are tweets created by the account owner and thus, can be considered as the most original form of activity. (In our sample, the number of organic tweets in 2013 was highly correlated with the number of retweets in 2013 ( $r = .471, p < .001$ ), with the number of replies in 2013 ( $r = .546; p < .001$ ) as well as the total number of tweets ( $r = .632; p < .001$ ). Control analyses with these alternative indicators for the quantity of information showed an analogous pattern of results.)

The distribution of the number of organic tweets in 2013 had a high skewness, that is, there were many accounts with a low number of tweets and only few accounts with a high number of tweets. In face of the skewness a median-split was not appropriate to divide low activity versus high activity. Thus, we used a Q3-split (with Q3 of the number of organic tweets in 2013 = 110.00). The test of the hypotheses was done by a  $2 \times 2$  multivariate analysis of variance (MANOVA) with the two factors activity (IV1 = “high” versus “low” defined by the Q3-split for number of organic tweets in 2013) and academic status (IV2 = “Prof” versus “PhD”). The five dependent variables were not independent from each other (since they were partly subgroups of each other) and therefore significantly correlated (correlations between  $r = .938$  and  $r = .373$  with  $p < .001$  for all correlations). Thus, we chose multivariate analysis of variance (MANOVA), that is, analysing all DVs simultaneously (instead of single ANOVAs for each DV) in order to reconsider the inter-dependencies between the DVs.

The statistical values of the MANOVA ( $F$ -value and  $p$ -value) are provided in the subsequent text. The descriptive statistics (means and standard deviations) for the dependent variables (in dependency of the independent variables) are listed in Table 4.<sup>5</sup>

## 5.2 Statistical Testing of the Hypotheses

### 5.2.1 Statistical test of H1 and H2 (influence of the information motive)

Both hypotheses were supported by the results of the MANOVA.

*H1 – quantity of information:* There was a main effect for the activity of the account (IV1) on the (DV1) number of researcher followers ( $F = 61.535$  and  $p < .001$ ). Analogous main effects were found for the (DV2) number of Prof followers ( $F = 48.490; p < .001$ ) and the (DV3) number of PhD followers ( $F = 59.529; p < .001$ ). Accounts with a high level of activity (compared to low activity) had a higher number of researcher followers including Prof followers as well as PhD followers.

<sup>5</sup>Please note: Table 4 lists the statistics for the identified researcher followers. Thus, the numbers are substantially lower than the total number of followers in Table 2. For details see Section 4.2.

*H2 – quality of information:* There was a main effect for the academic status of the account owner (IV2) on the (DV1) number of researcher followers ( $F = 78.620; p < .001$ ). Analogous main effects were found for the (DV2) number of Prof followers ( $F = 142.109; p < .001$ ) and the (DV3) number of PhD followers ( $F = 37.323; p < .001$ ). The accounts of professors (compared to PhD accounts) had a higher number of researcher followers including Prof followers as well as PhD followers.

To summarise, the number of researcher followers was significantly influenced by the quantity of information as well as the quality of information indicating a high importance of the information motive.

### 5.2.2 Statistical test of H3 (peer networking of Prof)

The results provided no clear support for H3: There was a main effect for the academic status ( $F = 95.960; p < .001$ ) and a main effect for the activity of the account ( $F = 52.764; p < .001$ ) on the (DV4) number of reciprocal Prof followers. Prof accounts (compared to PhD accounts) and accounts with a high level of activity (compared to low activity) had a higher number of reciprocal Prof followers. But there was only a non-significant tendency for the interaction between activity and academic status ( $F = 3.676; p = .055$ ).

The two main effects can be explained in the sense of H1 and H2, that is, the quantity and quality of information influences also the number of reciprocal Prof followers.

However, it is important to note that the main effect for the academic status (i.e., Prof accounts had more reciprocal Prof followers than PhD accounts) can be also interpreted as kind of peer networking. In addition, for the interaction between academic status and activity there was a marginal tendency (with  $p = .055$ ) in the sense of peer networking. Thus, the data provided some evidence for peer networking of professors.

### 5.2.3 Statistical test of H4 (peer networking of PhD)

The results provided no support for H4. There was only a main effect for the activity of the account ( $F = 63.828; p < .001$ ) on the (DV5) number of reciprocal PhD followers, but there was no main effect for the academic status ( $F = 0.119; p = .730$ ) and no significant interaction ( $F = 0.728; p = .394$ ).

Thus, the results provided no evidence for peer networking of PhD students.

Even though the results did not support H4, it was an interesting finding that for the number of reciprocal PhD followers, the academic status makes no difference.

### 5.2.4 Statistical test of H5 (career planning of PhD)

The results did support H5. As reported for H1 and H2 there was a significant main effect for the activity of the account ( $F = 59.529; p < .001$ ) and a significant main effect for the academic status ( $F = 37.323; p < .001$ ) on the number of PhD followers that were in line with the predictions of H1 and H2 on the influence of the information motive. In addition, there was a significant interaction ( $F = 5.903; p = .015$ ) that provided evidence for career planning/strategic politeness of PhD

Table 4: Descriptive statistics of the tested dependent variables: means and standard deviations (in brackets) for the experimental groups (activity  $\times$  academic status).

		DV1: number of researcher followers	DV2: number of Prof followers	DV3: number of PhD followers	DV4: number of reciprocal Prof followers	DV5: number of reciprocal PhD followers
activity low	PhD	9.82 (18.49)	1.11 (2.33)	2.32 (4.44)	0.90 (1.82)	1.44 (2.16)
	Prof	25.85 (47.33)	4.74 (7.98)	4.31 (9.14)	2.73 (4.61)	1.19 (2.52)
	all	15.17 (32.11)	2.32 (5.27)	2.98 (6.46)	1.51 (3.17)	1.36 (2.29)
activity high	PhD	23.23 (39.18)	2.84 (5.14)	5.18 (8.96)	2.14 (3.64)	2.93 (4.40)
	Prof	52.76 (83.90)	8.37 (12.30)	9.79 (17.61)	4.87 (7.14)	3.03 (6.52)
	all	38.19 (67.33)	5.64 (9.86)	7.52 (14.20)	3.53 (5.84)	2.98 (5.57)
PhD (all)		13.00 (25.61)	1.52 (3.31)	3.00 (5.96)	1.19 (2.44)	1.79 (2.92)
Prof (all)		36.33 (65.36)	6.16 (10.04)	6.44 (13.36)	3.56 (5.82)	1.91 (4.60)
total		21.98 (46.63)	3.31 (7.11)	4.32 (9.66)	2.11 (4.24)	1.84 (3.66)

students on Twitter. Professor accounts had in the case of low activity a comparatively higher number of PhD followers, that is, the significant interaction indicates that the number of PhD followers for professors' accounts could not be explained solely by the information motive.

As statistical evidence of the nature (direction/meaning) of the interaction we calculated one-way MANOVAs for the influence of the activity (IV1) separately for professor accounts and PhD accounts (using the academic status as filter). We found significant effects for the activity for professor accounts ( $F = 23.754; p < .001$ ) as well as for PhD accounts ( $F = 39.556; p < .001$ ). However, the effect size (partial eta squared) of the influence of the activity was lower for professor accounts ( $\eta^2 = .040$ ) compared to PhD accounts ( $\eta^2 = .042$ ). Thus, the identified interaction was in line with H5: The activity of the accounts of professors has a lower influence on the number of PhD followers (compared to the activity of PhD accounts). This provides evidence that the number of PhD followers of professor accounts can not solely be explained by the information motive but rather is also influenced by politeness/career planning of the PhD followers.

### 5.2.5 Additional findings in relation to H5 (career planning)

With respect to H5 on career planning it is important to note that we also found analogous interactions for the number of researcher followers ( $F = 6.902; p = .009$ ) and for the number of professor followers ( $F = 6.129; p = .013$ ). The statistical analysis of the direction of the interaction by separate one-way MANOVAs (for the IV1 activity) for professor accounts and PhD accounts revealed the same pattern as for PhD followers (described for H5). For the researcher followers we found a significant influence of the activity for professor accounts ( $F = 23.887; p < .001$ ) as well as for PhD accounts ( $F = 47.433; p < .001$ ), but the effect size was lower for professor accounts ( $\eta^2 = .040$ ) compared to PhD accounts ( $\eta^2 = .050$ ). Similarly, also for the number of professor followers, we found a significant influence of the activity for professor accounts ( $F = 18.274; p < .001$ ) as well as for PhD accounts ( $F = 47.124; p < .001$ ), but

the effect size for professor accounts ( $\eta^2 = .031$ ) was lower than for PhD accounts ( $\eta^2 = .049$ ). These results indicate that not only PhD students but also professors (and researchers) showed the same kind of career planning/politeness behaviour when following the accounts of professors.

### 5.2.6 Analyses of control variables

None of the control variables tested for changed the pattern of results reported in Section 5, except gender. In a separate analyses for males versus females (gender as filter) we found no significant interactions between academic status and activity for the dependent variables total number of researcher follower and number of PhD followers. This finding is somehow different from the reported findings but there was no difference between the findings for males and females. Thus, the non-significant interactions can probably be ascribed to the lower number of valid cases (for several accounts it was not possible to identify the gender of the account owner).

However, comparing male versus female Twitter users did reveal that the accounts of females (compared to the accounts of males) had less male followers but more female followers. Additionally, females' accounts had more reciprocal followers compared to the accounts of males. This indicated a gender-specific networking behaviour of males versus females. Even though the gender-related findings were interesting, they had only minor relevance to our study, as for most of the accounts the variable gender could not be identified and the portion of females is rather small.

## 5.3 Summary of the Findings

Overall, the results of the analyses of variance suggest that Twitter is mainly an information network with only secondary effects of community development in the sense of peer networking. We found only weak evidence that peer-networking may affect the following behaviour among professors, yet did not find analogous effects among PhD students. However, our data

also provided evidence for career planning/politeness in the following behaviour of computer scientists on Twitter. This holds true not only for PhD followers but also for professor followers (and the total number of researcher followers).

## 6 Discussion & Conclusion

Our study set out to analyse the role of the academic status in the reciprocity of following relationships among Twitter users in the field of computer science. We differentiate two key motives for following relationships, information and community development, and analyse their respective effect on following relationships on Twitter. Thereby, we contribute to the ongoing discourse on the effect of new ICTs on the openness, interconnectedness and the role of hierarchy in academia and the research process.

We found that interactions between computer scientists on Twitter were mainly driven by information motives. This part of our result conforms with Kwak *et al.* (2010) finding that Twitter is informational in nature, rather than a tool for maintaining relationships (cf. Johnson and Yang, 2009). However, our results also provided evidence for an effect of peer networking and strategic politeness on following decisions on Twitter: Our initial descriptive analysis of Twitter user data suggests an imbalance of following behaviour of professors versus PhD students, with only few professors unilaterally following PhD students, but a large number of PhD students unilaterally following professors. Based on our hypothesis test, it appears likely that this observed imbalance is largely due to the (assumed) quality of information provided by professors' Twitter accounts. Although we did not find a general social rule on following back for computer scientists on Twitter (Kwak *et al.*, 2010), we did find a higher degree of reciprocal following relationships among professors than among PhD students. Thereby, social norms may vary by academic position. Our results furthermore indicate that regardless of academic seniority, career planning/strategic politeness behaviour played a role when following the accounts of professors. This effect is particularly pronounced among PhD students when following professor accounts.

In summary, academic status appears to affect the following behaviour of computer scientists on Twitter twofold: First and foremost, (higher) academic status as an indicator of content quality induces following behaviour. Accordingly, professors attract significantly more unilateral followers than PhD students. Additionally, we find that the number of PhD followers for professors' accounts could not be explained solely by the information motive. Rather, strategic politeness appears to guide PhD following behaviour more strongly than that of professors. Secondly, while we found no evidence of peer networking among PhD students on Twitter, professors attract significantly more reciprocal following relationships by professors than do PhD students. On Twitter, professors seem to constitute a group of elite users, primarily following each other (Hopcroft *et al.*, 2011) and attracting more unilateral following relationships (cf. Xu *et al.*, 2013).

These findings go to the heart of our research question, as it indicates that academic hierarchies do in fact influence

the communication and networking among computer scientists on Twitter. Thereby, Twitter may provide different gratifications depending on a user's academic status: While information motives play an important role for both professors and PhD students in terms of following behaviour, these dynamics disproportionately provide networking opportunities for high-status academics. Also, while community development motives do not appear to motivate PhD students, we found some evidence of reciprocity norms, peer-networking and thereby community development among professors. Young researchers, instead, are geared more towards career development/strategic politeness as they unilaterally follow higher academic status accounts. Our study, thereby, provides little evidence in favour of the expectation that social media will render the research process less hierarchical. Of course, other social media like, for example, ResearchGate or Academia.edu, might provide more opportunities for fostering reciprocal academic relationships among all academic career levels and levelling traditional social hierarchies.

Accordingly, our findings have implications for the role of social media as an instrument for the perceived democratisation of academia, at least in relation to the specific case of Twitter and computer science. They can be interpreted against the background of social media exerting an influence on the speed and efficiency of scholarly communication, but not necessarily challenging academic hierarchies of distinction and seniority. Considering the fundamentally competitive nature of academia, changes in the available modes of communication are unlikely to affect the existence of academic hierarchies, even if their composition is ultimately impacted by the skill with which particular communication channels are used to amass social capital. The Utopian vision of democratised academia, therefore, appears unlikely to emerge due to technology alone. By contrast, Twitter serves as a powerful tool of making the social influence carried by academic seniority manifest and tangible, effectively reflecting it in the follower-follower graph.

Our analysis provides a valuable contribution to future analyses of Twitter use by researchers by introducing and comparing new approaches for identifying academic roles on Twitter. Compared to prior approaches (Wang *et al.*, 2010; Tang *et al.*, 2012) the profile-based method for the identification of the academic status is simple, but it also fits the available Twitter data that has only a sparse underlying co-author graph. Although the method could only identify the academic status for 17% of the researchers in the sample, a qualitative comparison with the year-based approach indicates much better precision. Further studies to evaluate the method and compare it with existing approaches were out of scope of this work but are a promising task for future research.

Our study features also some limitations that provide opportunities for future research: First, the findings of our study are limited to computer scientists that are using Twitter. A generalisation to the entire researcher population is difficult, since different disciplinary backgrounds and norms likely induce different behaviour on Twitter. An important next step therefore would be the comparison of our findings with different user communities. This includes other research areas but also communities of practice and the general Twitter population.

Similarly, a comparison of male- versus female-favoured re-

search domains presents in itself as an interesting research topic. Our control analyses on gender also provided first preliminary indications that following behaviour and rules of reciprocity might be different for male versus female academics. This also implies that the findings for a predominantly male domain like computer science might be different compared to a domain with gender parity. Since our sample had only a very small portion of female users, it was not possible to provide hard evidence for gender-related differences in following behaviour. However, this is an important issue for future studies, especially in the face of the ongoing discussion on the gender-gap in research and the general use of social media.

It should be noted that it was not the aim of our study to assess the wider dissemination of knowledge within civil society. Instead we limited ourselves to an analysis of hierarchies within (one discipline of) the scientific community. Additionally, we focused on following behaviour, which may serve other roles and be governed by different expectations than other means of interaction.

As this study is based on a cross-sectional analysis, we did not analyse who initiated a reciprocal relationship among academic Twitter users, since this information is not provided by Twitter. Recovering the temporal evolution of reciprocal relations could provide further insights into which users initiate relationships and why. Two approaches appear promising in this context that were out of scope for this work: The approach by Meeder *et al.* (2011) is based on the order of followers and information about their registration date as returned by Twitter. Zhang *et al.* (2014) proposed an optimisation framework based on the network topology. They outlined four consistency hypotheses to describe the observed directionality patterns in real-world directed social networks which were used to recover the hidden directionality of undirected social ties. Future research could use panel designs to collect longitudinal data and explain changes over time.

Another interesting open question for future research concerns the possible confluence in case of the usage of different social media by academics. Potentially, the academic usage of multiple social media (e.g., ResearchGate and Twitter in parallel) might influence each other. Given that different platforms are used for different purposes and might be associated with distinct usage norms, the simultaneous use of more than one platform might affect how these platforms are employed. It would be worthwhile to examine the role of academic hierarchies on social media platforms that are geared more towards community development, rather than information dissemination. Possibly, such networks might be more closely related to the underlying “real” social graph (cf. Cha *et al.*, 2010; Kwak *et al.*, 2012). Finally, our analysis focused on two motives derived from uses and gratifications theory: information and community development. Further in-depth analyses could differentiate the observable following behaviour into more detailed motives. Similar to Gruzd and Goertzen (2013) one could, for example, analyse aspects like “keeping up to date with topics”, “following other researchers work”, or “discovering new ideas and publications”. Identifying these aspects automatically is challenging but could be based on the content of the tweets. Ultimately, though, a comprehensive user survey would be required to obtain conclusive results.

## Appendix A Analysis of the Mention and Retweet Activities of Researchers

Similar to the analysis for the *follow* activity presented in Section 4, we performed an analysis for the common Twitter activities *mention* and *retweet* by counting pairs of users in which the first user mentioned (retweeted) the second user in some tweet and the second user did/did not mention (retweet) the first user in some tweet, which results in reciprocal/unilateral relationships, respectively. Table 5 shows the results for this analysis (for comparison we included the results for the follow activity that we presented in Table 3). We could observe that there were fewer mention relations than follow relations and even fewer retweet relations. Furthermore, the share of reciprocal relationships was decreasing from 27% for following over 23% for mentioning to 11% for retweeting in the complete dataset and analogously for the remaining subsets of users. This corresponds to the study by Wen *et al.* (2014) who found a higher share of reciprocal connections (17.2% to 23.7%) in a mention-based network than in a retweet-based network (5.3% to 7.0%). Again, all subsets of relationships among professors and PhD students exhibited a higher share of reciprocal relationships than on the complete dataset.

## References

- Anderson, C. M. 1997. “Computer-mediated communication, the invisible college and visiting scholars: An exploratory study”. *Journal of Computing in Higher Education*. 9(1): 27–48.
- Bader, A., G. Fritz, and T. Gloning. 2012. “Digitale Wissenschaftskommunikation 2010-2011: Eine Online-Befragung”. In: Gießen: Gießener Elektronische Bibliothek. 113.
- Borgman, C. L. 2007. “Scholarship in the digital age: Information, infrastructure, and the Internet”. In: Cambridge: MIT Press. 336.
- Cha, M., H. Haddadi, B. Benevenuto, and P. Gummadi. 2010. “Measuring user influence in Twitter: The million follower fallacy”. In: *Proceedings of the 4th International Conference on Weblogs and Social Media*. Menlo Park: AAAI. 10–17.
- Cha, M., A. Mislove, and K. Gummadi. 2009. “A measurement-driven analysis of information propagation in the Flickr social network”. In: *Proceedings of the 18th International Conference on the World Wide Web*. New York: ACM.
- Conway, J. and A. Rubin. 1991. “Psychological predictors of television viewing motivation”. *Communication Research*. 18(4): 443–463.
- Courtois, C., P. Mechant, L. D. Marez, and G. Verleye. 2009. “Gratifications and seeding behavior of online adolescents”. *Journal of Computer-Mediated Communication*. 15: 109–137.
- Desai, T., A. Shariff, M. Kats, X. Fang, and C. Christiano. 2012. “Tweeting the meeting: An in-depth analysis of Twitter activity at Kidney Week 2011”. *PLoS ONE*. 7(7).
- Dimmick, J., S. Kline, and L. Stafford. 2000. “The gratification niches of personal e-mail and the telephone”. *Communications Research*. 27(1): 227–248.

Table 5: Follow, mention and retweet relationships between the groups “PhD” and “Prof” and for all researchers in our dataset.

activity	follow		mention		retweet	
	unilateral	reciprocal	unilateral	reciprocal	unilateral	reciprocal
all researchers	75 684 (73%)	27 358 (27%)	52 915 (77%)	15 961 (23%)	39 447 (89%)	5 022 (11%)
Prof → PhD	298 (22%)	1 088 (78%)	485 (41%)	707 (59%)	357 (61%)	226 (39%)
PhD → Prof	2 585 (70%)	1 088 (30%)	1 407 (67%)	707 (33%)	1 123 (83%)	226 (16%)
PhD → PhD	1 095 (57%)	817 (43%)	719 (64%)	401 (36%)	593 (86%)	97 (14%)
Prof → Prof	1 478 (59%)	1 016 (41%)	1 288 (69%)	585 (31%)	1 059 (83%)	220 (17%)

- Ebner, M. 2009. “Introducing Live Microblogging: How Single Presentations Can Be Enhanced by the Mass”. *Journal of Research in Innovative Teaching*. 2(1): 91–98.
- Eysenbach, G. 2011. “Can tweets predict citations? Metrics of social impact based on Twitter and correlation with traditional metrics of scientific impact”. *Journal of Medical Internet Research*. 13(4).
- Ferguson, D. and E. Perse. 2000. “The world wide web as a functional alternative to television”. *Journal of Broadcasting and Electronic Media*. 44(2): 155–174.
- Gruzd, A. and M. Goertzen. 2013. “Wired academia: Why social science scholars are using social media”. In: *Proceedings of the Annual Hawaii International Conference on System Sciences*. 3332–3341.
- Hadgu, A. and R. Jäschke. 2014. “Identifying and analyzing researchers on Twitter”. In: *Proceedings of the Conference on Web Science*. New York: ACM. 23–32.
- Heider, F. 1958. *The psychology of interpersonal relations*. New York: Wiley.
- Hopcroft, J., T. Lou, and J. Tang. 2011. “Who Will Follow You Back?: Reciprocal Relationship Prediction”. In: *Proceedings of the 20th International Conference on Information and Knowledge Management*. New York: ACM. 1137–1146.
- Hovland, C., I. Janis, and H. Kelley. 1953. *Communication and Persuasion: Psychological Studies of Opinion Change*. New Haven: Yale University Press.
- Jackson, N. 2010. “Facebook vs. Twitter: Comparing Social Demographics”. *The Atlantic*. Dec.
- Johnson, P. and S. Yang. 2009. “Uses and gratifications of Twitter: An examination of user motives and satisfaction of Twitter use”. In: *Communication Technology Division of the annual convention of the Association for Education in Journalism and Mass Communication*. Boston.
- Karl, K. and J. Peluchette. 2011. ““Friending” professors, parents and bosses: A Facebook connection conundrum”. *Education for Business*. 86: 214–222.
- Katz, E., J. Blumler, and M. Gurevitch. 1974. “Utilization of mass communication by the individual”. In: *The uses of mass communications: Current perspectives on gratifications research*. Ed. by J. Blumler and E. Katz. Beverly Hills: Sage. 19–32.
- Krishnamurthy, S. and W. Dou. 2008. “Advertising with user-generated content: A framework and research agenda”. *Interactive Advertising*. 8(2): 1–7.
- Kumar, R., J. Novak, and A. Tomkins. 2006. “Structure and evolution of online social networks”. In: *Proceedings of the 12th SIGKDD International Conference on Knowledge Discovery and Data Mining*. New York: ACM.
- Kwak, H., C. Lee, H. Park, and S. Moon. 2010. “What is Twitter, a Social Network or a News Media?” In: *Proceedings of the 19th International Conference on the World Wide Web*. New York: ACM. 591–600.
- Kwak, H., S. Moon, and W. Lee. 2012. “More of a Receiver Than a Giver: Why Do People Unfollow in Twitter?” In: *Proceedings of the 6th International Conference on Weblogs and Social Media*. Dublin: The AAAI Press. 499–502.
- LaRose, R., D. Mastro, and M. Eastin. 2001. “Understanding Internet usage: a social-cognitive approach to uses and gratifications”. *Social Science Computer Review*. 19(4): 395–413.
- Letierce, J., A. Passant, J. Breslin, and S. Decker. 2010. “Understanding how Twitter is used to widely spread scientific messages”.
- Ley, M. 2009. “DBLP: some lessons learned”. *Proceedings of the VLDB Endowment*. 2(2): 1493–1500.
- Lin, Y.-R., B. Keegan, D. Margolin, and D. Lazer. 2014. “Rising Tides or Rising Stars?: Dynamics of Shared Attention on Twitter during Media Events”. *PLoS ONE*. 9(5): 1–12.
- Linek, S. and J. Baessler. 2015. “The role of libraries in Science 2.0 – focus on economics”. *D-Lib Magazine*. 21(7/8).
- Lupton, D. 2014. ““Feeling Better Connected’: Academics’ Use of Social Media”.
- Marwick, A. and D. Boyd. 2010. “I tweet honestly, I tweet passionately: Twitter users, context collapse, and the imagined audience”. *New Media & Society*. 13(1): 114–133.
- Mauranen, A. 2013. “Hybridism, edutainment, and doubt: Science blogging finding its feet”. *Nordic Journal of English Studies*. 12(1): 7–36.
- Meeder, B., B. Karrer, A. Sayedi, R. Ravi, C. Borgs, and J. Chayes. 2011. “We Know Who You Followed Last Summer: Inferring Social Link Creation Times in Twitter”. In: *Proceedings of the 20th International Conference on the World Wide Web*. New York: ACM. 517–526.
- Merton, R. 1968. “The Matthew effect in science”. *Science*. 159(3810): 56–63.
- Mislove, A., S. Lehmann, Y.-Y. Ahn, J.-P. Onnela, and J. Rosenquist. 2011. “Understanding the Demographics of Twitter Users”. In: *Proceedings of the 5th International Conference on Weblogs and Social Media*. AAAI. 554–557.

- Mortensen, T. and J. Walker. 2002. "Blogging thoughts: personal publication as an online research tool". In: *Researching ICTs in context*. Oslo: InterMedia. 249–279.
- Nawratil, U. 1997. *Glaubwürdigkeit in der sozialen Kommunikation*. Opladen: Westdeutscher Verlag.
- Neier, S. and L. T. Zayer. 2015. "Students' perceptions and experiences of social media in higher education". *Journal of Marketing Education*. 37(3): 133–143.
- Nentwich, M. and R. König. 2012. "Cyberscience 2.0: Research in the age of digital social networks". In: Frankfurt am Main: Campus. 237.
- Nielsen, M. 2011. "Reinventing Discovery: The New Era of Networked Science". In: Princeton: Princeton University Press. 280.
- Ostermaier-Grabow, A., E. Steinhagen, and S. B. Linek. 2016. "Researchers' use of Web 2.0 versus conventional media: perceived advantages and disadvantages". In: *Proceedings of the 8th annual International Conference on Education and New Learning Technologies*. Barcelona. 5855–5863.
- Park, N., K. F. Kee, and S. Valenzuela. 2009. "Being immersed in social networking environment: Facebook Groups, Uses and gratifications, and social outcomes". *Cyberpsychology & Behavior*. 12(6): 729–733.
- Peluchette, J. V. E., K. Karl, and J. Fertig. 2013. "A Facebook "friend" request from the boss: Too close for comfort?" *Business Horizons*. 56: 291–300.
- Priem, J. and K. L. Costello. 2010. "How and why scholars cite on Twitter". In: *Proceedings of the 73rd Annual Meeting on Navigating Streams in an Information Ecosystem*. Vol. 47. article 75. Silver Springs: American Society for Information Science.
- Priem, J. and B. Hemminger. 2010. "Scientometrics 2.0: New metrics of scholarly impact on the social web". *First Monday*. 15(7).
- Procter, R., R. Williams, J. Stewart, M. Poschen, H. Snee, A. Voss, and M. Asgari-Targhi. 2010. "Adoption and use of Web 2.0 in scholarly communications". *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*. 368(1926): 4039–4056.
- Pscheida, D., S. Albrecht, S. Herbst, C. Minet, and T. Köhler. 2013. "Nutzung von Social Media und onlinebasierten Anwendungen in der Wissenschaft". *Tech. rep.* Dresden: Technische Universität Dresden.
- Ross, C., M. T. and Warwick, and A. Welsh. 2011. "Enabled Backchannel: Conference Twitter Use by Digital Humanists". *Documentation*. 67(2): 214–237.
- Ruggiero, T. 2000. "Uses and gratifications theory in 21st century". *Mass Communication & Society*. 3(1): 3–37.
- Shao, G. 2009. "Understanding the appeal of user-generated media: a uses and gratification perspective". *Internet Research*. 19: 7–25.
- Shema, H., J. Bar-Ilan, and M. Thelwall. 2012. "Research blogs and the discussion of scholarly information". *PLoS ONE*. 7(5).
- Spinks, N., B. Wells, and M. Meche. 1999. "Netiquette: a behavioral guide to electronic business communication". *Corporate Communications*. 4(3): 145–155.
- Stafford, T., M. Stafford, and L. Schkade. 2004. "Determining uses and gratifications for the internet". *Decision Sciences*. 35(2): 259–288.
- Tang, J., T. Lou, and J. Kleinberg. 2012. "Inferring social ties across heterogeneous networks". In: *Proceedings of the 5th International Conference on Web Search and Data Mining*. New York: ACM. 743–752.
- Wang, C., J. Han, Y. Jia, J. Tang, D. Zhang, Y. Yu, and J. Guo. 2010. "Mining advisor-advisee relationships from research publication networks". In: *Proceedings of the 16th SIGKDD International Conference on Knowledge Discovery and Data Mining*. New York: ACM. 203–212.
- Weller, K., E. Dröge, and C. Puschmann. 2011. "Citation analysis in Twitter: Approaches for defining and measuring information flows within tweets during scientific conferences". In: *Proceedings of the ESWC Workshop on "Making Sense of Microposts"*. 1–12.
- Wen, X., Y.-R. Lin, C. Trattner, and D. Parra. 2014. "Twitter in Academic Conferences: Usage, Networking and Participation over Time". In: *Proceedings of the 25th Conference on Hypertext and Social Media*. New York: ACM. 285–290.
- Xu, B., Y. Huang, H. Kwak, and N. Contractor. 2013. "Structures of broken ties: exploring unfollow behavior on Twitter". In: *Proceedings of the Conference on Computer Supported Cooperative Work*. New York: ACM. 871–876.
- Zhang, J., C. Wang, and J. Wang. 2014. "Who proposed the relationship?: recovering the hidden directions of undirected social networks". In: *Proceedings of the 23rd International Conference on the World Wide Web*. New York: ACM. 807–818.