

Chapter 4

Learning in the Wild: Understanding Networked Ties in Reddit



Marc Esteve Del Valle, Anatoliy Gruzd, Priya Kumar, and Sarah Gilbert

4.1 Introduction

As social media platforms become more prevalent in our society, more educators are incorporating these platforms into their online and face-to-face classrooms. Some examples of social media use for teaching and learning include using Twitter as a backchannel to support synchronous discussions, blogging about current events related to a class or using Facebook to host asynchronous class discussions (Absar et al. 2015, Esteve Del Valle et al. 2017, Gruzd et al. 2018, Gruzd et al. 2016, Paulin et al. 2015). Previous research also found that social media use was beneficial to students' class engagement and academic performance (Collins and Gruzd 2017, Denker et al. 2018, Junco et al. 2013, Tang and Hew 2017). However, much of the research to date has focused primarily on providing insights about how social media platforms are used to support formal modes of learning. Few studies have examined the use of platforms like Twitter and Facebook to support informal learning outside the classroom (Gruzd et al. 2014, Gruzd and Conroy 2018). This chapter seeks to address this research gap by investigating what we refer to as 'learning in the wild'

M. E. Del Valle (✉)

Department of Media Studies, University of Groningen, Groningen, The Netherlands
e-mail: m.esteve.del.valle@rug.nl

A. Gruzd · P. Kumar

Social Media Lab, Ryerson University's Ted Rogers School of Management,
Toronto, ON, Canada

S. Gilbert

University of Maryland, College Park, MD, USA

© Springer Nature Switzerland AG 2020

N. Bonderup Dohn et al. (eds.), *Mobility, Data and Learner Agency in Networked Learning*, Research in Networked Learning,
https://doi.org/10.1007/978-3-030-36911-8_4

(following Hutchins' *Cognition in the Wild*, 1995), in which learning is not occurring in formal classes, guided by instructors, graded or structured around a syllabus. Instead, we are interested in understanding how learning is happening when social media users (regardless of their demographic characteristics, educational and professional background) turn to these platforms to pose and answer questions, comment, discuss, debate and argue. These new forms of learning echo Siemens' perspectives of connectivism (Siemens 2005), where social learning is integrated with information and communication technologies, and learning becomes a networked process. Such networked learning processes can be studied with the help of social network analysis (SNA), which 'provides a toolkit for exploring learning where connectivity is the major area of investigation' (Haythornthwaite et al. 2016, p. 253).

To study how learning is happening 'in the wild' (beyond formal classes), we explore interactions among users on Reddit, a popular social media site, in which we believe that informal learning is likely to happen. Specifically, we examine what structural configurations of resulting communication networks (for instance, the prevalence of mutual ties) or individual attributes (for instance, being a moderator) may predict the formation of 'learning' ties among Reddit users (known as Redditors). Our research was guided by the following two questions:

- How are learning processes taking place in informal social media environments such as Reddit?
- How do network configurations and/or individuals' attributes affect access to and the ability to act in informal networked learning environments?

Considering that there are over a million¹ of communities formed on Reddit (known as subreddits), we decided to adopt a case study approach by focusing on two communities known for their educational content: AskStatistics and AskSocialScience. While both subreddits are designed to support a question-answer type community, they are different in terms of the domain that each covers (statistics versus social science), as well as the number of members and moderators they have: AskStatistics has 13.1 k members and one moderator, as compared to 81 k users who joined AskSocialScience and 15 moderators (as of February 2019). Considering a small sample size of the studied population (two cases), the goal of this work is not to come up with generalizable results but rather to identify an initial set of factors that influence how learning ties are formed and maintained among Reddit users so that future work in this area can apply and validate our results in other communities and platforms.

In the next section, we conceptualise 'learning in the wild', a novel notion that is at the core of this research. We continue by reviewing studies that have used SNA to examine learning, followed by an explanation of Reddit. We then provide details on the data and methods used to answer our broad research questions. Last, we outline the results and discuss the factors explaining the formation of ties in both AskStatistics and AskSocialScience subreddit communities.

¹According to the Reddit metrics website: <http://redditmetrics.com/history/month>

4.2 Learning in the Wild

Internet technologies have broadened learning opportunities among their users by giving rise to the emergence of *networked learning communities*. In these communities, learners connect with others using ‘knowledge that is collaboratively constructed through their dialogues and social interactions’ (De Laat 2006, p. 123). At the core of our research are networked learning communities that are formed on social media. Such communities offer virtual spaces for open discussions where anyone can join and contribute their ideas and opinions, find and share relevant resources and connect with experts. Because of the open nature of these spaces, understanding the conduct of community members is important to comprehend how these networked communities operate. Rules and norms emerge from members’ interactions, with new users able to see and imitate observed practices. Subsequently, rules and norms determine not only what topics are appropriate for discussion but also what language and discourse practices should be used by community members. While community members are often able to flag or downvote a particular content that they find inappropriate, the enforcement of the rules usually depends on a limited number of users (e.g. moderators). And while joining such communities may be easy at first due to their open nature, it is much more difficult to stay and be active because rules and norms are in constant development and sometimes conflict with broader contexts. For example, Gilbert (2018) found that Reddit-wide norms of minimal moderation were problematic for members of a subreddit with strict rules and enforcement, largely because users who did not know or value the subreddit’s rules engaged in disruptive transgressions and rule-breaking behaviour. Successful integration into the community was often achieved through trial and error (during which users’ rule-breaking content would be removed) or prolonged passive participation while learning rules and norms.

In this chapter, we use the phrase ‘learning in the wild’ to explain interactions through social media and the key social and informal learning processes that lead to the emergence of relatively stable, *networked learning communities*. We view ‘learning in the wild’ as a form of social learning, which occurs through observation of and reaction to how others behave and interact (Bandura 1977). For example, as legitimate peripheral participants (Lave and Wenger 1991), new users learn and appropriate behaviours in keeping with group norms. Similar learning processes occur in social media sites, as individuals lurk before posting and as they observe others responding to and addressing inappropriate behaviour (Haythornthwaite and Andrews 2011). In these learning environments, social learning occurs through discussion. Online posting and reactions provide the material for learning about codes of conduct and community practices. Previous research has tried to unfold the learning occurring ‘in the wild’ by studying discursive practices among learners (Gunawardena et al. 1997, Chen and Resendes 2014) and by analysing the characteristics of their interactions (Gruzd and Haythornthwaite 2013, Schreurs and de Laat 2014). In general, results show that conversations among members of networked learning communities forge a web of social ties that contribute to both individual and group learning (Haythornthwaite 2011, Kumar et al. 2018,

Kumar and Gruzd 2019). Our research team has been working for several years on studying the practices of learning online (see, for instance, Esteve Del Valle et al. 2017 and Gruzd et al. 2018) by observing and researching the trends towards more learner-centred participation. We have developed a coding scheme to assess learning practices on social media (Haythornthwaite et al. 2018) and new models to understand the factors explaining networked interactions between learners (Esteve Del Valle et al. 2018). In doing so, our aim has been to understand learning processes in the social media age to suggest ways of improving and supporting current learning practices.

In addition to being *social*, learning ‘in the wild’ is also *informal*. Learning processes among social media users (usually) do not take place in institutionalised contexts (e.g. course units) and lead to formal certifications. In these platforms, users gain knowledge through their daily interactions and exposing themselves to the opinions of other users. As a consequence, learning becomes an unregulated, incidental and experiential process. Livingstone’s (1999) definition of informal learning helps us conceptualise the learning occurring ‘in the wild’ as the following:

Any activity involving the pursuit of understanding, knowledge or skill which occurs outside the curricula of educational institutions, or the courses or workshops offered by educational or social agencies. The basic terms of informal learning (e.g., objectives, content, means and processes of acquisition, duration, evaluation of outcomes, applications) are determined by the individuals and groups that choose to engage in it. Informal learning is undertaken on one’s own; either individually or collectively, without either externally imposed criteria or the presence of an institutionally authorized instructor. (p. 2)

In sum, learning activities associated with social media share both characteristics; they are social and informal. Users of social media can, for instance, post a question, and their peers can ignore this request or respond to the learning need. An answer to the user’s question will give rise to a networked tie that can be analysed through the lens of a social network analysis approach, as shown in the following section.

4.3 Social Network Analysis

Social network analysis (SNA) provides our study with theoretical lenses and measures for exploring collaborative learning activities in social media. The core concepts of SNA (such as *nodes*, *relations*, *ties* and *networks*) can be used to describe and study online learning processes in communities and wider networks (e.g. Rainie and Wellman 2012, Haythornthwaite 2014). Specifically, SNA can be employed to (a) develop interventions informing teachers (in their guiding role of the networked learning processes) and students about their social learning activities, (b) discover factors explaining the formation of online social learning activities, (c) predict learning outcomes, and (d) understand the nature and meaning of learning ties (Haythornthwaite et al. 2016). Below, we provide some examples of how SNA has been used to study learning processes occurring in online environments that are

especially relevant to the current study. The review below is not meant to be comprehensive. It is used as a starting point to demonstrate a variety of perspectives and questions that can be investigated in this area using SNA.

4.3.1 Network Visualisation and Data Exploration

In social learning analytics, researchers have designed interventions aimed to inform teachers and students about their online activities, such as experimenting with tools that visualise social learning activities automatically (Bakharia and Dawson 2011). An example of these tools is the Network Awareness Tool (NAT), designed by Schreurs and de Laat (2014). The tool aims to promote learner-centric reflection (e.g. how individuals use their peers for learning) and helps find peers who are engaging with the same learning topics online. Used as a plugin for online learning platforms, NAT visualises networked interaction (both actors and ideas) by identifying relations between people who interact around similar topics. In a related work, Comber, Durier-Copp and Gruzd et al. (2018) used network visualisations as a learning analytics tool to provide insights about student interactions in class-wide forum discussions. They confirmed that network visualisations are capable of ‘making the “invisible” visible to instructors’ by helping them to see who is engaged in online discussions and how. In our study, we experiment with Gephi (Bastian et al. 2009), a popular program for network visualisations, and Netlytic (Gilbert 2016), an SNA-based tool designed for the collection, analysis and visualisation of publicly available social media posts. We use Gephi and Netlytic to visualise and examine public interactions among members of the AskStatistics and AskSocialScience communities as networks.

4.3.2 Prediction

SNA has also been used to predict learning outcomes, such as discovering associations between students’ positions in a network and forecasting their successes in learning processes. For instance, Russo and Koesten (2005) found that prestige and degree centrality measures (i.e. the degree to which students are connected and engaged with others in the network) had a positive effect on classroom learning outcomes. Additionally, Cho et al. (2007) showed a significant association between students’ closeness centrality (a type of centrality measure that emphasises how easy is to reach a particular person based on their position in the network) and their final grades. While we do not study learning outcomes directly (in part due to the challenges of operationalising ‘learning in the wild’), we build on these studies by relying on network centrality measures to see why certain users are more central than others and if their position in the network affects who they interact with and how often they interact with them.

4.3.3 *Nature and Formation of Learning Ties*

Our work is especially close to the area of research that studies the nature of learning ties and how they are formed. In a related study, de Laat (2006) explored the gaps between social network data and learning processes using a multi-method approach that collected information on learning processes (who learned from whom?) and on the relational content of the learning ties (what were they talking about) and on combination-facilitated learning (why were they talking in such a way or in another?). Posing these questions allowed de Laat to triangulate data and explore learning processes by considering all relational aspects between networks and learning. In another study, Aviv et al. (2003) examined the process of knowledge creation in a formal and asynchronous online learning network (comprised of 18 participants) and in a more informal and asynchronous online learning network (comprised of 19 participants). The researchers found that the knowledge construction process in the formal and asynchronous online learning network reached a high phase of critical thinking, while in the asynchronous informal online learning network, the knowledge construction process reached a low phase of cognitive activity. The results of these studies serve as a guide to interpret the outcome of our investigation, specifically those of Aviv et al. (2003), concerning informal learning processes in online environments.

Research in networked learning has also aimed to discover the variables predicting the formation of learning ties. These variables can be based on individual or network characteristics. Individual attributes can include personal characteristics (e.g. age), and network characteristics can refer to one's position in the network (such as a centrality measure). Despite the relevance of this type of research, few studies have incorporated both individual and network characteristics to analyse learning processes in the online environment. A noteworthy example is Cho et al. (2007), who studied 31 learners working together to design an aerospace system using online collaboration tools. Their study showed that central individuals in the network remained connected to the same people over time, while individuals placed in the boundary of the network were more proactive in forming new learning ties with others.

Gilbert and Paulin (2015) used SNA to explore the role of experts, referred to as more knowledgeable others (MKOs), in conference Twitter networks. Using the (LAK) conference as a case study, the authors identified two types of MKOs: subject-specific MKOs who were involved in relevant professional organisations and past conferences and other MKOs with a relatively high h-index score of 23 or above (i.e. highly cited authors²). Both types of MKOs were found to have significantly higher levels of centrality and prestige than those who were not MKOs, suggesting that they are prominent members of the conference Twitter community and

²H-index is used to quantify the impact of someone's publication record by indicating how many of their publications are cited a particular number of times. For example, an author with the h-index of 23 has at least 23 publications that have been cited at least 23 times.

thus occupy positions in the network that allow them to make impactful contributions.

The results of the two studies described above (Cho et al. 2007, Gilbert and Paulin 2015) have relevant implications for learning purposes since they show that learners' attributes (such as being a subject expert) and network properties (such as one's position in the network) are important to consider when studying or designing networked learning activities. Indeed, we have considered the results of these studies when choosing the statistical models to examine Redditors' interactions in the AskStatistics and AskSocialScience subreddit communities. As discussed in the Methods section below, we chose to use exponential random graph models (ERGMs), a statistical approach capable of considering both individual and network characteristics, when studying tie formation. Building on the previous literature related to networked learning analysis, we seek to expand the understanding of what variables predict the formation of learning ties in online environments. The next sections provide more details on the social media site chosen for this study and our methods.

4.4 Reddit

Reddit is an online social content aggregation site that is commonly referred to as 'the front page of the Internet' for the way crowd-based voting raises the profile of user-submitted news or other items to a front-page equivalent. By its own account, 'Reddit bridges communities and individuals with ideas, latest digital trends and breaking news' (Reddit 2017). It has become increasingly popular since its launch in 2005, and it currently ranks 18th in terms of global traffic and sixth in the United States (Alexa 2019).

The basic framework of the Reddit system revolves around (a) subreddit communities, (b) posts, (c) comments, (d) votes, and (e) Karma (a popularity score earned by posting content that other users find engaging). Reddit is composed of millions of user-generated and user-moderated online communities on a wide range of topics (e.g. politics, economics, academia, etc.), which are called subreddits. Subreddits have their own norms and rules determining, for instance, what can and cannot be posted. Any registered user (Redditor) can create, comment and vote on posts. Comments are hierarchically threaded (root comment and subsequent comments) and can be in response to a general post (root comment) or in reply to another comment.

Redditors can upvote or downvote others' posts and comments. By default, posts and comments are displayed on the site according to the total vote ranking function; i.e. upvoted posts and comments rise to the top, while downvoted posts and comments are pushed to the bottom. Votes on Redditors' posts and comments contribute to their Karma score; posts and comments that are upvoted increase Redditors' Karma scores, while posts and comments that are downvoted detract from total Karma.

Redditors can also become Gold members³ and moderators. Gold membership grants access to extra features such as reading more comments per page or joining a private subreddit only available to those with Reddit Gold, among others. Holding a moderator role gives Redditors a range of controls for configuring the subreddits they moderate, for instance editing the rule page of the subreddit or banning specific users from participating in the subreddit.

Reddit is an ideal environment for our investigation of learning ‘in the wild’ because conversations emerge from the contribution and promotion of the members, combining perspectives of experts and non-experts (Gilbert 2018) outside traditional classroom settings. By analysing publicly available discussions on Reddit, we evaluate how the network configurations and Redditor’s individual attributes may influence the networked learning processes on this site.

4.5 Methods

We collected all posts and comments submitted to the AskStatistics and AskSocialScience subreddits in 2015. As part of this process, we also collected information about the (a) Karma points, (b) the Gold membership status (being or not being a Gold member) and (c) the moderator role (being or not being a moderator) of the Redditors of these two communities.

Data were collected using a custom application that relied on Reddit’s public application programming interface (API). In accordance with Canada’s Tri-Council Policy Statement on Ethical Conduct for Research Involving Humans (2014), review by a Research Ethics Board was not required as the research was non-intrusive and did not involve direct interaction between the researchers and subreddit members. The only data collected were publicly viewable discussions on the website. In addition, consent was not considered necessary to solicit from individual users or group moderators because of the lack of personally identifiable information (i.e. users on the platform do not use their real names) and low sensitivity of the discussion topics (Nissenbaum 2004). Finally, to further ensure the privacy of users whose data are included in our study, the results are presented in aggregate without identifying any particular username.

To discover the network characteristics and the Redditors’ attributes that were facilitating ties among the users of these two subreddits, we employed exponential random graph models (ERGM). Broadly speaking, ERGM are designed to test various network-based hypotheses by generating a large set of random networks, based on a chosen set of network configurations and node attributes, and comparing these networks to an observed network (Esteve Del Valle and Borge 2017, Gruzd and Tsyganova 2015, Lusher et al. 2013). In our case, we used ERGM to test whether certain network configurations (i.e. statistics in the ERGM terms) and node attri-

³After data collection, in August 2018, Reddit’s Gold Membership program was overhauled and renamed Reddit Premium.

butes can explain the formation of ties in the subreddit communication networks or whether ties in these networks are likely to occur by chance alone. To run ERGM, we used the ‘statnet’ package in R (Goodreau et al. 2008, Hunter et al. 2008).

When using ERGM, the process starts with building a null model (*net-edges*), which only accounts for the tie density of the observed network without considering any other predictors. Our subsequent model, Model 1, included three network parameters corresponding to reciprocity (*mutual*), which is a statistic counting the ratio of mutual replies (Goodreau et al. 2008); transitivity (*transitive*), which occurs whenever in a discussion thread User A replies to User B, User B replies to User C, then User C replies to User A; and, finally, popularity, which is based on the number of Redditors who replied to a user (*gwindegree*). The underlying idea of adding these network statistics to the model is to evaluate their effects in increasing or decreasing Redditors’ likelihood of establishing communication ties in both ‘Ask’ communities. Next, we expanded Model 1 by adding Redditors’ attributes that were available to the research team: ‘Gold Membership’, ‘Karma’ and ‘Moderator’. Specifically, Model 2 included Redditors’ ‘Gold Membership’ status (*nodefactor* [*‘Gold_Member’*]). Model 3 added Redditor’s Karma score as a popularity measure of the content that users shared in the subreddit (*nodecov* [*‘Karma’*]). Lastly, Model 4 added Redditors’ moderator role (*nodefactor* [*‘Moderator’*]) to indicate if they are a moderator in each studied community.

Finally, to determine the quality of the resulting model, randomly generated networks were compared to the observed networks by assessing the goodness of fit (Hunter et al. 2008, Li and Carriere 2013).

4.6 Results

4.6.1 Descriptive Network Statistics and Network Visualisations

Table 4.1 shows descriptive statistics of the AskStatistics and AskSocialScience networks. In the case of the AskStatistics subreddit, 1951 Redditors posted a total of 4301 replies, while for the AskSocialScience subreddit a total of 3689 Redditors posted 7723 replies. In both networks, the graph density is very low (0.001), meaning that only 0.1% of the total possible relations among the Redditors occur. This observation is also demonstrated by the fact that the average number of users a Redditor interacts with (average degree) is only 2.205 in AskStatistics and 2.094 in AskSocialScience. At the same time, the average path length (i.e., the average graph-distance between all pairs of nodes) is 4.409 for the AskStatistics subreddit and 5.232 for the AskSocialScience subreddit, indicating that the average distance between any pair of users is 4.4 steps in the former subreddit and 5.2 in the latter. Although the density in both networks is low, relatively short distances make it possible for Redditors to connect easily to others.

Table 4.1 Descriptive network statistics of the ‘AskStatistics’ and ‘AskSocialScience’ subreddits

	AskStatistics	AskSocialScience
P (number of posts)	2352	1523
N (number of nodes)	1951	3689
R (number of replies)	4301	7723
Graph density	0.001	0.001
Average path length	4.409	5.232
Average degree	2.205	2.094
Modularity	0.621	0.641

Another network statistic to examine is modularity (Gruzd et al. 2017). The modularity scores for both networks are 0.621 for the case of AskStatistics and 0.641 for the case of AskSocialScience. These scores indicate the existence of clusters, which may be formed around certain topics or threads of conversation. Notably, although we observe different clusters of conversations, the fact that the modularity scores are not closer to 1 (the maximum possible value for this metric) suggests that these clusters are interconnected, potentially through a core group of users who contributed to multiple different topics discussed by their group.

Lastly, for each subreddit, we built a network representing who replies to whom based on the collected posts and comments, where nodes are Reddit users (Redditors) and directed edges in the network represent their communication patterns. The network visualisation step allowed us to confirm the findings based on the descriptive statistics visually, as well as to continue our exploratory analysis of emerging communication networks in both communities. Figure 4.1 shows the visualisation of the networks. Node colours are assigned automatically to indicate highly connected nodes based on a community detection algorithm. The colour of the nodes indicates the existence of different clusters. The size of the nodes is equivalent to the degree centrality of the Redditors in the network. The ties among the nodes represent the replies between the Redditors (the thicker the line of the ties is, the stronger is the intensity of the relations – number of replies – among the Redditors).

4.6.2 Resulting ERG Models

Tables 4.2 and 4.3 summarise our results of running Model 1–Model 4. The selection criteria were driven by significance levels of the tested parameters and the iterative reduction in both Akaike information criterion (AIC) and the Bayesian information criterion (BIC) values, with smaller values indicating a better fitting model (Goodreau et al. 2008).

The last column of the two tables reports the estimates of Model 4, which includes all the variables of the analysis. In both tables, the edge parameter is negative, a common characteristic of sparse networks (Mai et al. 2015). The estimates suggest that reciprocity and transitivity remain positive and significant across all models, whereas popularity remains significant but negative. This means that reciprocity and transitivity increase Redditors’ likelihood of establishing networked ties, whereas popularity decreases their likelihood of forming new ties.

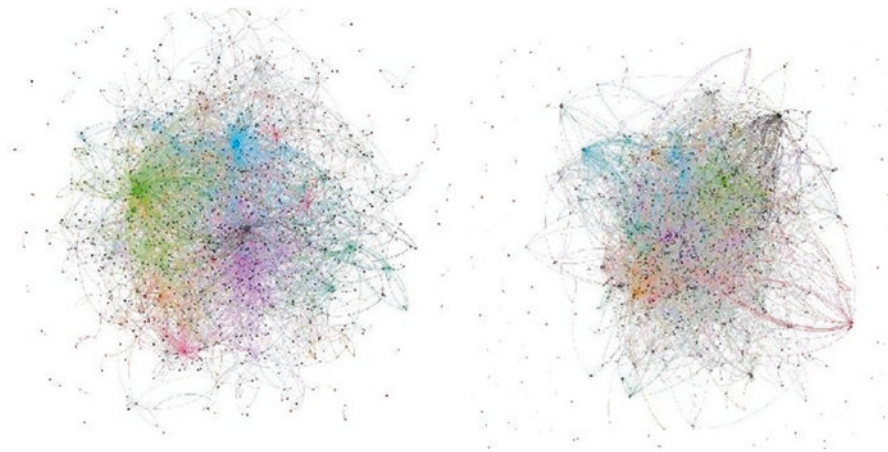


Fig. 4.1 Communication networks among Redditors in AskStatistics (on the left) and AskSocialScience (on the right)

Table 4.2 Factors underlying the formation of ties in the AskStatistics subreddit

	Model 1		Model 2		Model 3		Model 4	
	<i>EST</i>	<i>SE</i>	<i>EST</i>	<i>SE</i>	<i>EST</i>	<i>SE</i>	<i>EST</i>	<i>SE</i>
<i>Structural features</i>								
Edges	-7.631	0.021	-7.682	0.026	-7.318	8.297	-7.348	8.111
Reciprocity	6.523	0.039	8.096	0.034	6.826	3.310	6.789	3.427
Popularity	-0.751	0.022	-0.531	0.024	-1.064	1.624	-1.151	1.362
Transitivity	0.630	0.000	0.666	0.000	6.652	1.168	6.137	3.383
<i>Redditors' attributes</i>								
Gold member			-0.225	0.049	-2.340	1.552	3.939	9.670
Karma					3.341	1.077	-1.027	3.658
Moderator							9.343	9.779
<i>Akaike information criterion (AIC)</i>	58,577		58,406		58,427		58,294	
<i>Bayesian information criterion (BIC)</i>	58,630		58,471		58,505		58,386	

Note: Coefficients in bold are significant at the 99% level

Results from Model 4 also show the effects of Redditors' attributes in facilitating networked ties. The estimates of the 'Gold Membership' status are positive and significant for AskStatistics but negative and significant for AskSocialScience. This contradictory finding makes it difficult to reach any definitive conclusion on the effects of being a 'Gold Member' in increasing or decreasing the likelihood of replying or receiving a reply from another user. The estimates of Redditors' Karma scores are negative and significant for AskStatistics, but they are not significant for AskSocialScience. Again, these results do not allow us to draw conclusions about the effects of the Karma scores on forming ties on Reddit. Lastly, the estimates of the 'Moderator' attribute are significant and positive for both AskStatistics and

Table 4.3 Factors underlying the formation of learning ties in AskSocialScience subreddit

	Model 1		Model 2		Model 3		Model 4	
	<i>EST</i>	<i>SE</i>	<i>EST</i>	<i>SE</i>	<i>EST</i>	<i>SE</i>	<i>EST</i>	<i>SE</i>
<i>Structural features</i>								
Edges	-7.040	0.023	-7.102	0.024	-7.107	5.283	-7.174	5.830
Reciprocity	8.277	0.105	8.056	0.131	8.211	4.111	8.041	1.031
Popularity	-2.521	0.050	-2.398	0.067	-2.447	8.392	-2.375	8.604
Transitivity	0.377	0.040	1.003	0.003	3.799	1.521	3.940	1.644
<i>Redditors' attributes</i>								
Gold member			-0.075	0.028	4.914	5.300	-1.544	5.207
Karma					1.296	4.013	4.982	3.966
Moderator							9.903	4.430
<i>Akaike information criterion (AIC)</i>	120,183		120,022		120,127		119,680	
<i>Bayesian information criterion (BIC)</i>	120,241		120,094		120,213		119,781	

Note: Coefficients in bold are significant at the 99% level

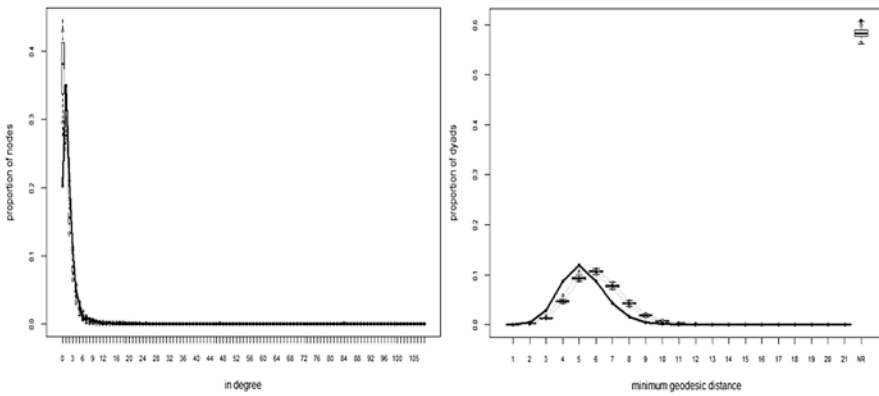


Fig. 4.2 Model 4 goodness-of-fit diagnostics for AskStatistics

AskSocialScience. This means that being a moderator highly increases the likelihood of establishing a tie in both networks.

To assess how well the final model, Model 4, captures the structure of the observed data, Figs. 4.2 and 4.3 show how the observed in-degree and minimum geodesic distance distributions reproduce the network’s statistics observed in the original data. This step allows us to see how similar networks simulated based on the final model (Model 4) are to the observed networks based on the two network statistics (in-degree and minimum geodesic distance), which have not been explicitly included in the tested model.

In the plots, the vertical axis is the relative frequency of nodes (in-degree) and dyads (minimum geodesic distance). The observed statistics in the actual network are indicated by the solid lines (thick black lines). The grey dotted lines represent

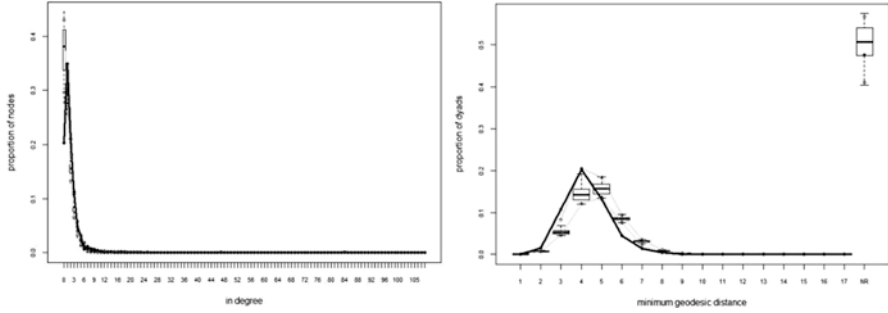


Fig. 4.3 Model 4 goodness-of-fit diagnostics for AskSocialScience

the range of 95% of the simulated statistics. The models perform relatively well for the in-degree and geodesic distance distributions. The observed distributions generally fall within the quantile curves for most of the range. The model overestimates the average in-degree distribution and geodesic distance, but overall it captures the shape of the distributions.

4.7 Discussion

This study sought to expand our current knowledge of learning processes in informal social media environments by discovering what factors may predict tie formation in two subreddits: AskStatistics and AskSocialScience. Using SNA, we analysed one year of data on all communication-related relations (based on posts and comments) and three Redditor attributes (Gold membership, Karma scores and moderator status). Our review of some commonly used descriptive network statistics revealed a low connectivity among Redditors of these two communities. This is because of relatively low values of the graph density (0.001 for both networks) and the average number of connected Redditors (2.2 for AskStatistics and 2.1 for AskSocialScience). However, the results of the average path length (4.409 for AskStatistics and 5.232 for AskSocialScience) show that despite the low connectivity in both networks, shorter distances between Redditors make it possible for them to connect and share information with one other efficiently (see also Esteve Del Valle and Borge 2017).

Based on the results of the ERG models, we found that in both subreddits, the likelihood of establishing networked learning ties greatly increases with Redditors' reciprocal posting behaviour (i.e. when User A and User B reply to each other) and increases with the existence of transitive replying behaviours (i.e. when User A replies to User B, User B replies to User C and User C completes the cycle by replying to User A). A possible explanation of this transitive replying behaviour among Redditors is the existence of a clustering effect facilitating interactions between users of the same conversational threads. Our findings also show that the likelihood

of establishing a tie decreases when users' posts are very popular (i.e. received a lot of replies). This suggests that there may be an upper limit on the number of replies and connections that a user can get on Reddit, which may be due to the platform's interface affordances or other factors requiring further research. We see three inter-related explanations for this result: (1) because Reddit collapses comments as threads get larger, users who are quickly scrolling through comments may not click on collapsed comments to view, read and respond to them; (2) reading popular posts with many comments may cause fatigue, and thus others are more likely to respond to comments displayed at the top of the page rather than at the bottom; (3) knowing that comments visible at the top of the page are more likely to be read, users may be more inclined to respond to top-level comments as a way to increase their Karma scores.

Finally, at the individual level, contradictory results concerning the Redditors' Gold membership status and their Karma scores do not allow us to draw clear conclusions on the effects of these two attributes in establishing communication ties. It is possible that these individual characteristics and their role in tie formation are subreddit specific. Gold membership status could be indicative of an active Reddit user since Gold members must either have purchased this status for themselves or have been gifted it by another user. But it appears not to be the case in at least one of the studied subreddits.

Our results regarding Karma scores also differ from conclusions drawn by Kilgo et al. (2016), who suggested that Karma scores may be used to identify opinion leaders (i.e. highly connected individuals). This may be explained by two factors. First, as relatively small subreddits, the primary form of recognition among users may be qualitative, such as giving thanks when answers are provided, rather than quantitative, such as upvoting and adding to Karma; this pattern of recognition has been noted as a key characteristic of online communities, particularly in those with a core intent of knowledge exchange, such as academic communities (Haythornthwaite 2009). Second, Karma scores are derived from participation across Reddit, rather than through participation in individual subreddits. Thus, a high Karma score may not be reflective of users' expertise in a given community. Hence, future research should examine more cases to clarify the role of these two attributes in establishing networked ties among Redditors.

While we found some conflicting results regarding the effects of Gold membership and Karma on the formation of communicative ties (and what we believe to be *learning* ties as well, considering the educational focus of both communities), we did find evidence to support the claim that being a moderator increases the likelihood of establishing ties in both subreddits. Unlike Gold membership, which is available to any user, being a moderator is likely indicative of more active and regular participation in a subreddit. In Reddit's topic-based 'Ask' subreddits, moderators are also often subject experts and thus also contribute to the community by responding to questions, thereby establishing learning ties. In 2015, AskSocialScience had 13 moderators, eight of whom were noted as experts in a specific topic area—'flaired' in Reddit terms. In AskSocialScience, a 'flair' is granted when a Redditor provides evidence that they have a university degree in the area and/or has a proven record of contributing high-quality social science comments. While AskStatistics

has no flair system, the subreddit has only one moderator, who is also the subreddit's creator, suggesting that the moderator has an interest in providing users with responses to statistics questions.

4.8 Conclusions

This chapter contributes to the understanding of learning 'in the wild'; that is, learning taking place on social media, beyond institutionalised curricula and formal classes. We chose Reddit as the platform for our investigation, and more specifically the subreddits AskStatistics and AskSocialScience, because their educational nature suggested that we would observe informal learning processes.

The main goal of our study was to predict the factors explaining interactions among the users (Redditors) of the AskStatistics and AskSocialScience *networked learning communities* (De Laat 2006). To do so, we used ERGM to examine interactions among the users of these two communities for a period of one year (2015). The results show that relations among users of these two communities are determined by both network configurations (reciprocity and transitivity) and individual attributes (being a moderator). This means that for the users of these communities, the likelihood of establishing learning ties increases if a user maintains reciprocal interactions with other users, participates in multiple conversational clusters or is a moderator.

From a methodological point of view, our study demonstrates the usefulness of applying SNA-based concepts and measures to make sense of learning processes occurring 'in the wild'. We hope that our research will inspire other colleagues to study on learning occurring across various social media sites and not just Reddit. From a practical perspective, we expect teachers to bridge the gap between formal and informal learning by using social media for educational purposes. More broadly, we expect educational institutions to make use of the networked analytics derived from learning interactions on social media to better understand today's ever-changing learning tools and strategies that often combine formal and informal elements and are based on both self- and collaborative learning.

Acknowledgement This work was supported by the Social Sciences and Humanities Research Council of Canada (SSHRC) under Insight Grant Scheme (PI: Gruzd).

References

- R. Absar, A. Gruzd, C. Haythornthwaite, & D. Paulin, Media Multiplexity in Connectivist MOOCs. in *Proceedings of the Fifth International Conference on Learning Analytics and Knowledge* (ACM, New York, NY, USA, 2015), pp. 424–425. <https://doi.org/10.1145/2723576.2723654>
- Alexa.,Traffic, Demographics and Competitors – Alexa. Retrieved December 23, 2019, from <https://www.alexacom/siteinfo/reddit.com>

- R. Aviv, E. Zippy, G. Ravid, & A. Geva, Network analysis of knowledge construction in asynchronous learning networks. *Journal of Asynchronous Learning*, 7(3), 1–23 (2003).
- A. Bakharia, S. Dawson, SNAPP: A Bird's-eye View of Temporal Participant Interaction. in *Proceedings of the 1st International Conference on Learning Analytics and Knowledge* (ACM, New York, NY, USA, 2011), pp. 168–173. <https://doi.org/10.1145/2090116.2090144>
- A. Bandura, *Social learning theory* (Prentice-hall series in social learning theory). Englewood Cliffs: Prentice-Hall. (1977).
- M. Bastian, S. Heymann, M. Jacomy, Gephi: an open source software for exploring and manipulating networks. In *Third international AAAI Conference on Weblogs and Social Media*, 2009
- B. Chen, M. Resendes, Uncovering what matters: Analyzing transitional relations among contribution types in knowledge-building discourse. *Proceedings of the 4th International Conference on Learning Analytics and Knowledge*, ACM, 2014. pp. 217–225
- H. Cho, G. Gay, B. Davidson, A. Ingraffea, Social networks, communication styles, and learning performance in a cscl community. *Comput. Educ.* 49(2), 309–329 (2007). <https://doi.org/10.1016/j.compedu.2005.07.003>
- R. Collins, A. Gruzd, Learning within Digital Media: Investigating the Relationships Between Student Citation Networks, Assignment Structures, and Learning Outcomes. in *Proceedings of the 50th Hawaii International Conference on System Sciences*, 2017. Available at <https://scholarspace.manoa.hawaii.edu/handle/10125/41407>
- S. Comber, M. Durier-Copp, A. Gruzd, Instructors' perceptions of networked learning and analytics. *Can. J. Lear. Technol.* 44(3) (2018). <https://doi.org/10.21432/cjlt27644>. Available at <https://www.learntechlib.org/p/188299/>
- M. De Laat, *Networked Learning* (phd), 2006., Retrieved from <https://eprints.soton.ac.uk/20358/>
- M. Esteve Del Valle, A. Gruzd, C. Haythornthwaite, S. Gilbert, D. Paulin, Social media in educational practice: faculty present and future use of social media in teaching. in *Proceedings of the 50th Hawaii International Conference on System Sciences*, 2017. pp. 164–173
- K.J. Denker, J. Manning, K.B. Heuett, M.E. Summers, Twitter in the classroom: modeling online communication attitudes and student motivations to connect. *Comput. Hum. Behav.* 79, 1–8 (2018)
- M. Esteve Del Valle, R. Borge, Leaders or brokers? Potential Influencers in online parliamentary networks. *Policy Internet* 10(1), 61–86 (2017)
- M. Esteve Del Valle, A. Gruzd, C. Haythornthwaite, P. Kumar, S. Gilbert, D. Paulin, Learning in the wild: Predicting the formation of ties in 'Ask' subreddit communities using ERG models. in *Proceedings of the 11th International Conference on Networked Learning*, IEEE, 2018. pp. 157–164
- S. Gilbert, Learning in a Twitter-based community of practice: an exploration of knowledge exchange as a motivation for participation in #hcsma. *Info. Commun. Soc.* 19(9), 1214–1232 (2016). <https://doi.org/10.1080/1369118X.2016.1186715>
- S. Gilbert, *Motivations for participating in online initiatives: exploring motivations across initiative types* (Doctoral dissertation). 2018. Retrieved from <https://open.library.ubc.ca/cIRcle/collections/ubctheses/24/items/1.0372890>
- S. Gilbert, D. Paulin, Tweet to learn: Expertise and centrality in conference Twitter networks. in *Proceedings of the 48th Hawaii International Conference on System Sciences*, IEEE, 2015. pp. 1920–1929
- S.M. Goodreau, M.S. Handcock, D.R. Hunter, C.T. Butts, M. Morris, A statnet Tutorial. *J. Stat. Softw.* 24(9), 1–27 (2008)
- A. Gruzd, N. Conroy, Designing a learning analytics dashboard for twitter-facilitated teaching. in *Proceedings of the Fifth Annual ACM Conference on Learning at Scale (L@S '18)*. ACM, New York, NY, USA, 2018. Article 46. <https://doi.org/10.1145/3231644.3231704>
- A. Gruzd, C. Haythornthwaite, Enabling community through social media. *J. Med. Internet Res.* 15(10), e248 (2013). <https://doi.org/10.2196/jmir.2796>
- A. Gruzd, K. Tsyganova, Information wars and online activism during the 2013/2014 crisis in Ukraine: examining the social structures of pro- and anti-maidan groups. *Policy Internet* 7(2), 121–158 (2015)

- A. Gruzd, C. Haythornthwaite, D. Paulin, R. Absar, M. Huggett, Learning analytics for the social media age. in *Proceedings of the Fourth International Conference on Learning Analytics And Knowledge (LAK'14)*. ACM, New York, NY, USA, 2014. pp. 254–256. <https://doi.org/10.1145/2567574.2576773>
- A. Gruzd, D. Paulin, C. Haythornthwaite, Analyzing social media and learning through content and social network analysis: a faceted methodological approach. *J. Learn. Anal.* **3**(3), 46–71 (2016)
- A. Gruzd, P. Mai, A. Kampen, A how-to for using Netlytic to collect and analyze social media data: a case study of the use of Twitter during the 2014 Euromaidan Revolution in Ukraine, in *The SAGE Handbook of Social Media Research Methods*, ed. by L. Sloan, A. Quan-Haase, (SAGE, Thousand Oaks, 2017)
- A. Gruzd, C. Haythornthwaite, S. Gilbert, D. Paulin, M. Esteve Del Valle, Users and gratification factors for social media use in teaching: Instructors' perspectives. *New Media Soc.* **20**(2), 475–494 (2018)
- C.N. Gunawardena, C. Lowe, T. Anderson, Analysis of a global online debate and the development of an interaction analysis model for examining social construction of knowledge in computer conferencing. *J. Educ. Comput. Res.* **17**(4), 397–431 (1997)
- C. Haythornthwaite, Crowds and communities: light and heavyweight models of peer production. in *Proceedings of the 42nd Hawaii International Conference on System Science*, 2009. 1–10. <https://doi.org/10.1109/HICSS.2009.137>
- C. Haythornthwaite, Learning networks, crowds and communities. in *Proceedings of the Learning Analytics and Knowledge Conference*, Banff, Alberta, Canada, February, 2011. pp.18–22.
- C. Haythornthwaite, Learning networks, in *Encyclopedia of Social Network Analysis and Mining*, ed. by R. Alhaji, J. Rokne, (Springer, New York, 2014), pp. 785–793
- C. Haythornthwaite, R. Andrews, *E-Learning Theory and Practice* (SAGE, London, 2011)
- C. Haythornthwaite, M. de Laat, B. Schreurs, A social network analytic perspective on E-learning, in *The SAGE Handbook of E-learning Research*, (SAGE Publications Ltd., Thousand Oaks, 2016), pp. 251–269
- C. Haythornthwaite, P. Kumar, A. Gruzd, S. Gilbert, M. Esteve Del Valle, D. Paulin, Learning in the wild: coding for learning and practice on Reddit. *Learn. Media Technol.* **43**(3), 219–235 (2018)
- D.R. Hunter, M.S. Handcock, C.T. Butts, S.M. Goodreau, M. Morris, Ergm: a package to fit, simulate and diagnose exponential-family models for networks. *J. Stat. Softw.* **24**(3), nihpa54860 (2008)
- E. Hutchins, *Cognition in the Wild* (MIT press, Cambridge, 1995)
- R. Junco, C.M. Elavsky, G. Heiberger, Putting twitter to the test: assessing outcomes for student collaboration, engagement and success. *Br. J. Educ. Technol.* **44**(2), 273–287 (2013)
- D.K. Kilgo, J.J. Yoo, V. Sinta, S. Geise, M. Suran, T.J. Johnson, Led it on Reddit: An exploratory study examining opinion leadership on Reddit. *First Monday* **21**(9) (2016). <https://doi.org/10.5210/fm.v21i9.6429>
- P. Kumar, A. Gruzd, Social Media for Informal Learning: A Case of #Twitterstorians. in *Proceedings of the 52nd Hawaii International Conference on System Sciences*, IEEE, 2019. pp. 2527–2535
- P. Kumar, A. Gruzd, C. Haythornthwaite, S. Gilbert, M. Esteve Del Valle, P. Drew, Learning in the wild: Coding Reddit for learning and practice. in *Proceedings of the 51st Hawaii International Conference on System Sciences*, IEEE, 2018. pp. 1933–1942
- J. Lave, E. Wenger, *Situated learning: legitimate peripheral participation* (Cambridge University Press, Cambridge, 1991)
- Y. Li, K.C. Carriere, Assessing goodness of fit of exponential random graph models. *Int. J. Stat. Probab.* **2**(4), 64 (2013)
- D. Livingstone, Exploring the icebergs of adult learning: Findings of the first Canadian survey of informal learning practices. Nall Working Paper No: 10. Centre for the Study of Education and Work, OISE/IT, Toronto, 1999. Retrieved from <https://tspace.library.utoronto.ca/retrieve/4451>
- D. Lusher, J. Koskinen, G. Robins, *Exponential Random Graph Models for Social Networks: Theory, Methods, and Applications* (Cambridge University Press, Cambridge, 2013)

- B. Mai, J. Liu, S. González-Bailón, Network effects in the academic market: mechanisms for hiring and placing PhDs in communication (2007–2014). *J. Commun.* **65**(3), 558–583 (2015)
- H. Nissenbaum, Privacy as contextual integrity. *Wash. Law Rev.* **79**, 39 (2004)
- D. Paulin, S. Gilbert, C.A. Haythornthwaite, A. Gruzd, R. Absar, Beyond the backchannel: Leveraging Twitter to enact learning processes. *iConference 2015 Proceedings*, 2015. URI: <http://hdl.handle.net/2142/73779>
- H. Rainie, B. Wellman, *Networked: The New Social Operating System* (The MIT Press, Cambridge, MA, 2012). Retrieved from <http://search.ebscohost.com.proxy-ub.rug.nl/login.aspx?direct=true&db=nlebk&AN=452133&site=ehost-live&scope=site>
- Reddit, Reddit, 2017, September 7. Retrieved September 7, 2017, from <https://about.reddit.com/>
- T.C. Russo, J. Koesten, Prestige, centrality, and learning: a social network analysis of an online class. *Commun. Educ.* **54**(3), 254–261 (2005). <https://doi.org/10.1080/03634520500356394>
- B. Schreurs, M. de Laat, The network awareness tool. *Comput. Hum. Behav.* **37**(C), 385–394 (2014)
- G. Siemens, Connectivism: a learning theory for the digital age. *Int. J. Instr. Technol. Distance Learn.* **2**(1), 3–10 (2005)
- Y. Tang, K.F. Hew, Using Twitter for education: Beneficial or simply a waste of time? *Comput. Educ.* **106**, 97–118 (2017)
- Tri-Council, Ethical conduct for research involving humans. Ottawa, Ont, 2014. Retrieved April 14, 2019 from http://www.pre.ethics.gc.ca/pdf/eng/tcps2-2014/TCPS_2_FINAL_Web.pdf