

COMPUTATIONAL SOCIAL ROLES

Diyi Yang

CMU-LTI-19-001

Language Technologies Institute
School of Computer Science
Carnegie Mellon University
5000 Forbes Ave., Pittsburgh, PA 15213
www.lti.cs.cmu.edu

Thesis Committee:

Robert E. Kraut, Co-Chair (Carnegie Mellon University)
Eduard Hovy, Co-Chair (Carnegie Mellon University)
Brandy L. Aven (Carnegie Mellon University)
Dan Jurafsky (Stanford University)

*Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Language and Information Technologies*

Copyright © 2019 Diyi Yang

Keywords: Computational social science, conversational acts, computer-supported cooperative work, generative model, human-computer interaction, machine learning, natural language processing, neural network, online health communities, recommender system, social network, social support, semantics, self-disclosure, social roles, social science, well-being, Wikipedia

Abstract

Millions of people participate in online communities, exchange expertise and ideas, and collaborate to produce complex artifacts. They often enact a variety of social roles in the process of helping their communities and the public at large, which strongly influence the amount and types of work they do, and how they coordinate their activities. Better understanding members' roles benefits members by clarifying how they should behave to participate effectively and also benefits the community overall by encouraging members to contribute in ways that best use their skills and interests.

Social sciences have provided rich theoretic taxonomies of social roles within groups, while natural language processing techniques enable us to automate the identification of social roles in online communities. However, most social science work has focused on generic roles without accommodating the activities associated with tasks in specific contexts or automating the process of role identification. While there has been work to date about automatic role inference, identification of social roles has not had a corresponding strong emphasis in the language technologies community. A variety of methods were developed to extract specific "roles" or patterns in different contexts, lacking generalized definitions about what are roles and systematic methods to extract roles. Moreover, how roles change over time and how the awareness of roles influence role holders' performance and the group production, have not been adequately researched in both fields.

This thesis advocates for both theories of social science and models of text analysis to better define roles, develop ways to extract roles and optimally recommend roles to users. Concretely, this work defines what are social roles, introduces five measurable facets associated with social roles, and proposed a generic methodology for role identification. It also demonstrated how to computationally model social role and its facets in two socially important contexts - Wikipedia and Cancer Survivor Network. Via combining theories of social roles and computational models for role identification on those two large-scale contexts, this research reveals details about emergent, behavioral, and functioning roles, and a set of computational techniques to identify such roles via fine-grained operationalization of role holders'

behaviors. This work fills the longstanding gap in role theory and empirical modeling about emergent roles in online communities, and lays the foundation for future work to identify and analyze roles that people enacted in group processes both online and offline.

Acknowledgments

I'm always grateful to my adviser Bob Kraut — the best adviser one could ever have. He is a great scholar and acts as a tremendous role model for me. Over the last several years we work together, I learned from him what is research and how to do rigorous research. I was trained in computer science, but Bob equipped me well with knowledge and mindset from social science. Although the journey of PhD is like taking a roller coaster, Bob provides as much support as I needed no matter there were ups or downs. These years of working with Bob is the most valuable part of my experiences at CMU.

I'm very lucky to have Eduard Hovy as my co-adviser. He always encourages me to go bigger, deeper and broader of my research direction. He is knowledgeable, visionary, and teaches me on how to think critically. Our advise meetings were not only about specific research projects, but also discussions around research visions, career plans, personality, leadership, etc. He believes in me, sets the right goals for me although I felt lost and not confident about myself most of the times. Without his encouragement and support, I would have still been the shy and timid myself.

I'm deeply honored to have Brandy Aven and Dan Jurafsky on my thesis committee. I benefit a lot from Brandy's expertise, who always push me to think of the big picture around my research. Dan is supportive, insightful, brilliant, and one of the great leaders in our field. I'm grateful that he hosted my three-month visit to Stanford NLP group. It was such a joy working with Dan and he provided me a lot of guidance and help over our meetings.

I thank LTI and HCI faculty Alan Black, Jamie Callan, Chris Dyer, Alex Hauptmann, Alon Lavie, Louis-Philippe Morency, Graham Neubig, Carolyn Rosé, Yulia Tsvetkov and Laura Dabbish for their advice and help to my research or job-seeking. A special thanks to the administrative staff: Stacey Young, Christy Melucci, Laura Alford, and JaRon Pitts. I would like to thank Aaron Halfaker from Wikimedia Foundation, Bin Gao, Tie-Yan Liu, Scott Counts from Microsoft Research, Ryan Ritter and Mark Handel from Facebook Research, for hosting my internships.

I'm extremely grateful for numerous friends and collaborators, including David Adamson, Waleed Ammar, Danqi Chen, Jiaxin Chen, Vivian Yung-Nung Chen, Zihang Dai, Sauvik Das, Pradeep Dasigi, Yulun Du, Manaal Faruqui, Fang Fei, Kartik Goyal, Liangke Gui, Anhong Guo, Hyeju Jang, He He, Ting-Hao Kenneth Huang, Junjie Hu, Liang He, Zhiting Hu, Dirk Hovy, Haojian Jin, Yohan Jo, Guokun Lai, Huiying Li, Mu Li, Jiwei Li, Bin Liu, Hanxiao Liu, Zhengzhong Liu, Zack Lipton, Anna Kasunic, Xuezhe Ma, Elijah Mayfield, Will Monroe, Felicia Ng, Joseph Seering, Chenxi Song, Lu Sun, Di Wang, Haohan Wang, Ling Wang, Xu Wang, William Yang Wang, Yi-Chia Wang, Yu-Xiang Wang, Fan Wei, Miaomiao Wen, Yuexin Wu, Pengtao Xie, Qizhe Xie, Zhang Yang, Zhilin Yang, Zi Yang, Zichao Yang, Zheng Yao, Pengcheng Yin, Zhou Yu, Hao Zhang, Shikun Zhang, Yang Zhang, Guoqing Zheng, Haiyi Zhu, and many others. I want to specifically thank Chenxi, Jiaxin, Joseph, Shikun, Xu, and Zheng for being a listening ear and helping me overcome stress and hard times in life and work.

I would like to thank my undergraduate adviser Yong Yu (and our ACM Honored Class), mentors Tianqi Chen and Weinan Zhang who started me on my first several research projects.

I thank my parents for raising me up, for their unconditional support, and for guiding me to where I am today. I would like to thank my boyfriend Peter for his patience, encouragement, trust, and love along the way.

Finally, I thank Facebook PhD Fellowship, CMU Presidential Fellowship and funds from Google and NIH for generously supporting this research.

Contents

- 1 Introduction 1**
 - 1.1 Thesis Overview 4
 - 1.1.1 Role Theory 4
 - 1.1.2 Methodology for Identifying Roles 4
 - 1.1.3 Case Studies of Role Identification 4
 - 1.2 Research Impact 6

- 2 Role Theory 9**
 - 2.1 Theoretical Modeling 10
 - 2.2 Computational Modeling 12
 - 2.3 Role Framework 14
 - 2.3.1 Agent 14
 - 2.3.2 Interaction 16
 - 2.3.3 Goal 17
 - 2.3.4 Expectation 18
 - 2.3.5 Context 20
 - 2.4 Relevant Processes about Roles 21
 - 2.5 Summary 23

- 3 Methodology for Identifying Roles 25**
 - 3.1 Generic Methodology 25
 - 3.1.1 Role Postulation 26
 - 3.1.2 Role Definition 27
 - 3.1.3 Role Identification 30
 - 3.1.4 Role Evaluation 34
 - 3.2 Iterative Process of Identification 36
 - 3.3 Reflection 39

I	Role Identification on Wikipedia	41
4	Identifying Roles of Editors	43
4.1	Introduction	44
4.2	Related Work	46
4.3	Research Question and Data	48
4.4	Predicting Edit Categories	48
4.4.1	Edit Categories Construction	49
4.4.2	Feature Space Design	50
4.5	Modeling Editor Roles	52
4.5.1	Role Identification Method	53
4.5.2	Derived Roles Exploration and Validation	54
4.6	Improving Article Quality	57
4.6.1	Model Design	58
4.6.2	Result Discussion	62
4.7	Discussion and Conclusion	63
4.8	Reflection	64
5	Identifying Semantic Edit Intention	65
5.1	Introduction	66
5.2	Related Work	67
5.3	Semantic Taxonomy of Edit Intentions	70
5.3.1	Taxonomy of Edit Intentions	70
5.3.2	Corpus Construction	71
5.3.3	Corpus Expansion	72
5.4	Identification of Edit Intentions	73
5.4.1	Identification Result	74
5.5	Intentions, Survival and Quality	76
5.5.1	How Edit Intentions Affect Survival	76
5.5.2	How Intentions Affect Article Quality	79
5.6	Discussion and Conclusion	80
5.7	Reflection	81

II	Role Identification on Cancer Survivors Network	83
6	Identifying Roles on CSN	85
6.1	Introduction	86
6.2	Research Question	88
6.3	Research Site	88
6.4	Generative Model for Role Identification	89
6.4.1	Operationalizing Behavioral Features	90
6.4.2	Determining the Granularity of User Activity	95
6.4.3	Determining the Number of Roles	99
6.5	Discovered Roles in Online Health Communities	100
6.6	Evaluating Roles	103
6.6.1	Recommender System with Roles	103
6.6.2	Deployment Studies on Recommendation	105
6.6.3	Survey on Roles	107
6.7	Influence of Emergent Roles on Commitment	115
6.8	Stability and Dynamics of Roles	117
6.8.1	Community Level Stability	117
6.8.2	Individual Level Dynamics	117
6.9	Discussion	123
6.9.1	Implication	124
6.9.2	Limitations	125
6.10	Reflection	125
III	Envoi	129
7	Conclusion	131
7.1	Summary of Findings & Contributions	132
7.2	Limitation	134
7.3	Future Work	136
7.3.1	Cross-community Role	136
7.3.2	Role and Expectation	137
7.3.3	Role and Supervision	138
7.3.4	Role Generation	138
7.3.5	Role and Stereotype	139

7.3.6	Role Transitions	140
7.3.7	Micro Level Role Identification	140
7.3.8	Role Balance in Groups	142
7.3.9	Role Configuration in Groups	142
7.3.10	Role with Malicious Intent	143
7.3.11	Privacy, Ethics, and Role	143
Appendix A Edit Intention Annotation Training		145
A.1	Annotation Task	146
A.2	Frequently Q & A	146
A.3	The Annotation Interface	147
Appendix B Interview for Role Validation		149
Bibliography		151

List of Figures

- 3.1 Iterative methodology for role identification 26
- 4.1 **The taxonomy of edit categories.** Note: Insertion is abbreviated as I, Deletion as D and Modification as M 48
- 4.2 Distribution of occupied number of roles. 57
- 5.1 The relative frequency of each edit intention, and its F1 score provided by the **BR** model. 74
- 5.2 Interaction effect of different levels of edit intentions and different levels of previous article quality (**prev**) on article quality changes. All variables are standardized. The Y-axis measures the predictive margins and X-axis refers to different standardized levels of edit intention. 80
- 6.1 Statistics about users’ participated sessions (left), the number of distinct roles they occupied throughout their lifetime (middle), and their role occupation per user-session (right). 97
- 6.2 BIC scores for GMM model with different number of K 99
- 6.3 Recommendations when a user logged into CSN 106
- 6.4 Recommendations when a user was browsing a thread on CSN 107
- 6.5 The percentage of different role occupation from 2004 to 2017. 118
- 6.6 The percentage of role occupation for users who participated in CSN for at least 12 months in their different tenure. (0, 1] refers to members role occupation in their first month, with (1, 6] as their second months till six months. Similarly, (6, 12] denotes role percentages from their six months to one year and (12, +) means after one year. 119
- 6.7 Conditional probability of role transitions from one session (row) to another (column) after the first (left) and tenth (right) session. 120
- A.1 The annotation interface on Wikipedia 147

List of Tables

- 3.1 Sample operationalization for different facets of roles 29
- 4.1 Edit type dataset description 45
- 4.2 Edit categories prediction results 52
- 4.3 Derived editor roles and their representative edit types 53
- 4.4 Article Quality Prediction Performances. P-value: < .001 :***, < .01 :**, < .05 :* 61
- 5.1 A taxonomy of edit intentions in Wikipedia revisions, Cronbach’s α agreement and the distributions of edit intention before and after corpus expansion. 69
- 5.2 Performance comparison for predicting edit intentions from revisions. Best results are bold. 73
- 5.3 The edit intention distribution in the first sessions (Intention Dist) and the revert ratio comparison (Revert Ratio), among non-survivors (NS) and survivors (SS). The numbers are bolded if 1-way ANOVA tests for difference between two groups are significant, with $p < 0.05$ 77
- 5.4 Regression coefficients of different edit intentions for predicting Newcomer **Survival** and Article **Quality Changes**. Here, † means the coefficient is statistically significant ($p < 0.05$) 78
- 6.1 Definitions and examples of nine goal-oriented conversational acts. 92
- 6.2 The intra-class correlation and correlations between human decisions and predictions for 9 conversational acts 93
- 6.3 Derived roles and their representative behaviors ranked by their importance in descending order. 98
- 6.4 Performance comparison for different types of recommender systems. . . 105

6.5	Question that examines how much people self-identify themselves with different role behaviors.	109
6.6	Pearson correlations between members' self-reported roles and our predicted roles that members occupy.	112
6.7	Pearson correlations between members' self-reported role behaviors and their actual behaviors	114
6.8	Survival Analysis predicting how long members continue to participate in the community. $p < 0.001$: ***; $p < 0.01$ **; $p < 0.05$ *. Number of users = 66,246. Number of user-session records = 522,429	115
6.9	The top 9 most frequent role transition patterns.	119

Chapter 1

Introduction

*We have many concepts but few confirmed theories;
many points of view, but few theorems; many “approaches”; but few conclusions.
Perhaps a shift in emphasis would be all to the good.
– Robert King Merton*

Online production communities like Wikipedia aggregate the efforts of hundreds of millions of volunteers to product complex artifacts such as the largest encyclopedia in human history and the software that runs the internet. Despite their proliferation into diverse aspects of life, such communities are not always successful in soliciting contributions and producing anticipated outcomes. Two major challenges are: how to sustain members’ engagement and how to coordinate users’ activities to contribute to public goods and community needs. For instance, lack of appropriate contributions has left over 88% of the roughly 5.7 million articles in the English Wikipedia at “stub” or “start” quality levels (as of January 2019 shows), and 60% new editors do not come back ([Halfaker et al., 2013](#)). Sustaining member engagement and ensuring quality contributions is a challenge not specific to Wikipedia. In online health support groups, around 10% thread-starting messages get no replies and many of the replies are not relevant to thread-starting posts, for example providing emotional support when people were seeking information ([Wang et al., 2015](#)).

In order for such complex socio-technical organizations to succeed, online communities have to depend equally on the technical infrastructure on which they rest, the policies that govern participants to behave in ways consistent with community goals, and **the behavior, roles, and coordination of their members**. The goal of this

thesis is to study members' participation and coordination in online production communities, focusing on the social roles they enact, which link individual contributions with community-level coordination and outcomes (Stewart et al., 2005). To this end, I examine how to integrate computational methods with insights from social science to study social roles and the optimal organization of online communities.

Social science has provided rich taxonomies of social roles within production groups. They range from 27 roles that Benne's and Sheats identified as fulfilling a group's needs to accomplish its production tasks (e.g., "opinion seeker", "coordinator", "evaluator-critic"), to maintain itself as a functioning group (e.g., "encourager", "harmonizer", "gate-keeper") and to meet the needs of individual members (e.g., "blocker", "aggressor", "playboy") (Benne and Sheats, 1948) to a recent taxonomy of 10 group roles covering a similar set of functions (Mumford et al., 2006) (e.g., "communicator", "cooperator", "completer"). Natural language processing research provides us with a variety of techniques to automatically **identifying** social roles in online communities. For example, Bamman et al. (2013, 2014) leveraged probabilistic graphical models to learn personas in movie plot summaries and English novels. Previous work also discovers roles in social networks based on the network structure, and typically focus on roles such as centers of stars, members of cliques, peripheral nodes. For example, *RoleX* introduced a unsupervised approach to extract features for each node, group features and interpret clusters (Henderson et al., 2012); *struc2vec* uses heuristics to construct a multi-layered graph based on topological metrics and simulates random walks on the graph to capture structural information (Ribeiro et al., 2017). Other examples include such models as mixed membership stochastic block models (Airoldi et al., 2008), unsupervised matrix factorization methods (Hu and Liu, 2012), or semi-supervised role inference models (Zhao et al., 2013). Another line of work formulated the identification of (predefined) social roles as classification problems. For example, Welser et al. (2011) identified four roles in Wikipedia: substantive experts, technical editors, vandalism fighters and social networkers. Fazeen et al. (2011) classified Twitter users into leaders, lurkers, associates, and spammers. Other common roles identified in online media include experts (Zhang et al., 2007), opinion leaders (Bodendorf and Kaiser, 2009), and influential bloggers (Agarwal et al., 2008).

However, most social science work has focused on **describing** roles that are designed to be generic without accommodating the activities associated with tasks in specific

contexts or automating the process of role identification. Although utilizing the network structures guarantees generalizability when discovering structural roles, the central problem is how to construct a network that can reflect user interactions in a meaningful and representative manner. While there has been work mining semantic actions for automatic role inference, identification of social roles has not had a correspondingly strong emphasis in the language technologies community. Moreover, a variety of methods were developed to extract specific “roles”, patterns, or components in different contexts, largely ignoring the relevant social theories on roles and lacking generalized methods about how to extract roles, let alone examining systematic evaluation of roles, how roles change over time and how the awareness of roles influences role holders’ performance, the expectations of others, and the production as a whole.

This thesis presents a systematic identification of social roles from a combined view from social science and NLP by taking into account three major challenges: (1) In contrast to roles in conventional organizations, roles in online communities are often self-selected and emergent, without explicit expectations associated with them, and only a little prior research has provided consistent definition and methodology. (2) Members’ participation in online production communities are recorded in what they do, to whom, and why. Although numerous studies have discussed how to identify roles based on users’ behavioral regularities, most research classified users based on their repeating patterns of activities or social network signatures, failing to capture what type of work were actually performed and for what purposes users conducted such interactions. (3) Moreover, members move upward or downward, vertically or horizontally within the community, making their roles change as a function of the tasks they perform, their tenure and audience in these communities. Understanding the mobility and stability of roles requires accurately delineating the dynamics (paths, directions, and strengths) of role transitions.

To sum up, this thesis takes highly detailed views from both theories of social roles and models of text analysis to better understand social roles by examining how social science theories of roles can be applied to online communities, describing a general methodology for identifying roles in any given domain, and identifying roles in two distinct contexts and using them to better understand socially important behavioral questions in these communities.

1.1 Thesis Overview

This section presents a detailed view of computational social roles in Chapter 2-6.

1.1.1 Role Theory

Chapter 2 provides a comprehensive definition of social roles by first reviewing existing conceptualizations of social roles in both theory and empirical research. We then describe the development of our social role framework, which hangs on five core facets of roles: **agent**, **interaction**, **goal**, **expectation**, and **context**.

1.1.2 Methodology for Identifying Roles

In addition to this easy-to-operationalize theoretical role framework, Chapter 3 introduces a generic methodology to recognize social roles, which is a repeated cycle of role postulation, definition, identification, and evaluation. This generic framework can be applied to any types of online communities. We also present a set of general approaches for evaluating role identification, including quantitative measures, qualitative evaluation, validation with role holders, and evaluation via downstream applications.

1.1.3 Case Studies of Role Identification

Based on the theoretical framework of roles and generic methodology for role identification, we further demonstrate how such framework and methodology can be proactively utilized to understand social roles in two socially important environments:

1. **English Wikipedia**¹: Wikipedia is a large, task-focused community whose goal is to produce a free, high quality online encyclopedia. Wikipedia is among the seven most popular websites globally, with approximately 30,000 active monthly editors in the English version (Foundation, 2017). Given its scale and complex social dynamic, Wikipedia must manage many types of contributions, including administration, community-building, writing and copy-editing. Wikipedia has multiple metrics of success, including an automatically computable measure of article quality (Warncke-Wang et al., 2013).

¹<https://en.wikipedia.org>

2. **The American Cancer Society’s (ACS) Cancer Survivors Network (CSN)**²: The American Cancer Society hosts online health support groups where cancer patients, survivors and caregivers exchange information, emotional support, social comparisons and companionship (Wang et al., 2015). The site consists of 40 discussion forums organized around type of cancer (e.g., breast, colorectal), demographics (e.g., youth, caregivers) and overarching topics (e.g., grieving, long-term effects of treatment). Over 204,000 people have registered for these forums, with almost 12,000 visiting each day. Contributions here are the conversations that people have with each other.

The two communities differ in what they produce, the activities and roles common in them, how they coordinate members’ activities, and the metrics of success that can be applied to them. By examining two distinct communities, each with multiple sub-groups or communities, we can test the robustness and generalizations about our role framework and identification methodology, as well as the similarities and differences in the ways that roles operate in online communities. In detail, certain roles on Wikipedia have more explicit expectations documented in guidelines and policies than CSN does. Wikipedia is designed to be non-interpersonal, while CSN is about interpersonal relationships. The core activities of members in those two communities are largely different — editing articles and discussing edits on Wikipedia and exchanging social support on CSN, which raises the question of how a generic method can be used to discover roles in different communities.

Chapter 4-5: Role Identification in Wikipedia

We address the identification of editor roles on Wikipedia by examining two core facets of social roles: (1) behavioral edits (**Interaction**), and (2) intentions (**Goal**) of edits.

To better understand editors’ editing behaviors on Wikipedia, we proposed a taxonomy of edit categories and built machine-learning models to automatically identify these edit categories associated with editors’ edits. We then introduced a graphical model analogous to Latent Dirichlet Allocation to uncover the latent roles in editor’s edit histories. Applying this technique revealed eight different roles editors play. We also validated how our identified roles collaborate to improve the quality of articles,

²<https://csn.cancer.org>

and found that editors carrying on different roles contribute differently in terms of edit categories; articles in different quality stages need different types of editors.

While this above taxonomy categorizes edits into low level actions such as file deletion, simply understanding the syntactic edit categories cannot tell the difference between simplifying a paragraph and maliciously damaging that paragraph, since both involve deleting a sentence. Since this nuance can largely affect our role identification, in Chapter 5, we further modeled another dimension of emergent social roles - **Goals** behind roles' interactions by proposing a 13-category taxonomy of the semantic intention behind edits in Wikipedia articles. We use this model to investigate edit intention effectiveness: how different types of edits predict the retention of newcomers and changes in the quality of articles. Our analysis further validated that articles in different stages need different types of edits.

Chapter 6: Role Identification in Cancer Survivor Network

This part presents one empirical study as an effort to better examine the behavioral roles that members occupy when participating in online health support groups. Specifically, Chapter 6 operationalized the facets of **Context**, **Goal**, **Expectation** and **Interaction** to model the emergent roles that members take on when participating on CSN. We identified eleven roles that members occupy such as *emotional support provider*, *welcomer*, and *story sharer*. We also described member role dynamics interacting with long-term participation and dropout in the community. We further validated the effectiveness of our derived roles by correlating our predicted social roles with members' self-reported role behaviors, and by incorporating roles as additional features to help recommender systems to accurately match support seekers with support providers.

1.2 Research Impact

This thesis investigates social roles in online communities via developing theories about social roles and computational models for identifying roles.

This work makes significant contributions to theoretical framework of social roles. Different from most social science work that focused on generic roles without accommodating specific activities associated with tasks in different contexts, this work proposes

five generic and measurable components, which are relatively minimal but complete compositions of social roles. We also have successfully demonstrated how to utilize this framework to model roles in two socially important contexts. Future studies could build upon this framework to identify and analyze various roles that people actually enacted in group processes.

Our generic methodology for role identification, which is a repeated cycle of role postulation, definition, identification and evaluation, can be applied to any other type of community, both online and offline. Most existing empirical methods for identifying roles in other domains can be abstracted into our generic methodology framework for role identification. Our unsupervised approach for extracting roles also requires less resources and annotation compared to supervised role identification, and it allows easy-to-use integration or plug-in of different types of unsupervised methods, such as topic modeling, mixture model based clustering.

The present work presents a systematic overview of quantitative and qualitative evaluation approaches for unsupervised role identification such as the perplexity of the model on held-out data and human interpretation of the component coherence. For example, when examining members' roles on Cancer Survivor Network, in addition to quantitative validation of model fit, we followed thorough in-depth interviews with domain experts and used their input to support the validity and quality of our derived roles. Our identified role was further evaluated via a large-scale survey analyses that examines the correlation between members' self-reported role occupations and our predictions, and a deployed recommender system that uses these knowledge as the basis to build interventions for real world benefits. This system that matches users to roles and tasks was deployed in the live site of Cancer Survivor Network. Overall, this iterative role identification process is reproducible broadly within the HCI and NLP community, as are our mixed-methods for evaluating the quality of derived roles.

Practically, it expands our knowledge by revealing the behaviors that editors occupy when editing Wikipedia and the functioning roles that members enact when participating in online cancer support groups. This thesis also provides insights on how the presence of different types of roles and their interaction with various context factors including task level and tenure of members, influences the group performance such as the quality changes of Wikipedia articles. Our role modeling methods can be employed

to develop tools that detect members' needs, track their activities, and offer them help and task of interests. Such identified roles can help members in online communities better know themselves and others. This can provide guidance for UX research on how to incorporate this information into profile pages and other interface affordances. The derived roles can also be incorporated as additional features for connecting users to other users, content and tasks based on their roles along with other information about them (e.g., their disease, expertise or, emotional support needs). In addition to the potential in boosting recommendation performance, members' behavioral roles can also be used as explanations to users about why such recommendations are made. For example, instead of "*You might be interested in ...*", the recommendations can be explained like "*This is an information expert who can help you*" or "*This article needs help in copy-editing*". Online communities could also introduce some of these derived roles as badges to encourage users to assume these roles and reward those who do so.

Chapter 2

Role Theory

The ruler rules, the minister ministers, the father fathers, and the son sons.

– Confucius

The term *role* has its origins in the vocabulary of the stage. Early sociologists began using the term to draw metaphors to this use, illustrated in the work of [Goffman \(1949\)](#). In modern social psychology use, [Mumford et al. \(2008\)](#) defines a role as “a cluster of related and goal-directed behaviors characteristic of a person within a specific situation”. [Heiss \(1990\)](#) defined roles as behavioral expectations for what a person *should* do. The “*should*” comes from internal and external sources, both the expectations associated with established and recognized roles, and the person’s own self-concept and inclinations toward particular behaviors and characteristics. Theory on coordination in groups and organizations has emphasized role differentiation, along with division of labor and formal and informal management, as a major mechanism through which members coordinate complex activities ([Kozlowski and Klein, 2000](#); [Kittur and Kraut, 2010](#)).

In offline settings, roles are often formally assigned, with a formal job title and prescribed activities needed to fulfill the role well, as described in the traditional *structural* perspective ([Ebaugh, 1988](#)). Such roles are mainly based on formal and informal social expectations and norms along with positive and negative sanctions to support the norms. Thus, both the group and individual role incumbents are likely to have clear expectations of what the incumbents should do ([Akerlof and Kranton, 2000](#)).

However, in online communities, members’ emergent roles are not structurally de-

defined or constrained. They instead emerge from common patterns of members' behaviors. As in conventional organizations, members in online communities engage in a variety of emergent and informal *social roles* that define the set of activities they perform. These roles are poorly captured by earlier definitions, and have received relatively less attention, since most empirical studies of roles only looked at formal roles like leaders or moderators (Burke et al., 2006; Mumford et al., 2006; Arazy et al., 2017; Zhu et al., 2012). This makes self-organized online communities a novel area for theoretical exploration of emergent roles.

The study of social roles provides several advantages for understanding how individuals contribute to their organizations or teams (Mumford et al., 2008), and for avoiding losses associated with dysfunctional conflict, role ambiguity, and social loafing. This chapter explores one fundamental question: what are social roles? This will provide a foundation for the chapters to follow. We begin with a review of existing conceptualizations of roles from both theoretical and computational modeling perspectives.

2.1 Theoretical Modeling

Social psychology and organizational behavior have provided rich taxonomies of social roles. In a systematic summary, Biddle (1979) summarized four types of social roles: (1) *basic roles*, like gender and age roles, that are grounded in society at large; (2) *structural roles*, like occupational, family and recreational roles that are attached to position, office, or status in particular organizational settings; (3) *functional group roles*, like the "mediator" and "investigator", which are not formally designated or attached to particular group positions or offices, but are recognized items in the cultural repertoire; and (4) *value roles*, like the hero, traitor, criminal, and saint, which embody the implementation or the negation of some recognized value or value complex.

Within these categories, subdivisions can exist. Influential early work on group work from Benne and Sheats (1948) categorized functional roles for group members into three broad subsets. *Group task roles* are related to the task "which the group is deciding to undertake or has undertaken", and facilitates and coordinates group effort in the selection, definition, and solution of their common problems. *Group building and maintenance roles* are oriented towards the functioning of the group, and are designed to regulate, strengthen or maintain the group way of working. *Individual roles* are directed

toward the satisfaction of the individual needs, which is not relevant either to the group task or to the functioning of the group. Roles can also be specialized for a domain; for instance, [Belbin \(1993\)](#) identified a set of roles when researching executive management teams. Those roles include *Chairman, Shaper, Plant, Monitor-Evaluator, Company Worker, Resource Investigator, Team Worker, Completer-Finisher, and Specialist*.

A complementary line of work describes role *typologies*, across axes of variation and within hierarchies ([Parker, 1990](#)). For example, [Margerison and McCann \(1990\)](#) developed eight roles, varying across four dimensions: relationship, information, decision-making, and organization. More recently, [Mumford et al. \(2006\)](#) synthesized around 120 specific roles in the team role literature into 10 new and broader roles, which were then further grouped them into three parent categories. Similar to [Benne and Sheats \(1948\)](#), *task roles* include coordinating team members about the tasks and clarifying team members abilities, resources and responsibilities, such as “*Contractor*”, “*Creator*”, and “*Critic*”. *Social roles* involve maintaining the social environment of teams, such as paying attention to members’ feelings (e.g., “*Calibrator*”) and creating positive and open working teams (e.g., “*Communicator*”). *Boundary-spanning roles* focus on important behaviors that team members exhibit outside of their teams such as “*Coordinator*” and “*Consul*”.

In online environments, members sometimes have formal assignment to roles and clear expectations of responsibilities ([Akerlof and Kranton, 2000](#)): for instance, moderator roles in many online discussion sites or administrator roles in Wikipedia. However, the vast majority of online communities lack visible role structures. Instead, roles are emergent, self-selected and are often not formally recognized ([Arazy et al., 2016](#); [Yang et al., 2016a](#)). As a result, although these emergent roles constitute consistent patterns of behavior, neither the role occupant nor other community members may have a clear understanding of who is occupying which role or how role occupants should behave. This more closely matches the *interactionalist* view of roles, which has built on several decades of sociological theory research ([Goffman, 1959](#); [Biddle, 1979](#); [Turner, 1990](#)).

There is substantial need for further work on interactionalist roles in order to apply theory to practice. Although prior work has described many various roles that people might assume, the high level roles are often too vague for practical use. Additionally, although such taxonomies comprehensively describe what roles might and should exist in

organizations or teams, few have provided practical guidance on how to identify which members occupy specific roles, or how to develop testable predictions about roles' impact on group outcomes or production. These roles usually correspond to **prescriptive role** — roles with clear expectations about behavior (the norms of “*ought*”), and in such cases, roles cause behavioral regularities. For instance, the role of *president* requires an individual to veto and sign bills, nominate Cabinet members, appoint ambassadors, etc., and not to make laws or decide how federal money will be spent. Finally, the aforementioned lists of roles are typically not specific enough about the **behaviors** people take on within specific emergent communities as they assume a role. All of these factors are challenges in any attempt to define and recognize such sociologically-informed, emergent roles, as well as how to understand those roles' consequences and impact. With the recent blossom of online communities, the need is growing to understand how those typologies of roles can be applied to understand members' roles, recognize emergent behaviors, and address issues facing users online.

2.2 Computational Modeling

While most theoretical work from social psychology or organizational behaviors focuses on either the description of generic taxonomies of roles or case studies of specific roles, computational modeling of roles usually concentrate on discovering latent roles in specific environments in a bottom-up and unsupervised manner. Here, we describe two major types of empirical studies of roles in the computational literature.

Network Structural Roles

The task of role discovery has been richly studied in the context of social graphs and networks (McCallum et al., 2007). Different approaches have been used for role discovery based on the network structure, and typically focus on roles such as centers of stars, members of cliques, peripheral nodes. For example, Somaiya et al. (2010) used Bayesian frameworks with an MCMC sampling algorithm for learning multiple roles of data points. Another algorithm, *RoleX*, introduced an unsupervised approach to extract features for each node, group features and interpret clusters as roles (Henderson et al., 2012). Similarly, *struc2vec* uses heuristics to construct a multi-layered graph based on topological metrics and simulates random walks on the graph to capture structural information of the network (Ribeiro et al., 2017). Other examples include such models as mixed membership stochastic block models (Airoldi et al., 2008), unsupervised

matrix factorization methods (Hu and Liu, 2012), or semi-supervised role inference models (Zhao et al., 2013). What these examples share is a belief that algorithmic evaluation of the frequency and features of connections between members of a social network is sufficient to discover the roles these members play in that network.

Behavioral Roles

Elsewhere, natural language processing research provides us with a variety of techniques to automate the identification of social roles in online communities by looking at the *content* of interaction and the use of behavioral cues in users' language (Kittur and Kraut, 2008; Welser et al., 2011). For example, Bamman et al. (2013, 2014) leveraged probabilistic graphical models to learn personas in movies and novels from the language of plot summaries and dialogue. Another line of work formulated predefined roles as classification problems. For example, Welser et al. (2011) identified four roles in Wikipedia: substantive experts, technical editors, vandalism fighters and social networkers. Fazeen et al. (2011) classified Twitter users into leaders, lurkers, associates, and spammers. Other common roles identified in online media include experts (Zhang et al., 2007), opinion leaders (Bodendorf and Kaiser, 2009), and influential bloggers (Agarwal et al., 2008). Yang et al. (2016a) focused on types of edit behavior in Wikipedia and used an LDA-based model to derive editor roles from their editing behaviors. Yang et al. (2015) introduced a lightly supervised approach to extract discussion roles over sets of participants' contributions within discussions, where the supervision comes in the form of an outcome measure from that discussion. Ferschke et al. (2015) applied a similar approach to identify coordination roles that predict quality of the Wikipedia pages where the discussions take place, and found four important coordination roles including Workers, Critiquers, Encouragers, and Managers. Maki et al. (2017) proposed a supervised graphical model with an outcome measure to define editor roles based on talk page behaviors on Wikipedia.

For both network structural and behavioral roles, various sets of automated inference techniques have been developed to identify specific roles - defined here not from theory but from empirical evidence, as patterns or components of behavior associated with users in different online contexts. These derived roles are **descriptive roles** that are only defined by behavioral regularities (the norms of "*is*" following Cialdini et al. (1991)). In practice, we simply use features to identify the behavioral regularities that define an emergent role. With descriptive roles, the similarities in behavior across peo-

ple doesn't arise from expectations, but rather external processes such as common experiences in a community (e.g., newcomers vs old-timers) and common goals (e.g., the goal in Wikipedia to produce a good encyclopedia, which leads people to write, copy-edit, guard against vandalism, etc). Although computational study of social roles allows identifying roles at a large scale, most of them located themselves in a very specific context (e.g., Wikipedia (Welser et al., 2011) or movie plots (Bamman et al., 2013)), and have not attempted to demonstrate generalizability or repeatable patterns for studying roles across domains. Furthermore, little attention from computational studies has been paid to social psychological theory of social roles which usually facilitate role discovery. This has, so far, been a missed opportunity for computational research to leverage a rich body of existing knowledge. For instance, the discovery of "leader" roles may become more informative in the context of Twitter (Fazeen et al., 2011) or Wikipedia discussion pages (Ferschke et al., 2015) once different types of leadership styles have been considered (Huffaker, 2010). By unifying social psychology work on roles with modern computational techniques, there is room for a renaissance in online role discovery and understanding.

2.3 Role Framework

In this section, building upon theories in social science and empirical work from computational fields, we propose a generic role framework, which is broadly applicable to online environments where roles may be emergent. Here, we define a social role as **a cluster of interaction patterns regulated by explicit or implicit expectations and adopted by agents in a social context to achieve specific social goals**. Our definition hangs on five core facets of roles: agent, goal, interaction, expectation and context. In the remainder of this section, we elaborate on the details of each of these facets.

2.3.1 Agent

Roles are performed by *agents*. That is, roles are reflected by features of human beings, teams or organizations in some extreme cases (Biddle, 1979). Such features may include relatively static attributes of individuals, such as their sex, age, color, states of disability or disease. In most cases, individuals were born with particular physical features (e.g., sex, race) or have slowly developed it over time (e.g., disease, age), which are not under their controls. People's non-behavioral attributes such as their demographics like

gender, race, family relationships, accents, are sometimes predictive and informative of the roles that they occupy. As a result, we may be able to utilize them to design features and to predict behavior and posit such associations between roles and person attributes. Except in specialized cases (e.g., sex roles), these characteristics are typically not an intrinsic part of roles. However, they are often entwined with expectations. For example, HCI researchers have found that in design activities, participants cluster personas based on gender (Hill et al., 2017), and that students using educational technology can enact science-related roles more effectively when those technologies are responsive to their race (Finkelstein et al., 2013). The facet of agent may not be a must for identifying roles that individuals or organizations enact, however, it is necessary for better understanding and interpreting roles.

Agents' attributes can also influence people's choices of occupying specific roles and others' expectations towards such roles. For instance, in the role example of male nurse, the social expectations of maleness often carries expectations such as assertiveness, masculine strength and aggressiveness, which may conflict with the affective role of the nurse — caring, warm, tendering, and sympathetic (Bush, 1976). But these expectations can be recognized and addressed directly - for instance, within the growing number of initiatives today supporting women in tech or women in STEM. The mission of these projects is to help females in assuming roles in science and engineering. These have historically been seen as male professional roles, due to only a small number of females in that field. However, as there is no intrinsic reason for those roles to be dominated by males, direct intervention can change the distribution of personal facets of specific roles over time.

Broadly, roles can be associated with any arbitrary set of people, and may exist from macro-roles which are characteristics of the whole societies to micro-roles that associated only with individuals. A set of people who share an identity or social positions can be assumed to enact roles, such as American Asians or the CEO of a company. In other examples, we may use the term of role to talk about role models, such as a Lincoln, a Mandela and a Hitler. Here, the role model occurs because a single person's behavior is sufficiently archetypal and representative that it causes us to discuss them, singularly, as a model for characteristic behaviors. For instance, there was a person named as Lincoln who actually demonstrated a set of characteristic behaviors, building on which we retained our expectations of such roles. We then use such roles in our

vocabulary because of this reinforced strong association between a person's name and his/her expected characteristic behaviors (Biddle, 1979).

2.3.2 Interaction

Roles are based on role holders' characteristic *interactions*, which can happen when role holders engage with other persons or objects, within or outside the context where the role is enacted. These interactions make up the core content of online communities - the threads and comments where discussion takes place. But these interactions also take place when role holders interact with the user interface of the community's website, or when they speak with their spouse or friends outside. Such interactions are observed by role holders and repeated over time (Turner, 1990). Whether or not each interaction is expected, valued, or approved by a role holder, each interaction shapes the roles they will choose to enact in the future.

The facets of interaction and agent are both closely related to individuals who exhibit them, but they are distinct. First and most importantly, most of agents' attributes are slow to change; attributes like race or sex are normally unchanging over the lifespan of the individual, and can usually be judged or self-reported once, and attributed to a person for a long period of time. Interactions, however, occur constantly, change often, and need to be observable from behavioral patterns multiple times in order to shape an agent's role.

Second, agents' attributes are passive and have no intrinsic effects or implications, although they can be viewed and responded by others, based on Biddle (1979). Sometimes reactions to such attributes carry strong individual preferences and culture prejudice. In contrast, interactional behaviors are transitory, transactive from one person to another, and usually change over time. As opposed to personal attributes, behavioral patterns can have intrinsic effects, such as to accomplish goals, to interact with others, and to facilitate other behaviors. Furthermore, interactions are not "born", and role holders are conceived to be responsible for choosing specific behaviors that they perform. Specific interaction behaviors may be performed one or multiple times by the individual. In order to determine that someone occupy a role, we must observe his/her behaviors multiple times over some period of time. These interactions occur at many granularities. A role may be occupied by a single person, in one specific context, or even exhibit one

single small behavior. These characteristics might consist of both core aspects of an interaction, or more peripheral, unnoticed features. But a role must be based on at least multiple characteristic observed goal-oriented interactions, otherwise it is impossible to determine a trend.

This does not mean that we need to exhaust the realm of interaction behaviors for role discovery. Individuals' behaviors that are indicative and characteristics of their roles are usually mixed together with other types of irrelevant behaviors. Thus, we must be clear on what aspects of behaviors we want to study, and discard behaviors that are not an essential part of the expected roles. For example, to understand the role of professor, attention should be paid to how he/she advises and teaches, rather than whether he/she eats a vegan diet. Oracle computational techniques may be able to filter out irrelevant behaviors automatically in the process of role identification, however, in practice, too noisy representation may mislead the role extraction models to make inaccurate predictions. Although we assume most roles to be closely related with the behaviors of individuals, there are exceptions. For instance, value roles (Biddle, 1979) such as "hero", "villain", "fool", are not the behaviors of specific individuals, but the values of the society. Other types of roles may appear through portrayal in literature and movies rather than in real life, and sometimes real persons may exhibit patterns of behaviors of those roles or figures in fiction.

2.3.3 Goal

Roles are associated with specific social *goals*, and are not isolated phenomena. Goals may serve the individual interests of the role occupant, role partners or the groups in which the roles are embedded (Mumford et al., 2006). For example, specific roles may be adopted to facilitate collective effort toward completion of a task, such as a devil's advocate in a course project team. Roles can also be oriented toward the long-term functioning of the group as a whole, such as "vandal fighter" in Wikipedia (Welser et al., 2011). Finally, people may take on some roles to satisfy their individual needs or desires, such as newcomers acting as information seekers to understand what the group has to offer or more senior members experiencing pleasure in mentorship. For these kinds of goals, members who enact the same type of role usually demonstrate the same set of goals, and their goals may be largely influenced by the outcomes of the groups or platforms. Sometimes, as role holders go through the process of accomplishing goals,

they may better understand why a role is organized in such ways.

As in the interaction facet, the goal facet of a role is also closely related to behaviors. Like the previous section, role holders are not born with characteristic patterns of behaviors. Rather, their behaviors come through internal operations which are learned through experience and help role holders to accomplish specific functions or goals. A clear recognition of goals can help distinguish characteristic role behaviors from irrelevant behaviors, a difficult filtering step for studying the interactions of a person assuming a particular role.

Sometimes, goals are reflected in role holders' behaviors. As a result, a set of low level interactions can be utilized to infer goals, as we demonstrate in Chapter 4 which uses low level syntactic actions to predict editors' intentions behind their edits on Wikipedia. Note that, this does not imply that goals and behaviors are identical, although a set of low level behavioral patterns are used to predict goals. The intuition or focus here is that explicitly inferring goals from behaviors makes the process of role identification interpretable from a basic scientific role understanding perspective. From a practical problem-solving perspective, the differentiation of goals and behaviors become less important, if a research aims at accurately predicting individuals' role occupation.

2.3.4 Expectation

Roles also involve *expectations* — norms, preferences, and belief — about typical interaction patterns of agents (Goffman, 1959; Jahnke, 2008). Adherence to or departure from these understandings can result in positive or negative sanctions from others (Blumer, 1986; Mead, 1934). Expectations are bidirectional: from role holders to others, and from others towards role holders, about what actions or behaviors “should” be associated with a particular role. Role expectation produces conforming behaviors - people who hold an expectation may behave in conformity with it (expectation for self) or take actions to ensure conformity in others (expectation for others). Expectations may be held or expressed by a single person or shared among individual, and may be understood or misunderstood by people who enact them.

In conventional organizations offline, roles are assigned and associated with strong expectations; managers in corporations speak differently when speaking to their em-

ployees than they do when speaking to executives, for instance (Bramsen et al., 2011). In many online communities, though, roles are emergent. In these cases, there may exist explicit regulations or informal or implicit “negotiated understandings” among individuals about how role occupants should or must conduct themselves, or no expectations at all.

Explicit expectations usually happen on assigned or formal roles such as administrators or moderators. In the context of Wikipedia, there exists explicit norms¹ developed by the community to describe the principles and agreed-upon best practices, including *Policies*, *Guidelines*, and *Essays*. Such community standards especially Wikipedia Policies² and Guidelines³ have wide acceptance among editors and describe practices that all users should normally follow. For example, in order to become the role of “administrator”, the Wikipedia community expects⁴ editors to demonstrate a set of evidence about their abilities, experiences and trustworthiness. Sample expectations for administrators include strong editing history, varied experience, helping with chores, observing consensus and talking to other editors (Burke and Kraut, 2008). Via semi-structured interviews with 56 volunteer moderators of online communities across three platforms, Seering et al. (2018) looked at how moderators engaged with communities, and revealed a set of tasks and expectations associated with being a moderator, such as approving new members, defending community, and critiquing offenders.

While those assigned or formal roles in online communities have clear written expectations, most “roles” are emergent and do not have explicit criteria of how they should behave. For instance, Wikipedia does not have written standards for defining “senior editors” in terms of the number of edits or tenure, although many editors perceive themselves to be. Similar cases apply to members who often welcome new editors, people who fight with vandals, and people who enact roles such as story sharers or support seekers in online health communities. Because these informal understandings can often be implicit or known only to long-time members, they can create barriers to community participation; for instance, on Stack Overflow, fear of hostile feedback for improperly meeting expectations of information seekers can prevent new users from asking questions or joining the community in the first place (Ford et al., 2016).

¹https://en.wikipedia.org/wiki/Wikipedia:Policies_and_guidelines

²https://en.wikipedia.org/wiki/Wikipedia:List_of_policies

³https://en.wikipedia.org/wiki/Category:Wikipedia_guidelines

⁴https://en.wikipedia.org/wiki/Wikipedia:Guide_to_requests_for_adminship

2.3.5 Context

Roles can be very broadly applicable or limited to specific *contexts*. These contexts set **boundaries** for role holders, i.e. delimiting the perimeter or setting the scope of roles. For example, *information provider* is a common role in many groups, including social Q&A websites, health discussion forums, and problem-solving groups. In contrast, the *committer* role (Wagstrom et al., 2012) is limited to open-source development communities. Within a community, roles can also appear based on privacy - a user may take on one set of roles in public, while enacting different roles in private discussions with peers.

Context can be relatively specific, and is often treated as the “containers” of roles. The entire platform can be one single context, or may include multiple contexts with different size. For instance, on Wikipedia, there exists a *context* of editing articles in the Main Namespace⁵ which may incubate a set of editing roles such as copy-editors or substantive experts; discussions on the choice of contents or topics go on in the *context* of talk pages associated with articles where editors can play different discussion roles such as facilitators, opinion providers and leaders. A context may produce another newer context, for example, the context of discussion forum may include the context of information seeking or the context of social chatting. The granularity of context largely depends on the intended roles to expect. In addition to the “container” function, contexts could also be “triggers” of roles. For example, for a mother who also enacts the role of a professor, the context of schools may trigger her perceptions of being a professor and start to behave as professors, while the context of home may cause the professor role to disappear and the role of mother to occur.

We do not aim to provide an exhaustive overview of how context influences role differentiation. Instead, we emphasize that roles are contextualized, and the identification of roles needs a specification of context, otherwise roles may not occur or we may end up with a mixture of roles from different contexts.

To sum up, it is generally difficult to separate one facet from another in our role framework. Instead of being mutually exclusive, these five facets — agent, interaction, goal, expectation, and context — are **mutually implicated**. Specifically, *agent* is usually associated with specific *context*, and *context* may guide *expectation* and *goals*. *Goals* itself

⁵https://en.wikipedia.org/wiki/Wikipedia:What_is_an_article%3F#Namespace

is strongly related to *expectation*, while both *expectation* and *goals* can affect the behavior or *interaction* of individuals. Despite potential overlaps, these five facets still can serve as the representative basis for defining and understanding roles.

2.4 Relevant Processes about Roles

The previous section defined the boundaries of how we will construct roles in this thesis. But roles do not encompass the entirety of social systems, and are not static over time. Understanding social roles also involves an understanding of how roles are integrated with other factors in those systems, the many other social processes associated with roles, and the way roles change over time. Briefly, we provide an overview of relevant processes of social roles in this section, some of which are further examined in our subsequent studies.

Role Transitions

Role transition occurs when people either move from one role to another (*interrole transition*), or change their orientations toward a role already assumed (*intrarole transition*) (Ashforth, 2000). There are several types of role transitions especially about inter-role transitions. *Macro role transitions* examine entry or reentry, transit, and exit from organizations such as promotion or transfer in a company. *Micro role transitions* investigate psychological or physical movement between simultaneously held roles such as switching between one's different roles. For example, President Obama alternates between his President role and father role, each of which is deeply defined by context, interaction, goals, and expectations. Compared to shifting roles within individuals, macro role transitions have received relatively more attention from empirical research, such as the transitions from readers to leaders in online communities (Preece and Shneiderman, 2009). For example, Arazy et al. (2016) found that emergent roles on Wikipedia are transient and editors frequently transit to other roles throughout their life-cycles. We further examine these micro role transitions in Chapter 6 to understand how members transit to other roles over their participation on CSN.

Role Configuration

Extensive literature has focused on what roles function in processes of group discussion, but do not answer the question of what roles are required for "optimum" group success

from a theoretical perspective. There has been some empirical work on which sets of roles or role behaviors promote successful group outcomes (Isotani et al., 2009). Higgs et al. (2005) found that team composition (diversity) is positively related to performance for complex tasks and negatively related for straightforward tasks. Wen et al. (2015) found that the leader's behaviors are more predictive of team performance than activity count of a whole team. Instead of revealing "optimum" team compositions, there has been some empirical work about at which conditions some roles must meet. For example, Benne and Sheats (1948) suggested that the combination and balance of role requirements is a function of the group's stage of progress with respect to its task, and also a function of its level of group maturity. Exploring role configuration patterns associated with successful teamwork deserves much attention in order to provide guidance for unsuccessful teams or organizations. As an initial effort to understand role configuration, we investigate at which conditions a set of roles is needed in our empirical study of roles on Wikipedia, and the composition of roles at a community level in our study of CSN roles.

Role Conflict

Conflicts of roles can occur on an individual level and an organization level. The role conflict at an intra-role level might happen when a person occupies multiple different roles. For example, a person who is a professor and a mother may have to satisfy both expectations from her school, students and her families on her after-office hours - one is to further research and advise, while the other is to care families. In terms of teamwork or organizations, role conflicts also exist. For instance, certain roles may not work well together (Belbin, 1993): dominant roles and coordinators may have problems with their equals, and an employee may have to report to and receive orders from several superiors. One main cause of such role conflicts is role ambiguity - the lack of "certainty about duties, authority, allocation of time, and relationships with others; the clarity or existence of guides, directives, policies; and the ability to predict sanctions as outcomes of behavior" (Rizzo et al., 1970). The solution to role ambiguity could come from role clarity, in terms of both objective presence of adequate role-relevant information and subjective feeling of having enough role relevant information (Lyons, 1971). However, in most online platforms, roles are emergent and self-selected, with no explicit expectations associated with them. As a result, neither the role occupant nor other community members may have clear expectations of who is occupying which role and how they should behave to do their jobs well. Improved role clarity is a key

potential practical application of the research findings in this thesis for improving user experience in online communities.

2.5 Summary

To sum up, the present chapter provides our working definition of social roles, working in the gap between the high level roles described in the social science literature and the low-level, empirical roles identified through automatic clustering of activities. The high level roles are often too vague. Because they are not specific about the behavior they encompass, they don't give community members guidance about how to behave in various contexts, and they prevent scientists from developing testable predictions about their impact. In contrast, computational models identifying roles from low level actions are specific, but they rarely generalize beyond a specific context. For roles to be effective constructs to explain and improve contribution and coordination in production communities, a good role theory must manage the tensions between describing general roles, like a task leader, that apply across many different communities, and concretely defined ones, like a copy-editor in Wikipedia, that are well-defined within a specific context. Our facet-based framework for defining roles is well positioned to bridge this gap and make social science theory usable for computational means. Through Chapter 4 to Chapter 6, we demonstrate how to utilize this role framework to understand different facets of social roles and to discover hidden roles in online communities.

Chapter 3

Methodology for Identifying Roles

By three methods we may learn wisdom: First, by reflection, which is noblest; Second, by imitation, which is easiest; and third by experience, which is the bitterest.

– Confucius

To test theories of roles, to empirically model roles, and to build interventions to improve communities based on roles require techniques to identify such roles at scale. This chapter gives an overview of our generic methodology for role identification and evaluation. Following this, our empirical studies in the subsequent Chapter 4 and Chapter 6 are specific studies showcasing the use of the framework as described here.

3.1 Generic Methodology

Our generic methodology for identifying emergent social roles in online communities is a **repeated cycle of role postulation, definition, identification and evaluation**. Intuitively, identifying roles usually starts with postulation about what roles might exist in any given communities. This continues on to defining the space of features that allow capturing of role dynamics, identifying specific roles by clustering over such constructed features, and then measuring performance of those roles both with quantitative measures as well as by influencing and improving downstream tasks with the addition of role-based knowledge. Additionally, all of the facets in the role framework from Chapter 2 can be used as motivators for specific feature design or choices in the modeling process. In the following sections, we describe each principle in this iterative process to design robust role identification models.

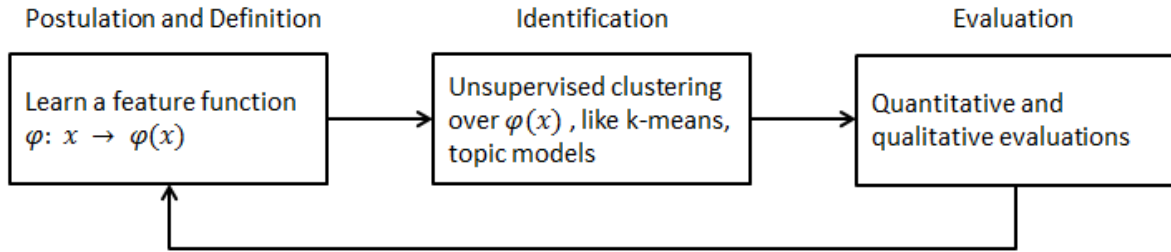


Figure 3.1: Iterative methodology for role identification

3.1.1 Role Postulation

To automatically identify social roles, one needs to postulate what roles might exist in a specific context. Intuitively, *postulation* means that if we have some knowledge of potential roles or role type behaviors, that knowledge can serve as a premise or starting point for further role identification. These early assumptions about the existence of certain types of roles are largely regulated by domain expertise or related theories in social psychology, sociology or linguistics. Frequently, rather than starting from a vacuum, cornerstone work in role theory such as [Benne and Sheats \(1948\)](#) or [Mumford et al. \(2008\)](#) can motivate this postulation process. For instance, when studying groups in the context of teamwork, it is a safe starting point to postulate that predefined and theory-backed roles like “*task leader*”, “*information provider*”, “*encourager*” will exist. In addition to utilizing the insights from theories in social psychology or social science, findings from empirical studies in a particular domain can also help postulate roles. For example, prior work on understanding users’ roles on Wikipedia revealed a set of editing roles ([Welser et al., 2011](#)) such as “*copy editor*” and “*substantive content provider*”. In Chapter 4, we were able to leverage this pre-existing knowledge to assist our definition, identification, and modeling of editors’ roles. Though the computational modeling of social roles often requires discovery in an unsupervised manner, where no prior role labels exist to learn from, this postulation process allows us to build and draw from domain expertise, supervising the selection of features to identify facets, the filtering of irrelevant information when it does not align to theory, and the selection of downstream tasks where performance should improve with a better role representation.

Based on the research presented in subsequent chapters, we can make two practical suggestions on how to use domain expertise during the process of role postulation.

First, we recommend looking at the postulated roles and their expected behaviors, and designing corresponding features to capture them. For example, to extract the leader *role* in teamwork settings may require designing features that capture different types of leadership *behaviors* associated with leaders. Examples of such behaviors include transnational and directive leadership types (as pointed out by [Bass and Stogdill \(1990\)](#) and [Zhu et al. \(2012\)](#)). Ensuring that features selected for role definition are capable of capturing these behaviors improves the ability to successfully discover those roles in data. Second, we recommend incorporating the postulated roles as “seed roles” in the modeling process of role identification. For instance, topic models such as Latent Dirichlet Allocation (LDA) can be used to automatically uncover latent topics or clusters in data unsupervisedly, but if there are a set of topics (clusters, roles) already known by researchers, these can be used as seeds to guide the topic discovery process. Technical approaches to seeding topic models exist — for instance, SeededLDA ([Jagarlamudi et al., 2012](#)) ensures that documents are a mixture of both standard latent topics, as well as pre-designated seed topics.

Our first recommendation requires somewhat laborious feature engineering and is more suitable for pipeline-based role identification systems¹. The latter recommendation requires relatively sophisticated modification of the inner workings of role identification algorithms in order to incorporate human knowledge at modeling time. Of course, while injecting domain knowledge helps in modeling, and can facilitate the validation of existing role theories, it is neither required nor does it need to be exhaustive. A thoughtful, targeted use of domain knowledge and past theory during modeling can be more effective than a larger quantity of less precise feature engineering.

3.1.2 Role Definition

Here, `definition` refers to the operationalization of all kinds of characteristics related to role holders. This part closely relates to our five facets theoretical framework of social roles; operationalizing each facet in the five facets framework can easily produce a relatively complete representation of typical behaviors associated with roles and people. Mathematically, the goal of role definition is, for each agent in any given context, to learn a feature function $\phi : x \rightarrow \phi(x)$. Intuitively, if a person named Alice participates in the

¹Pipeline-based role identification systems refer to systems that have a chain of algorithmic processing components arranged so that the output of each part is the input of the next.

sub-forum of breast cancer on CSN, x could be a message that Alice posted — “... *I had my surgery after 18 weeks of chemo/radiation...I’m having a hard time. I just burst out in tears at anything. They started giving me a antidepressant. Did anybody else have a problem like this?*”. Here, one example $\phi(x)$ could contain the bag-of-words representation of this message — $\phi(x) = [\text{uni-grams}, \text{bi-grams}, \dots]$. Practically, this operationalization includes designing features to capture who are the role holders, what they do, for what purposes, and in which places. This can be achieved in at least two distinct ways.

The first approach is to manually construct features to model role holders such as their personal attributes, behaviors, goals of interaction, as pointed out by our five facets role framework, which provides reasonable interpretability. For text-based environments, this can require understanding conversations between users by designing linguistic features to capture users’ language styles (Danescu-Niculescu-Mizil et al., 2011), emotions, opinions and feelings (Pang et al., 2008), topical interests (Blei et al., 2003), their usage of specific words or phrases such as personal pronouns (Pennebaker et al., 2015) or modal verbs, and their choice of entities or events. For instance, in Chapter 4 we look at work done by Wikipedia editors, developing a taxonomy of edit categories within articles. This taxonomy supported our ability to represent editors for role identification based on behaviors that were relevant to the roles they assume. Example edit categories that describe editors’ interactions include inserting a sentence, deleting an image, modifying an external URL. Similarly, Liu and Ram (2009) looked at the behavioral edits from editors to define categories of editors on Wikipedia.

In addition to behavioral measures, utilizing structural associations between these attributes from a network perspective is another effective source of feature construction for role definition. Social network analysis (Wasserman and Faust, 1994) has already been conducted to profile users (Mislove et al., 2010), model users’ positions and structures within their interactions with peers (Welser et al., 2011; Henderson et al., 2012; Bamman et al., 2013; Fisher et al., 2006). For example, Welser et al. (2007) utilized the visualized structural signatures to represent users and to extract their associated roles in Usenet groups.

Following both bodies of work, in Chapter 6 we give a demonstration of how to define functional roles in text-based cancer support groups by modeling content as well as participant interaction structures. We examine (1) nuanced intentions in language

Agent	Profile attributes like gender, age, job, race, title, interests, diseases
Interaction	<ul style="list-style-type: none"> • Elements from text such as words, entities, facts, events, topics • Structure signals from constructed networks like degrees, cliques • Multi-modal evidences from audio, images
Goal	Intentions, motivations, purposes
Context	Communication channels, topical areas
Expectation	Norms, guidelines, policies

Table 3.1: Sample operationalization for different facets of roles

that members exhibit to exchange social support such as seeking informational support and providing emotional support, (2) linguistic indicators of members’ interests by comparing their word usage with semantic categories provided by the psycho-linguistic lexicon LIWC (Pennebaker et al., 2015) such as members’ usage of words related to family or religious orientations, (3) content-based topics such as radiation, clinical trials from topic modeling, as well as users’ inclusion of external knowledge or citations in their messages, and (4) network analysis-based structure regularities, constructing a user-reply network and measuring in-degree, out-degree, and other structural features of interactions. To sum up, Table 3.1 provides a set of example elements that can be considered or extracted from role holders for a better representation.

A second approach exists. This is to make us of recent advances in machine learning, especially deep learning based techniques, to learn user embeddings or representations in an end-to-end manner, without feature construction. For instance, neural units like Recurrent Neural Networks (RNN) (Mikolov et al., 2010; Gers et al., 1999) or Convolutional Neural Networks (CNN) (Kim, 2014; Krizhevsky et al., 2012) can be used to process users’ language in messages or activities to obtain latent representations (Zheng et al., 2017). A recent proposed language representation model — Bidirectional Encoder Representations from Transformer (BERT) (Devlin et al., 2018) — has been the most recent of several neural representations of language that is powerful for a wide range of tasks using natural language, such as sentiment analysis and question answering. Our recent work (Yang et al., 2019a) shows how neural methods may be used to better define categories of users. We utilize a hierarchical Long Short Term Memory (LSTM) architecture to encode all messages made by a user, including both threads and comments, as his/her input. We then aggregate the message level information to represent

each user and predict his/her personal attributes, such as gender and disease type. This work demonstrates that such an approach, enabled by neural methods, requires less domain knowledge and less manual effort in feature construction. However, it also demonstrates the downfall of these methods, a loss of explanatory power, transparency, and interpretability of model outputs. A hybrid of the two approaches is also widely used for learning representations of users; along with careful evaluation steps involving subject matter expertise, these hybrid approaches are often the suitable choice for a new domain.

Overall, this approach to role identification, based on theoretical postulation and operationalization of representations, can be viewed as a *bottom-up* approach to uncover roles directly from data. There are also *top-down* steps that can be taken during modeling. Especially when working with generative models, it is possible to manually inject domain knowledge at learning time, providing insights for role postulation. An example of this is to determine the *number* of roles to seek out. Domains vary from only a handful of appropriate roles to dozens or even hundreds; top-down constraints based on theory and domain expertise can inform us how many components can fit the setting well while providing utility for downstream tasks.

3.1.3 Role Identification

`Role identification` refers to the computational process for extracting roles over a set of user data points after `postulation` of possible roles and `definition` of their characteristics have been completed. That is, after we have determined our feature definitions for $\phi(x)$, we next perform an unsupervised clustering to group together users whose behavior is similar, to define and discover roles. Here, we purposefully and closely relate “**clustering**” to role identification. In the term of clustering, users within the same cluster share high similarity, while users belong to a different cluster are dissimilar from each other (Xing et al., 2003). This aligns well with the underlying assumption in our definition of social roles — “a cluster of related and goal-directed behaviors characteristic of a person within a specific situation” (Mumford et al., 2008)). Thus, a role should be based on at least multiple characteristic observed interactions, both within a single role holder’s tenure in a community, and across different role holders. Without such continuity, follow-up identification of users as assuming a particular role becomes impossible.

Clustering analysis is a computational technique that allows us to group a set of data points into clusters. Data points that are similar to one another fall within the same cluster, while dissimilar data points fall into different clusters. The notion of a cluster varies significantly across clustering algorithms, and appropriate use of clustering methods is key to effectively identifying the hidden roles that users occupy. There are five classical unsupervised methods:

(1) *Centroid models* represent each cluster by a centroid vector, for instance, k-means clustering partitions n observations into k clusters in which each data point belongs to the cluster with the nearest mean. This centroid serves as the “prototype” of the cluster that data points are compared to. This approach is enormously widely used, but comes with drawbacks; for instance, k-means needs to specify the number of clusters, and is sensitive to outliers.

(2) *Distribution models* define clusters via statistical distributions such as Gaussian mixture models, which assume all observed data points are generated from a mixture of a finite number of Gaussian distributions with unknown parameters, and Latent Dirichlet Allocation (LDA), which allows sets of observations to be explained by a mixture of groups, where observations within each group are similar. Those distribution models often strongly rely on the prior assumption of the data. For instance, GMM assumes data from Gaussian distribution and LDA assumes Dirchelet prior. If data does not follow such Bayesian prior, the distributional models can result in sub-optimal clustering results.

(3) *Density models* define clusters as connected dense regions in a constructed feature space, such as DBSCAN (Density-based spatial clustering of applications with noise) (Ester et al., 1996), which discovers clusters of arbitrary shape of high density and expands clusters from them. This type of model does not need the specification of the number of cluster, and is robust to outliers and non-linear decision boundaries. However, it is highly sensitive to hyper-parameters and cannot handle data with varying densities, making it hard to determine the correct set of parameters.

(4) *Connectivity models* create a hierarchical decomposition of a set of data by some criterion. Hierarchical clustering, the prototypical example of these models, seeks to

build a hierarchy of clusters. Clustering can be either bottom-up, with each observation starting its own cluster followed by pairing and merging of similar clusters to form the hierarchy, or top-down, where all observations start in one big cluster and are then split recursively as one moves down the hierarchy. Hierarchical models are useful in producing meaningful taxonomy and compression, and do not need the design of the number of clusters. However, it lacks an explicit objective for optimization, and usually could not handle clusters with imbalance sizes.

(5) *Graph models* represent data as a graph, where a vertex denotes a data point, and the weight of an edge denotes the similarity between two data points connected by the edge. Clusters can then be formed by graph analysis, such as highly connected sub-graph clustering or spectral clustering. Graph models can handle heterogeneous data well, but are generally computationally expensive as it often requires solving the eigen decomposition problems.

In addition to these methods, dimensionality reduction techniques can also be used to group data points into different clusters. For example, principal component analysis (Jolliffe, 2011) can be used to reduce a large set of variables to a small set of variables called principal components, which still contains most of the information in the larger original feature space. Similarly, independent component analysis finds the latent independent components by maximizing the statistical independence of the estimated components.

As with role definition, new approaches in neural methods can provide benefits during role identification. Conventional clustering methods often have poor performance on high-dimensional data, due to the inefficiency of similarity measures used. Furthermore, those classical methods largely rely on the original feature space, which requires role operationalization component to be representative and comprehensive. In general, these methods also suffer from high computational complexity on large-scale datasets. With the recent development of deep learning, deep neural networks can be used to transform the data into more clustering-friendly representations, and most neural-network models can mostly be characterized as similar to one or more of the above models. Classical neural-based clustering models begin with training a deep neural network for representation learning first and then using hidden representations as input for certain clustering methods. As summarized in Aljalbout et al. (2018), deep

clustering first transforms input data into a latent representation, which can be used for clustering. Different types of neural network architectures can be used for this purpose, including multi-layer perceptrons (which use several layers of feed-forward networks), Convolutional Neural Networks, Deep Belief Networks (a generative graphical model consists of several layers of latent variables), Generative Adversarial Networks (a system with two competing neural network models that engage in a zero-sum game where the generator G learns a distribution to generate samples and the discriminator D learns to distinguish between real samples and the generated ones) (Goodfellow et al., 2014), and Variational Autoencoders, which learn the distribution of data via an autoencoder architecture. Once the input has been transformed into a high-dimensional representation, it can then be taken for clustering via one or more layers of the deep neural network. The learning objective of different types of deep clustering methods could come from clustering loss functions. Examples include k-means loss, which minimizes the distance between each data point and its assigned cluster center, non-clustering loss, which is independent of clustering algorithms and usually enforces a desired constraint, such as reconstruction loss associated with autoencoders, or multi-task loss, which predicts additional information from a combination of both clustering and non-clustering losses. These combination approaches are often highly effective. For example, training a deep autoencoder on a graph and then running K-means algorithm on the output can produce high-quality cluster assignments (Song et al., 2014). Deep representations and cluster assignments can also be learned simultaneously, such as in Xie et al. (2016). These approaches are effectively largely because clustering or role identification based on deep neural networks can learn non-linear mappings. This allows data to be transformed into more clustering-friendly representations without manual feature extraction.

Methods that learn latent representations first or learn both latent representations and cluster centers together may demonstrate more predictive power in downstream applications, like node classification or link prediction (Perozzi et al., 2014; Grover and Leskovec, 2016). However, as alluded to in the previous section on role definition, it is often challenging to figure out what clusters mean from models with hidden representations based on neural methods, compared to models with manually constructed features. This is especially true when there is no ground-truth information about cluster labels. In contrast, role identification with manually constructed feature definitions demonstrates more interpretability. Again, we do not aim to provide an exhaustive

list of role clustering techniques here. Role identification is exploratory — in most cases, there is no single “true” set of roles, and thus there may be no single “correct” method for role clustering. Different role clustering techniques will produce different outputs; we suggest using multiple approaches on a dataset, allowing exploration and learning about the domain and the appropriateness of particular methods, as well as optimization of task performance based on the roles that are identified.

3.1.4 Role Evaluation

Measuring the quality of derived roles through `evaluation` is as difficult as role identification itself. Researchers have a need to evaluate the identified components in a scientific way. The unsupervised nature of role identification methodology makes model selection and the specification of parameters (like the number of roles to discover) challenging. There is no absolute schema which should be used to measure the derived roles. The choice of a suitable role identification algorithm and of a suitable evaluation measure depends on the people who hold these roles, the specific task to perform with the learned roles, and the context in which the data was collected. As with previous sections, this section does not aim at providing an exhaustive list of clustering evaluation metrics, but to provide an overview of some widely used measures. Practically, one has to carefully compare and choose the measures that are applicable to their role identification task. With that being said, various measures exist for evaluating the quality of derived roles, which can be categorized into four distinct types of methods.

(1) Quantitative Measures

Here, evaluating derived roles is simplified into the task of measuring and evaluating of clusters. If we view the role identification as a clustering process, evaluation measures for classical clustering analyses can be used to evaluate the derived roles. Many researchers have defined methods summarizing the quality of clusters into a single quality score — a process known as *internal* evaluation, as defined in [Feldman et al. \(2007\)](#). Example measures include sum-of-squared-error, which sums over the squared distances between data points and the cluster centroids, silhouette coefficient, which measures how similar an object is to its own cluster (cohesion) compared to other clusters (separation), and cluster/topic coherence by measuring the degree of semantic similarity between high scoring words or data in a cluster.

Alternatively, an *external* evaluation can compare the clustering to existing ground-truths, if any are present, as a classification problem. However, this requires the existence of suitable gold standard labels. A variety of measures can be used to evaluate how good a role clustering is compared to such gold standards, including accuracy, precision, recall, and rand index (which measures the percentage of correct decisions produced by the algorithm). The generalizability of clusters on a held-out test set can also tell the quality of the role identification models. Examples of such measures include observing the model fit performance in perplexity scores (a measure of how well a model fits the data distribution; the lower the perplexity, the better the model), log-likelihood of a held-out test set, and measuring information criterion like Bayesian information criterion or Akaike information criterion.

(2) Qualitative Assessment

Qualitative evaluation (manual evaluation) refers to looking into derived components to see whether we can tell a story about these latent roles, which may be **highly subjective**. One direct approach is to ask human to judge whether they can use a label to interpret a latent component. For example, [Blei et al. \(2003\)](#) were able to attach labels to their topics that correspond well with the top ranked words in a latent topic. Similar sense-making approaches can be employed to assess whether the derived roles are interpretable, such as visualizing the top ranked behaviors or features associated with a role and asking domain experts to come up with a meaningful label for the collected behaviors. Indirect evaluations on whether a given latent component accords with manual judgments include the “intruder” human judgment tasks (introduced by [Chang et al. \(2009\)](#)), where annotators were asked to identify an intruder topic word for a given topic or an intruder topic for a given document, and the “observed coherence” task, in which human judges rated component coherence directly on an ordinal 3-point scale ([Newman et al., 2010](#)). Different from classical clustering tasks, the process of unsupervised role identification/clustering prefers more about whether the derived roles conform to human intuition, rather than measuring the effectiveness of a cluster to perform any particular task or describe data in any quantifiable way. For instance, [Chang et al. \(2009\)](#) found that surprisingly quantitative measures of models like held-out likelihood and human judgement are often not correlated and even sometimes negatively correlated, which further confirms our suggestion above that qualitative measure may be a better choice for role evaluation.

(3) Validation with Role Holders

These quantitative and qualitative methods work well with most settings, but when predicting roles enacted by people in online platforms, it remains unclear whether such evaluation metrics (either automatic quantitative measures or annotators' qualitative judgments) have reasonable correlation with users' own perceptions of their roles. To address this, surveying or interviewing users who tend to occupy the roles being studied may help address this issue. For example, researchers may design a survey to ask role holders about their perceived role occupations, and correlate survey responses with model predictions. In this thesis, when evaluating the derived roles on CSN, we conduct a large scale behavioral survey to ask role holders what types of roles they perceive themselves to occupy, as described in Chapter 6.6.3. Directly talking to users in the form of structured, unstructured or semi-structured interviews can also assist the process of evaluating the quality of role identification.

(4) Evaluation via Downstream Applications

The quality of roles can also be evaluated via their utility in an intended direct application, or their effectiveness and influences in improving any related downstream applications (indirect evaluation). In measuring model performance, for example, we can evaluate whether including latent clustering of subjects into roles can better help user classification or relationship prediction. These identified roles may be incorporated as additional features to help recommender systems connect users to other users, through which we can measure roles' contribution to recommendation accuracy. As an example, Chapter 6.6.1 introduces a recommender system that utilizes social role information associated with members to better connect help seekers with support providers, which demonstrates better performance compared to models without roles. In addition to performances, broadly, the derived roles can be used as part of the design of user profile pages and other interface affordances, like badges. In these design-oriented cases, the evaluation result of role identification can be tied to users' experience and satisfaction in using the role-enhanced functionality of a website.

3.2 Iterative Process of Identification

The process of role identification in Figure 3.1 is an exploratory data analysis, including an iterative understanding and analysis of what is the number of roles, how many

people occupy a role (the size of the clusters), and how a user is represented via what features. These parameters need to be tuned carefully, and each specific setting of parameters may need to go through the whole process of role discovery. Here, we describe a set of principles used to set parameters for robust role modeling as follows.

Unit of Analysis

Determining the unit of analysis for appropriately representing users is a key decision in modeling roles. On one hand, treating users as an aggregation of all their historical actions may prevent one from examining the evolution of roles or transitions between them. On the other hand, employing very small time intervals such as a single user action, may miss important larger constructs like a cluster of actions needed to achieve a goal. Thus, different temporal units — all activities within each calendar day, week, or month — can be explored to determine the appropriate granularity of user activity. For instance, in our role identification in Chapter 6, we used aggregated data from each user session, defined as a time interval in which the time gap between any two adjacent actions is less than 24 hours. We also explored other temporal units and found that the roles that emerged using a calendar day as the unit of analysis were very similar model to those emerging from session-level modeling, likely due to the similar time-scale. We also noticed that as the temporal unit increased from a day to a week to a month, the derived roles became harder to interpret, which may be because emergent roles in this community are more variable over time, unlike assigned roles in offline organizations (e.g., professor in a university).

Choice of Model and Evaluation

Model choices for identifying roles vary depending on many factors including the nature of the context, available computational time, and the expected roles based on theory and domain expertise. For example, identification of structural roles might require a deeper network analysis and graph-based clustering techniques, while identification of behavioral roles in text-based settings may need better language analysis and scalable clustering algorithms. Computationally, there are advantages and differences of different types of role identification methods. For instance, role-based clustering methods such as K-means and GMM may not be able to handle large and high-dimensional data well, while role identification based on LDA assumes an a priori distribution associated with roles. Similarly, when choosing evaluation metrics there is no absolute conclusion

on which one will most accurately judge the quality of the roles discovered by a model. We recommend trying different combinations from the evaluation measures we introduced before, such as a combination of both quantitative held-out performance and qualitative human judgment.

Number of Roles

The number of roles is a free parameter, and is the element most susceptible to over-tuning. Quantitative measures to mitigate this susceptibility exist. The goodness-of-fit of LDA models is often used to decide on a suitable number of topics, such as calculating the perplexity of a held-out set, component coherence score and other quantitative methods as we mentioned previously. Our iterative approach works well for unseen and unstructured data with trial and error evaluation, i.e., presenting different models with different number of roles and selecting the number of roles for which the model works “best” on the test set. However, this iterative process may be time-consuming with multiple iterations of trial and error. External knowledge, as discussed in our **Role Postulation** section, can further assist this step in determining the number of roles — if experts have already know how many roles they want to extract or how many exist in a given context. A deep analysis might determine the optimal number of topics (Teh et al., 2005) by utilizing variants of Hierarchical Dirichlet Process (HDP). In such cases, a Dirichlet process is used to generate the number of topics or roles, and no manual specification of roles is needed. That being said, we suggest to run a few iterations of models with different number of roles, manually inspect the clusters it identified, decide whether to increase or decrease the number of clusters, and continue iterating until a relatively optimal and satisfying level of granularity is produced.

Multifaceted Roles

Methodologies for identifying social roles should also take into account the multifaceted property of social roles. Put simply, one can perform multiple social roles simultaneously and over time. Hard clustering methods such as K-means or PCA assume that each user does or does not belong to a role, while soft clustering models such as LDA and GMM assign likelihoods or probabilities of a user belonging to different roles, often guaranteeing that users will be represented as a mixed combination of roles. Depending on the specific context, one might be preferred over another. For example, for functioning roles relevant to role holders’ expertise, a mixture of roles (soft

clustering) may profile the versatility and dynamics of users better, while for the role of a mother or a president, assigning a single fixed role to a user may be a better choice.

3.3 Reflection

Summary

The iterative process of role identification described in this chapter, including both clustering techniques and evaluation metrics, can be applied and extended to many contexts to cluster coherent sets of heterogeneous features into reliable user roles. The chapters that follow in this thesis involves in a set of empirical studies on Wikipedia and CSN that illustrates this iterative process of role identification. Specifically, in terms of **role postulation and definition**, we will demonstrate how to represent editors on Wikipedia via a fine-grained edit type taxonomy on Wikipedia, and how to define functional roles in text-based cancer support groups by modeling content as well as participant interaction structures. We will utilize two specific distribution based generative models for **role identification** - a graphical model that uncovers the hidden roles among users, and a mixture model that clusters heterogeneous user representations into a set of coherent roles - for role identification in Chapter 4 and Chapter 6. Unlike traditional unsupervised learning such as K-means clustering, our methods allows the acquisition of multiple roles per user representation, which are quite reasonable in profiling the versatility and dynamics of users. Furthermore, when identifying editors' roles based on their edits, we employed a distribution based generative model (LDA) since our feature representation of editors are relatively homogeneous in terms of edit types. The identification of members' roles on CSN is based on a set of heterogeneous features from both textual and network analyses, thus we chose GMMs together with the assumption that a user is a mixture of multiple roles.

Our Chapter 4 and Chapter 6 will also demonstrate **role evaluation**, focusing especially on how to select the number of topics via both quantitative and qualitative measures. Specifically, to determine the number of editor roles, we run our role identification model — a variant of LDA — multiple times with different numbers of roles and manually inspect the derived clusters with the help of domain experts in addition to the model perplexity. When identifying participants' roles on CSN, we search the number of roles over a bounded range from 1 to 20 in our GMM model and use a Bayesian

information criterion to obtain a focused range. We further validate the output from models with different number of roles with in-depth interviews with 6 domain experts who have a deep understanding of CSN. The results of these interviews support the validity and quality of our derived roles. Throughout our studies, we also demonstrate how to conduct sensitivity analyses to design role identification models with the appropriate unit of analysis and number of roles. To sum up, the generic methodology for identifying and evaluating roles is reproducible broadly, and can be applied to any types of communities, both online and offline.

Reflection on Unification

This thesis unifies theories and computation by introducing a five facets framework to describe what social roles are and proposing a generic methodology to utilize the role framework for identifying roles. Compared to existing empirical studies in the computational literature, both our role framework and identification methodology are generic enough to extract repeated patterns to study roles in different domains — for instance, different theoretic facets of roles can motivate the feature representation of role holders. In contrast to theoretical work introduced in Chapter 2, the present chapter provides a set of principles on how to specifically represent individuals’ behavioral regularities and identify their roles in different contexts. Here, the unification refers to unifying role theories and different identification techniques into our generic, iterative role identification process. Doing so enables us to utilize insights from theories and external knowledge to extract roles that are generic across similar contexts. For example, roles derived from CSN that model social support exchange may be comparable to roles that exist in other type of online health communities, and editors’ roles on Wikipedia represented by edit types may be similar to roles in other collaborative writing contexts such as Overleaf or Google Docs. Although this series of case studies demonstrate the success of using our role framework and methodology to identify social roles, those roles are primarily community specific, and we have not yet demonstrated how to identify some general roles that are not limited to a particular context. For instance, the role of vandal or leader exists in many online communities. However, jointly identifying such trans-community roles has received little attention. We discuss this as a future direction in Section [7.3.1](#).

Part I

Role Identification on Wikipedia

Chapter 4

Identifying Roles of Editors

It's not who I am underneath, but what I do that defines me.

– Batman Begins

Understanding the social roles played by contributors to online communities can facilitate the process of task routing. This chapter mainly focuses on developing new techniques to identify roles that editors enact when editing Wikipedia articles and on investigating how work contributed by people from different roles affects article quality. From a theoretical perspective, this chapter looks at the facet of **Interaction** in our five-facet role framework to represent editors, and **strictly** follows the generic role identification method to postulate roles (Section 4.4), identify roles (Section 4.5.1) and evaluate roles (Section 4.5.2). Specifically, we first introduce a taxonomy of editing types to capture what editors did — role definition — as a way to operationalize editors' behavior, and built machine-learning models to automatically identify the edit categories associated with edits. We then applied a graphical model analogous to Latent Dirichlet Allocation, a distribution based clustering model, to uncover the latent roles in editors' edit histories. The derived roles were evaluated via both perplexity scores on a held-out test set and experts' judgment. Applying this technique revealed eight different roles editors play. Finally, we examined the utility of the derived roles by measuring their influences on article quality changes. The results demonstrate that editors carrying on different roles contribute differently in terms of edit categories and articles in different quality stages need different types of editors. Implications for editor role identification and the validation of role contribution are discussed.

4.1 Introduction

Distributed work teams in online communities have become increasingly important in creating innovative products, such as GNU, Linux and Wikipedia. Millions of volunteers participate in the online production communities, exchange their expertise and ideas, and collaborate to produce complex artifacts. Better understanding of the participants and how they behave can make these communities more successful. For example, in Wikipedia, editors take up different responsibilities, when editing articles, based on their interest and expertise. Some, for example, might add substantive new content to articles while others may focus on copy-editing. Systems designed to route work to appropriate Wikipedia editors have focused on matching editors to articles that are topically similar to ones they have already worked on (Cosley et al., 2007). These task recommenders, however, have for the most part ignored the type of work that the editors can do. This paper develops new methods to identify roles that editors exhibit when contributing to Wikipedia and then tests whether work done by editors occupying different roles affects article quality. This knowledge can then be used to create more sophisticated task recommender systems that take both article content and editing skill into account.

The problem of identifying editors' roles in Wikipedia has attracted significant attention. Numerous studies have discussed how to identify roles based on users' behavioral regularities and social network signatures (Welser et al., 2007). Most research classifies editors based either on their edits in different namespaces (Welser et al., 2011) or via the user attributes such as access privileges (Arazy et al., 2015), personalized barnstars (Kriplean et al., 2008), etc. Classification based on users' attributes is relatively accurate, but this information is not available for many active editors and is insufficient in explaining the nature of an editor's work. While classification based on edit histories can be constructed for most active editors, current approaches focus on simple edit counts and access privileges fail to provide a finer grained description of the work actually performed in an edit. For example, it cannot tell the difference between an editor who copy-edits or rephrases a paragraph and an editor who inserts markup, template or information to an article.

In this work, we extend Daxenberger's (Daxenberger and Gurevych, 2012) fine grained taxonomy of edit types to differentiate editors who occupy different editing roles. In our

Dataset	# Revisions	# Editors	# Article	Time Period
Annotated Edit Category Corpus	953	728	-	2014.06.10 - 2015.06.10
Editor Modeling Revision Corpus	626,761	38,520	172,740	2014.12.01 - 2014.12.31
Article Quality Prediction Dataset	-	22,633	151,452	2015.01.01 - 2015.06.30

Table 4.1: Edit type dataset description

taxonomy, edits are distinguished contextually in terms of the object being edited (e.g. information, template, reference, etc.) and functionally, in terms of the edit operation (e.g. insert, delete, modify, etc.). Specifically, we developed 24 edit categories to understand how different users perform the editing task collaboratively. We then described the development and validation of methods for the automated measurement of these edits categories revealed in users' edits.

Building on this automated measurement of edit types, we use a graphical model analogous to LDA topic modeling analysis to identify the latent roles editors occupy, much as documents comprise topics. Just as documents are mixtures of topics, editors are mixtures of roles. The roles that editors occupy generate the edits they perform, just as the topics that comprise a document determine the works in it. In contrast to studies that employed either clustering analysis or principle component analysis to extract user roles (Liu and Ram, 2009), our role modeling treats an editor as comprising multiple roles at the same time. This approach makes the role more interpretable in capturing the versatility and dynamics of editors.

The collaborative contribution and interaction behaviors of such roles matters a lot in shaping the health and sustainability of Wikipedia. As a further step, we investigated how the collaboration of editors carrying on different roles predicted the quality changes of articles and some differences in the number of requisite roles for improving the quality of articles. The results demonstrated that different sets of roles are needed in the different quality stages of article. In detail, articles in Start or Stub¹ stages require more Substantive Expert to help with the content; articles in A or Good stages show

¹ https://en.wikipedia.org/wiki/Template:Grading_scheme

a lack of Wikipedia Gnomes² to repair the broken links and make things run more smoothly.

To sum up, this work lays a foundation for future research to automatically identify a fine granularity edit types for Wikipedia editors, to extract a mixture of editor roles and to encourage specific role setting to improve the quality of articles. It also helps in how to develop intelligent task routing systems to recommend users to tasks that match their expertise.

4.2 Related Work

A role is a bundle of tasks, norms and the behaviors that are expected of those who occupy a position in a social structure (Biddle, 1979). Roles are major mechanisms through which project members, including volunteers in large online communities, coordinate complex activities. Theory on coordination in groups and organizations emphasized role differentiation, division of labor and formal and informal management (Kittur and Kraut, 2010).

Previous social roles studies in online communities can be understood through the content of interaction and through the use of behavioral and structure cues (Kittur and Kraut, 2008; Welser et al., 2011). For example, a variety of roles have been identified in online discussion forums (Fisher et al., 2006; Yang et al., 2015; Welser et al., 2007), including *answer people*, *questioners*, *leaders*, etc. Another similar line of work studies the identification of roles in the context of a social network Bamman et al. (2013), e.g. *celebrity*, *newbie*, *lurker*, *troll*, etc.

In the context of Wikipedia, Welser et al. (2011) used both qualitative and quantitative methods to identify four roles in this online community: substantive experts, technical editors, vandal fighters, and social networkers. In contrast, Arazy et al. (2015) utilized the access privileges in Wikipedia and developed a set of twelve roles based on Wikipedia's organizational structure. Kriplean et al. (2008) showed that informal awards can be used to encourage and reward different types of valued work, and suggest that these Barnstars might be a good way to identify emerging types of works and different roles in Wikipedia. However, such role discoveries based on superficial

² <https://en.wikipedia.org/wiki/Wikipedia:WikiGnome>

edit types, structural signatures or access privileges suffer from either weak ability in differentiating editors or not readily accessible profile information. They are also inadequate in capturing what is actually edited and how editors collaborate in the construction process (Qin et al., 2014; Liu and Ram, 2009).

Existing studies on capturing the intentions behind a textual change (Faigley and Witte, 1981) suggest that edit types that each editor contributes to an article can also be considered to uncover the expected and enacted behaviors of an editor (Liu and Ram, 2009). For example, Daxenberger and Gurevych (2012) automatically assigned edit categories such as grammar, paraphrase or vandalism to edits in a document. Their taxonomy of edit categories (Daxenberger and Gurevych, 2013; Pfeil et al., 2006) is acquired through the differentiation and operationalization of surface edits and text based edits. However, relatively little research except (Liu and Ram, 2009) has gone into how such edit categories define and interpret specific roles in their coordinative contribution to editing articles.

Researchers have developed a number of techniques for identifying social roles online, generally employing either clustering analysis or principle component analysis. For example, Welser et al. (2011) grouped editors based on the types and content of their edits, as well as their user pages. Liu and Ram (2009) utilized a K-Means approach to classify contributors based on their actions in editing article pages. However, relatively little research has discussed the multi-faceted property of a user, namely, one can perform multiple social roles simultaneously. Graphical models used in uncovering the hidden topics in a document (Blei et al., 2003) can be leveraged here to acquire a mixture of user role representation, which are quite reasonable in profiling the versatility and dynamics of editors.

Our research also extends earlier research on role modeling by introducing evaluation criteria. Although earlier attempt to deduce the roles structure in Wikipedia have generated roles with face validity that are loosely consistent with expert's classifications, they provide no metrics to evaluate the quality of the roles. In the current paper we validate the methods we used by (a) estimating the percentage of the variance in low-level editing behavior the roles account for and (b) examining whether roles are useful for predicting changes in the quality of articles.

4.3 Research Question and Data

Our major research goal is to find a set of social roles associated with editors in Wikipedia based on our developed taxonomies of edit categories. Then we plan to investigate how these roles and their collaborative participation affect the quality and coordination of users' contribution. Our analysis is conducted on three datasets from English edition of Wikipedia, as shown in Table 4.1. Specifically, we will train a multi-class classifier to assign edit types for edits inside a revision on the Annotated Edit Category Corpus. Then apply the learnt model to the Editor Modeling Revision Corpus and identify editors' repeating patterns of activity. The Article Quality Prediction dataset is used to investigate how the collaboration of editor roles affects the changes of article quality.

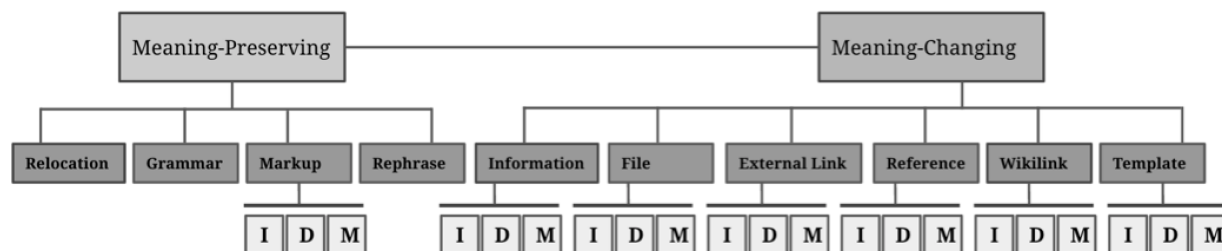


Figure 4.1: **The taxonomy of edit categories.** Note: Insertion is abbreviated as I, Deletion as D and Modification as M

4.4 Predicting Edit Categories

Previous research to identify editors' roles in Wikipedia based these assessments primarily used edit counts in different namespaces, structure signatures (Welser et al., 2011) and access privileges (Kriplean et al., 2008), without making assumptions about the type of work that a particular edit entailed. To address the inadequacy, we first introduce a fine-grained taxonomy of the types of edits editors make to Wikipedia articles (i.e., pages in Wikipedia namespace 0). We then describe new a machine-learning model to automatically identify the semantic edit categories (e.g., adding new information versus vandalizing and article) associated with each edit. These classifiers map low-level features of the edits, including the number of added or removed tokens, misspelling words, and comment length to a multi-label classification, representing the edit categories which an edit belongs. We then use this classification of edit types as

well as other information about the type of work editors do in Wikipedia as input into our role classifier. The development and validation of this machine-learning model are described in more detail as below.

4.4.1 Edit Categories Construction

Basing our research on [Daxenberger and Gurevych \(2012\)](#), we distinguished between revisions and edits. A Revision is created whenever an editor makes changes to a Wikipedia page. An Edit is a coherent local change and regarded as one single editing action. Each edit is associated with a set of labeling of edit categories, representing in which aspects it has been changed. A revision can contain multiple edits. For each pair of adjacent revisions, we collected a set of edits that has been made to transform from its parent revision into this revision.

Figure 4.1 provides an overview of our edit taxonomy, on the basis of these studies ([Daxenberger and Gurevych, 2012, 2013](#)). In this work, we annotated a set of edits rather than revisions. In general, this taxonomy considers actions (insert, delete, modify) applied to different objects in Wikipedia (e.g., information, templates or references), leading to 24 distinct edit types. The two top-level layers summarize whether these edit categories are meaning-preserving or meaning-changing.

Of the meaning-preserving edits, *Grammar* (G) means the edit is correcting spelling or grammatical errors, as well as fixing punctuation. When an edit attempts to paraphrase words or sentences, it is categorized as *Rephrase* (P); if such edit only moves entire lines without changes, it is defined as *Relocation* (R). For edits that try to operate with the markup segments, such as “===History===”, depending how it affects the markup, we divide them into three sub-categories, *Markup Insertion* (M-I), *Markup Deletion* (M-D) and *Markup Modification* (M-M).

Meaning-Changing edits depends upon how an edit affects the textual information content, we generated three categories: *Information Insertion* (I-I), *Information Deletion* (I-D), and *Information Modification* (I-M). Similarly, we acquired the remaining categories *Template Insertion* (T-I), *Template Deletion* (T-D), and *Template Modification* (T-M), *File Insertion* (F-I), *File Deletion* (F-D), *File Modification* (F-M), *External Link Insertion* (E-I), *External Link Deletion* (E-D), *External Link Modification* (E-M), *Reference Insertion* (R-I), *Reference*

Deletion (R-D), Reference Modification (R-M), Wikilink Insertion (W-I), Wikilink Deletion (W-D), and Wikilink Modification (W-M).

Our taxonomy breaks Daxenberger’s ‘Reference’ category (Daxenberger and Gurevych, 2012) into three finer-grained categories: External Link refers to links from articles to web pages outside Wikipedia, Wikilink refer to links to another page within the English Wikipedia and Reference describes the source of the information, to help the reader who wishes to verify it, or to pursue it in greater depth³. Note that we utilized the Revision Scoring package⁴ to identify Relocation, and did not include the category of relocation into our prediction stage.

4.4.2 Feature Space Design

The Annotated Edit Category Corpus contains 1997 edits. We annotated it based on a written annotation guideline. The annotation task is framed as a multi-label classification. That is, each edit will be assigned to one or more edit categories. For example, if an edit added a sentence to an article, this edit might involve insertion of information only or the insertion of information, a Wikilink insertion and a reference simultaneously. An edit containing the three components would be multi-labeled as I-I, W-I and R-I. To assess the validity of the annotation, we compared the annotations of 63 randomly sampled revision edits made by the first author and by an expert Wikipedian. Despite the difference in Wikipedia editing experience between the hand coders, the agreement between the annotations was substantial (Cohen’s Kappa = 0.723; see (Landis and Koch, 1977) for rules of thumb for evaluating strength of agreement using Kappa).

The machine learning goal was to classify an edit into one or more of the edit categories based on characteristics of the text changed, the comments editors used to describe their edits, and characteristics of the edit. To capture these characteristics, we developed the following features⁵:

³https://en.wikipedia.org/wiki/Help:Referencing_for_beginners

⁴<http://pythonhosted.org/revscoring/index.html>

⁵Here, Operation represents the action (Insert or delete) of an edit. Segment means the textual content that has been operated by a user. Segment Context is a piece of article content where the Segment is situated in (we collect the Segment content together with around 100 characters before and after its content).

- **Is minor:** whether the revision is marked as minor change.
- **Comment length:** the number of characters in the revision comment.
- **Typo mention:** whether the comment contains “typo” or “grammar”.
- **Is user registered:** author is registered or is IP user.
- **Number of edits:** the number of edits in this revision.
- **Number of tokens, capitals, digits, and whitespace:** the number of tokens, capitals, digits, and whitespace in a segment.
- **Types of POS tag:** the number of distinct POS tags.
- **Semantic similarities:** the maximum, minimum and average semantic similarities between segments within an edit.
- **Misspelling words:** the number of misspelling words in the segment.
- **Operation type:** the number of *insert* and *delete* operations.
- **Segment length:** the length of *insert* and *delete* segments.
- **Operation in template:** whether the edit happens in the segment context of template such as “{{{}}”.
- **Operation in file:** an edit happens in the segment context of file such as “[File /Image/ Media:]”.
- **Operation in markup:** an edit happens in a markup⁶ segment context, such as “===”, “==”, “<div>”, “</div>”, “”.
- **Operation in reference:** an edit happens in a reference⁷ segment context “<ref>”, “</ref>”.
- **Operation in external link:** an edit is performed in the segment context of external link⁸ such as “www:”, “http:” or “https:”.
- **Operation in wikilink (internal) link:** an edit happens in an internal link⁹ context such as “[[’, ‘]]”.
- **Template/markup/reference/file/external/wikilink in segments:** the number of designed markers related to template, markup, reference, file, external, wikilink that are contained in the segment.

⁶https://en.wikipedia.org/wiki/Help:Wiki_markup

⁷<https://en.wikipedia.org/wiki/Template:Ref>

⁸https://en.wikipedia.org/wiki/Help:Wiki_markup#Externallinks

⁹https://en.wikipedia.org/wiki/Help:Interwiki_linking

Given the input feature representation of an edit, we then built a machine-learning model for this multi-label classification (Yang et al., 2016b). Specifically, we used two of the multi-label classifier implemented in Mulan (Tsoumakas et al., 2010) using ten fold cross validation. We used the RAKEL ensemble method classifier, described in (Tsoumakas and Vlahavas, 2007). It randomly chooses a small subset with k categories from the overall set of categories. We compared this with the MLkNN classifier, which is based on K Nearest Neighbor method. Table 4.2 shows the evaluation metrics including Recall, Precision, micro-averaged F1 score and AUC (Area under Curve). Both methods gave classifications that agreed with the human judgments, indicated by the AUC score of 0.865 and 0.906 respectively. We chose to use RAKEL method in order to acquire a relatively better performance in terms of F1 Score.

	Recall	Precision	F1	AUC
RAKEL	0.575	0.730	0.643	0.865
MLkNN	0.363	0.724	0.482	0.906

Table 4.2: Edit categories prediction results

4.5 Modeling Editor Roles

Our edit taxonomy and its automated measurement only describe the types of work that an editor does when writing or revising the article pages that the general public associates with the encyclopedia. However, in addition to what Kittur and colleagues call this “direct production work” (i.e., edits to articles) (Kittur et al., 2007, 2009), Wikipedia requires a lot of behind-the-scene administrative and coordination work to be successful, and what might be termed the indirect work has been increasing as a percentage of all work done in Wikipedia (Kittur et al., 2007). To a first approximation, one can identify indirect work by the namespace in which it is done. For example, discussion of changes to articles is typically done in namespace 1 (article talk pages), discussion and changes to Wikipedia policies are done in the Wikipedia talk and Wikipedia name spaces (5 and 4 respectively), and much editor-to-editor communication occurs in the user talk namespace (namespace 3). To allow our role models to represent indirect work, such as social interaction, community support, and maintaining standards in our role models, we included the number of edits editors made in each Wikipedia namespace¹⁰

¹⁰<https://en.wikipedia.org/wiki/Wikipedia:Namespace>

Derived Roles	Representative Behavior
Social Networker	Main talk namespace, user namespace, reference modification
Fact Checker	Information deletion, wikilink deletion, reference deletion, file deletion, markup deletion, external link deletion
Substantive Expert	Information insertion, wikilink insertion, markup insertion, reference insertion, external link insertion, file insertion,
Copy Editor	Grammar, paraphrase, relocation
Wiki Gnomes	Wikilink modification, template insertion, markup modification, wikipedia talk namespace, category namespace
Vandal Fighter	Reverting, user talk namespace, reference insertion, external link deletion, paraphrase
Fact Updater	Template modification, reference modification, file namespace
Wikipedian	Wikilink insertion, Wikipedia namespace, template namespace,

Table 4.3: Derived editor roles and their representative edit types

into the role models.

We also include the number of reverts (i.e., returning a Wikipedia to a prior state) and vandalistic edits editors made in the role model. Unlike (Daxenberger and Gurevych, 2012), we did not create new classifiers to infer these edit types from editing activity. Rather we take advantage of two utilizes written by the Wikimedia Foundation that accurately measure this activity. Mediawiki-utilities Revert Check API ¹¹ measures revert. The Vandalism API ¹² returns the probability that a given revision is vandalism; we considered revisions with a vandalism probability scores larger than 0.85 to be vandalism. Reverts and vandalism was assigned to each of the edits comprising a single revision (i.e., all the edits done between consecutive saves to a Wikipedia page).

4.5.1 Role Identification Method

Our objective is to identity the roles that editors play, clustering editors who share patterns of work, using the types of edit they make in articles, their revert and vandalism, and edit counts in other namespaces. For this purpose, we used the graphic model

¹¹<https://pythonhosted.org/mediawiki-utilities/lib/reverts.html#mw.lib.reverts.api.check>

¹²http://ores.wmflabs.org/scores/enwiki/?models=reverted&revids=revision_id

underlying the Latent Dirichlet Allocation (LDA) method. LDA is an unsupervised, statistical generative model that can be used to discover hidden topics in documents as well as the words associated with each topic (Blei et al., 2003). It assumes that each document is generated as a mixture of latent topics and each topic is characterized by a multinomial distribution over words. In the context of Wikipedia, an editor, represented by his or her edit history of edit types, reverts, vandalistic edit and edits in non-article namespaces, is analogous to a ‘document’. The number of edits of different edit types, reverts, vandalistic edits and edits in non-article namespaces, is analogous to word frequency within the editor “document”. The latent roles derived are analogous to an LDA topic.

Here, roles are based on repeating patterns of activities or ‘structural signatures’ and are analyzed in action, based on the work itself. Roles that editors occupy generate the edits they perform; editors occupying the same roles have similar patterns of work. Unlike the use of the term role in sociology, our definition did not include expectations from role partners (Orlikowski, 2000) because in Wikipedia informal roles do not include strong expectations. Just as in an LDA topic model, where each document comprises multiple topics and each word can appear in multiple topics, an editor in Wikipedia comprises a mixture of roles, which may vary from one article to another, from one namespace to another or even within a single article. This approach is more realistic than previous ones that assumed that each editor occupies only a single role at a time and renders our extracted social roles more interpretable when describing editors’ versatility and dynamics.

4.5.2 Derived Roles Exploration and Validation

We trained a LDA model on the Editor Modeling Revision Corpus. We experimented with driving from 5 to 15 roles (i.e., topics in the LDA software) and evaluated the interpretability of the produced latent roles based on human judgment. Qualitatively, we first visualized the top ranked edit types for each role, and then authors interpreted the results based on whether such work types are coherent in explaining the given roles. We ended up with 8 roles and selected the edit-types and namespaces that are most likely to correspond to a role. We summarized the results in Table 4.3. Two experts familiar with Wikipedia applied a label to each topic, based on the behaviors most heavily associated with each role. Detailed discussion of these roles identified via the

LDA method is presented as below.

1. **Social Networker.** These editors make frequent edits in Wikipedia's communication spaces and their profile page but rarely edit articles. As demonstrated in Table 4.3, social networkers utilized 'Main Talk' and 'User' namespaces extensively. Instead of contributing to articles, social networkers tend to discuss article content and build profiles that show their interests and group membership.
2. **Fact Checker.** The most defining characteristics of these editors are the removal of content. Fact Checkers have extensive activities related to information, markup, and wikilink deletion, etc. While this may seem counter-productive on the surface, removing unnecessary content is part of Wikipedia's fact-checking process.
3. **Substantive Expert.** Substantive expert contributes by adding substantive content to articles, including providing references to increase the reliability of articles and inserting new knowledge to articles etc. They are the main content producers, engaging in many types of creations, and perform actions more frequently than average contributors.
4. **Copy Editor.** Editors who make contribution to improve the format, punctuation, style and accuracy of articles are referred as copy editors. They copy edit Wikipedia articles to make them clear, correct and concise mainly through checking grammar, paraphrasing and adjusting sentences to proper positions.
5. **Wiki Gnome.** "Wiki Gnome" is a term used by Wikipedians to refer to uncontroversial, but productive work. These editors make smaller contributions that tend to be focused towards making the content in Wikipedia cleaner by fixing issues with markup and easier to find by fixing and disambiguating links. These editors mainly work on Template insertion, Wikilink modification and Markup modification.
6. **Vandal Fighter.** These editors are the gatekeepers of Wikipedia. They monitor the feed of edits as they saved, identify vandalism and revert it (Reverting) and also post warnings to editors who vandalize Wikipedia (User Talk namespace).

7. **Fact Updater.** This group of editors contributes mainly to the template content of articles (e.g. Infoboxes – Boxes containing statistics and quick facts that appear on the right-hand side of most Wikipedia articles). Since Wikipedia covers topics that change over time, a lot of work needs to be done to keep these articles up to date. For example, when a company’s CEO changes or when a popular band releases a new album.

8. **Wikipedian.** Editors in this group contribute to a diverse namespaces such as file, template, draft, etc., some of who might belong to the administrators. These editors work in spaces that are seldom seen by readers to keep the hidden order ordered (Viégas et al., 2007). Although Wikipedian have limited activities in editing articles, they invest a lot of time to help organize and standardize Wikipedia.

Our findings of Substantive Expert, Vandal Fighter and Social Networker roles are consistent with the roles discovered by (Welser et al., 2011) and Wikipedian role is similar to the Quality Assurance role defined in (Arazy et al., 2015). However, the difference is that our Copy Editor, Wiki Gnome, Fact Checker, Fact Updater roles are obtained through a fine-grained analysis of editors’ edits types, which are not directly reflected by simple edit counts in different namespaces.

We also represented how mixed editors are by computing a Gini coefficient based on how many roles an editor has occupied. A user is considered as occupying a role if he/she has a probability higher than $\frac{1}{8}$ (0.125). The Gini coefficient is 0.3, indicating that editors do occupy different number of roles. This is consistent when we visualized how much percentage of editors occupies a certain number of roles, as shown in Figure 4.2.

To evaluate the validity of our identified editor roles our methods identified, we estimated the percentage of variance across editors in the number edits of each edit type the roles accounted for. This metric is analogous to communalities in a factor analysis or principal components analysis. In this regression model, the input is an eight dimensional vector indicating how likely the editor belongs to each role and the output indicates how many edits an editor contributes to a specific edit category. We built 24 regression models to predict edit counts in each individual edit category from editors’ role distribution. The average R-squared score for these models weighted by

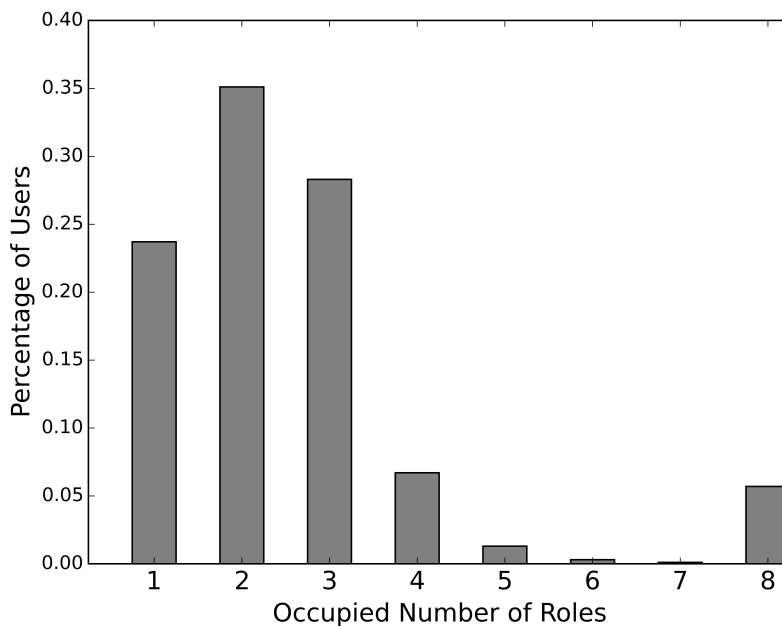


Figure 4.2: Distribution of occupied number of roles.

the frequency of the predicted behavior is 0.562, indicating that editor roles can explain over 56% of the variability in the numbers of edits of a certain type that an editor makes. The editor roles were especially successful in predicting grammar edits (81% of variance explained), modifying templates (76%), insertion of Wiki links (73%), and additions (62%) and deletions (52%) of information. Roles were poor in explaining insertion, modification and deletion of files, external link deletion and paraphrasing (all with less than 9% of variance explained).

4.6 Improving Article Quality

The quality of Wikipedia articles varies widely. Although there are over 4.5 million articles in the English Wikipedia, as of September, 2014 Wikipedians have evaluated fewer than 0.1% of them as good articles or better and over 88% of them as start or stub class articles (the two lowest quality categories). Collaboration among editors with different skills is essential to developing high quality articles ([Kittur and Kraut, 2008](#)). This section of the paper attempts to determine how contribution by editors occupying different roles at distinct times in an article’s history influence changes in its quality. Doing so will allow us to better understand the causes of quality variance in Wikipedia ([De la Calzada and Dekhtyar, 2010](#)) and will demonstrate the utility of our role .

To do so, we first measured the how much contribution made by a specific role to an article page during a certain time period. Then we explored the correlations between the coordination of editor roles and article quality, controlling for the number of editors, the total number of edits, etc. This analysis is conducted on Article Quality Prediction Dataset. Identifying roles entailed first applying our multi-label classification model of edit categories to categorize the work done during this work and then using LDA techniques to derive the roles from the edit categories performed by each of the editors.

4.6.1 Model Design

We modeled editor roles during the month of Dec 2014 and change in article quality in the first half of 2015 so that the data for modeling roles did not overlap with the data for computing changes in article quality. We measured the contribution of each role in the following six months by summing up all the work of editors who take up that role. Since each editor is a mixture of roles, we attributed the contribution of different roles to an edit in proportion to the probability that that the editor belonged to a specific role. For example, consider editor A who belongs with 80% probability to the Copy Editor role, with 10% to Social Networker and with 10% probability to Vandal Fighter. In this case, we consider one of A's edit consists of 0.8 edits contributions by the copy editor role, 0.1 edits by the social networker role and 0.1 by the vandal fighter role.

Dependent Variable

- **Article Quality Changes:** We validated how our extracted roles and their collaborative interaction contribute to article qualities by framing it as an article quality prediction task. Past work exploring the dynamics of article quality in Wikipedia used assessments applied by Wikipedia editors to articles [Hu et al. \(2007\)](#); [Lipka and Stein \(2010\)](#). However, these assessments are rarely updated and therefore are often out of sync with the quality level of the article at any given time. To get around this problem, we opted for a different strategy. Researchers have developed robust machine learning strategies for predicting the quality level of an article that do not suffer from such staleness. There are many models to choose from in the literature (e.g., [Anderka et al. \(2012\)](#) quality flaw model and [Lipka and Stein \(2010\)](#), which used writing styles to identify featured articles). However, we chose to use the model developed by [Warncke-Wang et al. \(2013\)](#) because it

focuses exclusively on current features of the article itself as opposed to the history of activity on the article. This model is currently used by Wikipedia editors and updated by members of the Wikimedia Foundation Staff to measure article quality and identify articles with stale assessment tags¹³. This model classifies articles into the Wikipedia's article assessment scale based on article length, number of headings, number of references, completeness (Warncke-Wang et al., 2013), etc. This classifier is highly accurate, with a mean agreement with classification made by Wikipedia editors of 0.609. Consistent with past work (Kittur and Kraut, 2008), we measured article quality using this classifier at two time points six months apart, Jan 1, 2015 and July 1, 2015 (denoted as previous quality score and end quality score respectively). In order to measure sub-class changes in quality we applied a simple weighted sum-based aggregation to the article quality scores such that Stub (the lowest class) was assigned a score of zero and Feature Article (the highest class) was assigned a score of 5 and multiplied the probabilities returned by the classifier by each score and summed the result. With this strategy, if 100% of the probability were centered on Stub, we would arrive at a score of zero. If 100% of the probability were centered on Featured Article, we arrived at a score of five. We calculated change in article quality by subtracting the previous quality score from the end quality score. Spot-checking by comparing changes scores with an examination of the two versions of the article revealed that even small increases in the change score represented clear improvements in the coverage and quality of the article, while decreases represented vandalism and other types of damage.

Control Variables

- **Previous Quality Score:** This is the article quality score in the beginning of Jan 2015. We controlled this variable to validate how role coordination affects the article quality in different stages of an article.
- **Article Registered Edits:** the total number of edits contributed by registered editors (not IP users) to an article page during the six-month time period.
- **Article Registered Editors:** the number of unique registered editors involved in

¹³https://meta.wikimedia.org/wiki/Research:Screening_WikiProject_Medicine_articles_for_quality

the past six months. Wikipedia is easy to edit does not mean that editors carrying different roles contribute with the same intensity or are needed in the same way.

- **Talk Registered Edits:** This is the total number of edits contributed by registered editors to the article talk pages.
- **Article Bytes Changed:** This variable summed the added (removed) bytes to an article page that increase (decrease) its length. Then we calculated the length increment by subtracting the removed bytes from the added bytes.

Independent Variables

- **Contribution of Social Networker (Social Networker):** We summed all the edits contributed by editors who take up the social network role in the past six month, dividing by the total number of edits in this article.
- Similarly, we obtained other seven dependent variables, including Contribution of Fact Checker, Copy Editor, Substantive Expert, Vandal Fighter, Fact Updater and Contribution of Wikipedian.

Variables	Model 1		Model 2		Model 3	
	Coef.	SE	Coef.	SE	Coef.	SE
Previous Quality Score	-.183***	.001	-.188***	.001	-.140***	.003
Article Registered Edits	.129***	.000	.125***	.000	.128***	.000
Article Registered Editors	-.046***	.000	-.045***	.000	-.045***	.000
Talk Registered Edits	-.031***	.000	-.030***	.000	-.030***	.000
Social Networker			.015***	.006	.023***	.014
Fact Checker			-.009***	.005	-.026***	.013
Substantive Expert			.058***	.005	.017***	.013
Copy Editor			.013***	.003	.029***	.009
Wiki Gnomes			-.033***	.005	-.073***	.012
Vandal Fighter			.008***	.006	.009	.014
Fact Updater			.006*	.005	.012*	.012
Wikipedian			.013***	.005	.047***	.012
Previous Quality Score × Social Networker					-.008	.005
Previous Quality Score × Fact Checker					.021**	.005
Previous Quality Score × Substantive Expert					-.139***	.005
Previous Quality Score × Copy Editor					-.017*	.005
Previous Quality Score × Wiki Gnomes					.049***	.005
Previous Quality Score × Vandal Fighter					.001	.005
Previous Quality Score × Fact Updater					-.008	.005
Previous Quality Score × Wikipedian					-.039***	.005
R-Squared			0.224		0.228	

Table 4.4: Article Quality Prediction Performances. P-value: < .001 .***, < .01 .** , < .05 .*

4.6.2 Result Discussion

Results of four regression models are shown in Table 4.4. Regression Coefficient (Coef.) is reported, which represents the main change in the dependent variable for one standard deviation of change in the predictor variable while holding other predictors constant in the model. Model 1 reports the effects of the control variables.

The strongest predictors were the previous score (-.183) and the article bytes changed (.409). The negative correlation of pretest score with change score reflects both regression towards the mean and the substantive phenomenon that as articles rise to higher quality levels, it is more difficult to increase their quality further. The positive coefficient for edits by registered may simply reflect that more edits generally leads to higher quality or may reflect the distinctive importance of registered as opposed to anonymous editors. The number of editors working on the article (-.046) and the amount of activity on the talk page (-.031) were negatively correlated with quality which may confirm prior work's conclusions (Kittur and Kraut, 2008) about cost of coordination in influencing article quality. Model 2 adds roles' activity to the model and achieves a boost of .005 to the R-Squared. Examining this result in more detail suggests that more activity by substantive expert (.058) and less activity by Wiki gnomes (-.033) predicts of quality improvements. The value of substantive experts is that they add substantive information to an article. In contrast, Wiki gnomes contribute Wikipedia specific cleanup edits. This type of work may be unimportant to article quality or even detrimental, at least based on our automated measures. Alternatively, Wiki gnomes might be drawn to articles whose quality is declining because of the work of other editors.

To determine if the effect on quality of contribution by different roles depends upon the initially quality of the article, Model 3 adds the eight interaction terms between the previous quality score and the contribution of different roles (e.g., *Social Networker* × *Previous Score*). Again we see an improvement to the R-Squared, suggesting that the activities of different types of editors are needed at different stages of article development. The negative coefficient for *Substantive expert* × *Previous score* (-.139) suggests that, as articles increase in quality, the substantive content provided by substantive experts is needed less. In contrast, the positive coefficient for *Wiki gnomes* × *Previous score* (.021) suggests that, as articles increase in quality, activity the cleanup activities by Wiki gnomes become more important. Although one might have expected the cleanup work

done by copy editors, who in conventional publishing are most heavily involved in the final stages of manuscript production, would also become more important for higher quality, more complete articles, the negative coefficients disconfirm this conjecture.

4.7 Discussion and Conclusion

This paper focused on identifying editors' roles in Wikipedia and how the work contributed by editors in different roles affect the article quality. To achieve these goals we introduced a fine-grained taxonomy of edit types to characterize users' edits and built machine learning models to automatically identify the edit categories in each edit. We appropriated LDA-like graphical models to extracted latent roles from editors' history of edit activities. Finally, we examined the relationship between contributions of different types of editor to the improvement of article quality.

This research is an initial step in understanding the nature and value of social roles in online production and leaves much room for improvement, which we hope to address in future research. First, our role labeling is based on first identifying semantically meaningful edit types (e.g., adding information or paragraphing). The entire role modeling pipeline depends on creating an appropriate taxonomy of edit types, of accurately classifying each type of edit, of developing models that can account for each edit type. Each of these steps could be improved. Second, our role models take into account only the types of edits editors make and the namespaces where they work. Differentiating types of edits in other namespaces could be valuable (e.g., differentiating supportive versus critical comments in user and article talks pages (Zhu et al., 2011)). In addition, other features used by prior researchers should be included as input to the editor roles models, including user attributes, their social network signatures, users who edit multiple language editions (Hale, 2014), and the length of time spent editing (Geiger and Halfaker, 2013). Future work can extend ours by including a more comprehensive set of relevant features as input to latent role representation. Third, although our findings suggest eight informal editor roles, whether a role accurately represents an editor is not clear. A natural next step is to conduct surveys or interviews, which ask Wikipedians whether our descriptions of them are reasonable. Fourth, our measurement of article quality comes from Wikipedia's Article Quality Predictor. This predictor may be accurate enough in matching human judgments, and because the judgments it is attempting to match are those of committed Wikipedia editors, it may not reflect the characteristics

of articles that ordinary readers consider important to quality, such as the recency of the information cited or its accuracy.

We embarked on this research with the hope that automated identification of editors' roles would be useful in building recommender systems to better match editors to work. Although we have demonstrated the promise of social role modeling in Wikipedia, we believe that this approach could be applied to other online production communities, if they require a variety of skills from different contributors to be successful.

4.8 Reflection

This work mainly looked at a **single, simplified, and specific** context on Wikipedia - the context of editing main articles, because our expected roles are functioning *editing roles* that are essential for task-routing. As a result, we did not examine other available contexts on Wikipedia, such as the talk pages associated with articles (Maki et al., 2017; Ferschke et al., 2015). This work can be easily extended via an introduction of more contexts; for instance, including the context of topical areas may result in finer-grained editing roles, such as *substantive expert* × *biology* or *fact checker* × *politics*. Similar extensions also apply to the facet of **agent**. The profile attributes of editors such as gender or geo-location were not modeled, which are not necessary for functioning editing roles. We also did not investigate expectations since there are no explicit guidelines or norms associated with the derived roles. Despite lacking the modeling of several facets of roles, this empirical work reasonably demonstrates the effectiveness of our role identification framework and methodology.

Chapter 5

Identifying Semantic Edit Intention

We think in generalities, but we live in detail.

– Alfred North Whitehead

Most studies on human editing focus merely on syntactic revision operations, failing to capture the intentions behind revision changes, which are essential for identifying functioning roles that editors enact and facilitating the collaborative writing process. The present chapter models the facet of **Goal** in our role framework to improve role representation (postulation and definition), in order to potentially help the identification of roles occupied by editors on Wikipedia. It works as a complement to the modeling of editors' roles in Chapter 4. Specifically, in this work, we develop in collaboration with Wikipedia editors a 13-category taxonomy of the semantic intention behind edits in Wikipedia articles. Using labeled article edits, we build a computational classifier of intentions that achieved a micro-averaged F1 score of 0.621. That is, we predict editors' goals of editing via their observed behaviors (e.g., low level syntactic actions), because goals are manifested in the core characteristic behaviors of role holders. We further use this model to investigate edit intention effectiveness: how different types of edits predict the retention of newcomers and changes in the quality of articles, two key concerns for Wikipedia today. Our analysis shows that the types of edits that users make in their first session predict their subsequent survival as Wikipedia editors, and articles in different stages need different types of edits.

5.1 Introduction

Many online text production communities, including Wikipedia, maintain a history of revisions made by millions of participants. As Wikipedia statistics as of January 2017 show, English Wikipedia has 5.3 million articles with an average of 162.89 revisions per article, with revisions growing at a rate of about 2 revisions per second. This provides an amazing corpus for studying the types and effectiveness of revisions. Specifically, differences between revisions contain valuable information for modeling document quality or extracting users' expertise, and can additionally support various natural language processing (NLP) tasks such as sentence compression (Yamangil and Nelken, 2008), lexical simplification (Yatskar et al., 2010), information retrieval (Aji et al., 2010), textual entailment recognition (Zanzotto and Pennacchiotti, 2010), language bias detection (Recasens et al., 2013), spelling errors and paraphrases (Zesch, 2012; Max and Wisniewski, 2010).

To avoid building different approaches to extract the information needed by different NLP tasks (Ferschke et al., 2013), a unified framework to recognize edits from revisions is needed. A unified framework for identifying from revisions the types of edits people make in a variety of texts would simplify different natural language processing (NLP) tasks and improve comparability among them (Ferschke et al., 2013). Prior research on revision editing primarily develop syntactic edit action categories, from which they try to understand the effects of edits on meaning (Faigley and Witte, 1981; Yang et al., 2016a). For instance, Daxenberger and Gurevych (2012) categorized edits based on whether edits affect the text meaning, resulting in syntactic edit categories such as file deletion, reference modification, etc. However, simply understanding the syntactic revision operation types does not provide the information we seek: *why* do editors do what they do? *how effective* are their actions? For example, syntactic edit type taxonomies cannot tell the difference between simplifying a paragraph and maliciously damaging that paragraph, since both involve deleting a sentence.

In this work, we focus explicitly on revision intention. We introduce a fine-grained taxonomy of the reasons why an author in Wikipedia made an edit. Example edit intentions include copy editing, elaboration, verification, and simplification. Compared to taxonomies that either focus on low-level syntactic operations (Faigley and Witte, 1981) or that mix syntactic and semantic classes (Daxenberger and Gurevych, 2013), a

clean higher-level semantic categorization enables us to easily identify textual meaning changes, and to connect revisions to “what happens in the mind of the revising author during the revision” (Fitzgerald, 1987; Daxenberger, 2016). In order to capture the meaning behind edits, we worked with 13 Wikipedians to build a taxonomy that captured the meaning of an revision, which we term *edit intention*, and hand-labeled a corpus of 7,177 revisions with their edit intentions. We then developed an automated method to identify these edit intentions from differences between revisions of Wikipedia articles. To explore the utility of this taxonomy, we applied this model to better understand two important issues for Wikipedia: new editor retention and article quality. Specifically, we examined whether edit intentions in newcomers’ first editing sessions predict their retention, and examined how edits with different intentions lead to changes in article quality. These analyses showed that specific types of editing work were positively correlated with newcomer survival and articles in different stages of development benefited differently from different types of edits.

5.2 Related Work

Wikipedia revision histories have been used for a wide range of NLP tasks (Yamangil and Nelken, 2008; Aji et al., 2010; Zanzotto and Pennacchiotti, 2010; Ganter and Strube, 2009; Nelken and Yamangil, 2008). For instance, Yatskar et al. (2010) used Wikipedia comments associated with revisions to collect relevant edits for sentence simplification. Max and Wisniewski (2010) constructed a corpus of rewritings that can be used for spelling errors and paraphrases (Zesch, 2012). Similarly, Zanzotto and Pennacchiotti (2010) used edits as training data for textual entailment recognition, and Recasens et al. (2013) analyzed real instances of human edits designed to remove bias from Wikipedia articles. Most of these work employed manually defined rules or filters to collect relevant edits to the NLP task at hand.

Towards analyzing revisions and developing unified revision taxonomies (Bronner and Monz, 2012; Liu and Ram, 2009), Fong and Biuk-Aghai (2010) built machine learning models to distinguish between factual and fluency edits in revision histories. Faigley and Witte (1981) made a distinction between changes that affect meaning, called *text-base changes* and changes which do not affect meaning, called *surface changes*. The two categories are further divided into formal changes, meaning-preserving changes, micro-structure changes and macro-structure changes. This taxonomy was later extended by

Jones (2008) to take into account edit categories such as significant deletion, style, image insertion, revert, etc. Pfeil et al. (2006) proposed a 13-category taxonomy based on the data and performed manual annotation to compare cultural differences in the writing process in different versions of Wikipedia. Daxenberger and Gurevych (2013) introduced a finer-grained edit taxonomy, and performed multi-label classification to extract edit categories based on unparsed source text (Daxenberger and Gurevych, 2012). However, most taxonomies of edit categories contain only syntactic actions or a mixture of syntactic and semantic actions, failing to capturing the intention of revisions.

In terms of revision intentions, Zhang and Litman (2016) incorporated both argumentative writing features and surface changes from Faigley and Witte (1981) and constructed eight categories of revision *purposes*, such as claims, ideas, warrant, reasoning, backing, rebuttal, reservation, organization, clarify, etc. Tan and Lee (2014) used revisions to understand statement strength in academic writings. There are multiple works on the detection of specific subsets of revision intentions in Wikipedia, such as vandalism detection where the goal is to classify revisions as vandalized or non-vandalized (Harpalani et al., 2011; Adler et al., 2011) and language bias/neutral point of view detection (Recasens et al., 2013). Instead of recognizing a specific type of revision intention each time, our work aims at designing a systematic and comprehensive edit intention taxonomy to capture intentions behind textual changes.

Label	Description	α	Before	After
Clarification	Specify or explain an existing fact or meaning by example or discussion without adding new information	0.394	0.7%	4.1%
Copy Editing	Rephrase; improve grammar, spelling, tone, or punctuation	0.800	11.8%	14.8%
Counter Vandalism	Revert or otherwise; remove vandalism	0.879	1.9%	1.5%
Disambiguation	Relink from a disambiguation page to a specific page	0.401	0.3%	1.8%
Elaboration	Extend/add substantive new content; insert a fact or new meaningful assertion	0.733	12.0%	12.0%
Fact Update	Update numbers, dates, scores, episodes, status, etc. based on newly available information	0.744	5.5%	5.2%
Point of View	Rewrite using encyclopedic, neutral tone; remove bias; apply due weight	0.629	0.3%	2.2%
Process	Start/continue a wiki process workflow such as tagging an article with cleanup, merge or deletion notices	0.786	4.4%	5.8%
Refactoring	Restructure the article; move and rewrite content, without changing the meaning of it	0.737	1.9%	2.9%
Simplification	Reduce the complexity or breadth of discussion; may remove information	0.528	1.6%	4.6%
Vandalism	Deliberately attempt to damage the article	0.894	2.5%	2.0%
Verification	Add/modify references/citations; remove unverified text	0.797	5.4%	9.8%
Wikification	Format text to meet style guidelines, e.g. add links or remove them where necessary	0.664	33.1%	33.6%
Other	None of the above.	0.952	1.2%	-

Table 5.1: A taxonomy of edit intentions in Wikipedia revisions, Cronbach’s α agreement and the distributions of edit intention before and after corpus expansion.

Prior work also used edit types and intentions to better understand the process of collaborative writing, such as article quality improvement (Kittur and Kraut, 2008). For example, Liu and Ram (2009) found that Wikipedia article quality correlates with different types of contributors; similarly Yang et al. (2016a) pointed out articles in different quality stages need different types of editors. However, there are few studies examining the specific types of edits that are predictive of article quality. Recent research shows that the number of active contributors in Wikipedia has been steadily declining since 2007, and Halfaker et al. (2013) suggested that the semi-automated rejection of new editors' contributions is a key cause, but they did not explore whether or not specific types of newcomers' work got rejected at different rates and how that affects retention. In this paper, we take advantage of this new taxonomy to explore correlations between edit intentions, newcomers' retention, and article quality.

5.3 Semantic Taxonomy of Edit Intentions

A *revision* is created whenever an editor saves changes to a Wikipedia page. As one revision could contain multiple local changes, each revision can be labeled with one or more edit intentions, representing the purposes of why an editor made that change. Different from prior research (Daxenberger, 2016; Yang et al., 2016a), we do not distinguish between revisions and edits. Although an edit is a coherent local change and might belong to any edit categories, it cannot be used to represent the intentions of editors during the revision. For example, it might be difficult to recognize *Refactoring* if only one single edit is present. Since relocation or reorganization might involve several changes in the article, looking at one might lose the whole picture and lead to information loss. Moreover, edit types simply extracted from an edit is inadequate in outlining the correct intentions, for instance, adding a sentence could be *Clarification*, *Elaboration*, or *Vandalism*.

5.3.1 Taxonomy of Edit Intentions

Our semantic taxonomy of edit intentions builds on prior literature on collaborative writing (Faigley and Witte, 1981; Fitzgerald, 1987), research on document revision analyses (Bronner and Monz, 2012), studies on edit categories (Daxenberger and Gurevych, 2012; Fong and Biuk-Aghai, 2010), and work on purpose/intention classification (Zhang and Litman, 2016). In order to ensure that our taxonomy captured the *intentions* that

Wikipedians would find meaningful, we set up discussions with a group of 12 interested editors on a Wikipedia project talk page, and iteratively refined our taxonomy based on their feedback. Our discussion with Wikipedia editors is in this page¹. We also analyzed which intentions get more confused with which and used that to guide the refinement.

We define a top level layer for the revision intention taxonomy: intentions that are common in general revisions: **General Revision Intentions**, and intentions that are specific in Wikipedia: **Wikipedia Specific Intentions**. This categorization leads to 13 distinct semantic intentions, and Table 5.1 provides detailed descriptions. The percentage in each row represents what percentage of revisions are labeled with this edit intention. The percentages do not sum up to 100% because one revision could belong to multiple categories. The *After* corpus is used for all our analyses. Corpus size refers to the number of revisions.

Specifically, general revision intentions include: *Clarification*, *Copy Editing*, *Elaboration*, *Fact Update*, *Point of View*, *Refactoring*, *Simplification* and *verification*, and can be applicable to other contexts. *Counter Vandalism*, *Disambiguation*, *Process*, *Vandalism*, and *Wikification* are edit intentions related to Wikipedia. We also propose an *Other* category, intended for edits that cannot be labeled using the above taxonomy.

As the first work to model intentions of revisions, our taxonomy distills and extends existing edit type taxonomies. For instance, our intentions of “elaboration” and “verification” are extensions of “evidence” type proposed by (Zhang and Litman, 2016), and a syntactic category of “information deletion” in (Daxenberger and Gurevych, 2013) could be an instance of our “vandalism” or “simplification” depending on the context.

5.3.2 Corpus Construction

To construct a reliable, hand-coded dataset to serve as ground truth for automatic recognition of edit intentions, we employed four undergraduate students who had basic Wikipedia editing experience to label edits using our intention taxonomy, based on written annotation guidelines (see Appendix A) vetted by Wikipedia editors and provided examples². Moreover, to expose annotators to more working knowledge of Wikipedia,

¹https://en.wikipedia.org/wiki/Wikipedia_talk:Labels/Edit_types/Taxonomy

²https://en.wikipedia.org/wiki/Wikipedia:Labels/Edit_types/Examples

we provided three one-hour training sessions where annotators were asked to label a small set of revisions (around 50 each time) and to discuss their disagreements until consensus.

We randomly sampled 5,000 revisions from Jan, 2016 to June 2016 from the recent changes table³ in the Wikipedia database. For each revision, we displayed the content difference⁴ before and after the change to annotators, via a labeling interface that we developed. Because an editor could make several different types of edits within a single revision, we asked four RAs to label each revision with one or more of the possible semantic intentions. We collected four valid annotations for 4,977 revisions. We used Cronbach’s α , a measure of internal consistency, to evaluate agreement among the annotators. The overall agreement α score was 0.782, indicating substantial agreement between different annotators; The rule of thumb (Cortina, 1993) suggests that Cronbach’s alpha scores larger than 0.7 are considered as acceptable. The inter-annotator agreement per semantic intention is described in column α in Table 5.1.

5.3.3 Corpus Expansion

As shown in column *Before* in Table 5.1, some types of edit intentions, such as *disambiguation* and *clarification*, were very rare in the random-sample corpus. As a result, this corpus would not have enough positive examples on which to train a machine-learning model for some edit intentions. To address this under-representation problem, we used the text of editors’ comments to expand the corpus by retrieving 200 more revisions for each edit intention except Vandalism and Counter-Vandalism, resulting in 2,200 revisions⁵. More precisely, as a common practice (Zanzotto and Pennacchiotti, 2010; Recasens et al., 2013), we utilized regular expressions to match the text from the comments, which editors often wrote when saving their revisions, to the edit intentions. For example, editors might be signalling that they were intending to fix problems of *Point of View* when their comments contained keywords such as “npov” or “neutral”. Even though the comments sometimes signal the editors’ intents, they are not infallible, editors may fail to complete the comment field, may only label one of the multiple edit intentions for a single revision, or write comments that are inaccurate, irrelevant,

³https://www.mediawiki.org/wiki/Manual:Recentchanges_table

⁴en.wikipedia.org/wiki/?diff=712140761

⁵We used a practical and economic way to expand the corpus, and this made the intention distribution skewed away. We acknowledge this expansion as a limitation.

or incomplete. Thus the first author annotated the 2,200 revisions from the expanded corpus and merged it with the randomly sampled corpus. The frequency of the edit intentions before and after the expansion is in Table 5.1. We used the majority voting to resolve the disagreement. That is, if at least 3 out of 4 annotators picked an intention for a revision, it will be selected as the ground-truth. The final corpus contains 5,777 revisions, and can be downloaded from here⁶.

	Metric	Random	Majority	CMT	BR-	BR	MLKNN	RAKEL
Example	Exact Match	0.052	0.284	0.352	0.391	0.426	0.452	0.292
	Accuracy	0.052	0.283	0.428	0.498	0.540	0.542	0.338
	Precision	0.084	0.417	0.479	0.626	0.586	0.599	0.381
	Recall	0.052	0.285	0.458	0.562	0.611	0.578	0.344
	F1 Score	0.052	0.285	0.455	0.536	0.580	0.574	0.354
Label	Macro F1	0.060	0.042	0.310	0.487	0.597	0.576	0.385
	Micro F1	0.074	0.370	0.528	0.583	0.621	0.613	0.441
Ranking	One Error	0.920	0.583	0.415	0.400	0.358	0.320	0.434

Table 5.2: Performance comparison for predicting edit intentions from revisions. Best results are bold.

5.4 Identification of Edit Intentions

We frame automated identification of edit intentions as a multi-label classification task. We designed four sets of features for identifying edit intentions from revisions. Set I comprised two features associated with the **Editor**: *user registration* indicating whether the editor of a particular revision was registered or anonymous and *tenure*, which refers to the elapsed months between the current revision and editors’ registration date. Set II comprised 16 features associated with the **Comment** written by the editor to describe the revision, including *comment length* and a set of regular expressions to match intentions such as **pov**, **clarify**, **simplif**, **add link**, etc. Set III comprised 198 features associated with the **Revision Diff**, based on content differences between current revision and the previous one. They are similar to textual features defined in Daxenberger and Gurevych (2013), but we considered a wider range of objects be-

⁶http://www.cs.cmu.edu/~diyiy/data/edit_intention_dataset.csv

ing modified. In particular, we computed the difference in the number of characters, uppercase words, numeric chars, white-spaces, markups, Chinese/Japanese/Korean characters, HTML entity characters, URLs, punctuations, break characters, etc. We also considered languages features, such as the use of stop words, obscene words and informal words. Set IV comprises two features associated with **Vandalism** and **Revert**. We utilized the Wikipedia API to extract whether a revision was likely to be vandalism⁷ or reverting revisions⁸.

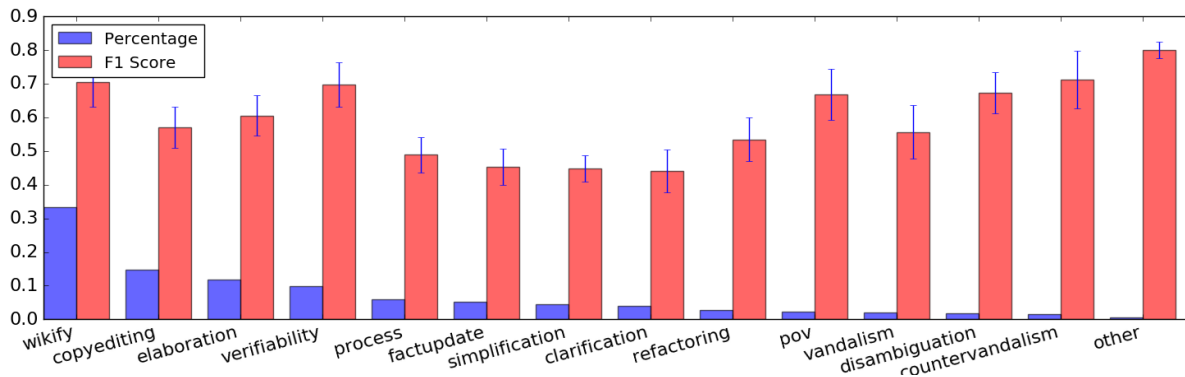


Figure 5.1: The relative frequency of each edit intention, and its F1 score provided by the **BR** model.

5.4.1 Identification Result

We extracted the input features with the help of Revision Scoring package⁹ and framed this task a multi-label classification problem. For multi-label classification, we considered solving them by using single-label classification algorithms and by transforming it into one or more single-label classification tasks. We used the multi-label classifiers implemented in Mulan (Tsoumakas et al., 2011), with 10-fold cross validation. We utilized Binary Relevance (**BR**) to convert our multi-label classification into 13 binary single-label problems. Similar to Daxenberger and Gurevych (2013); Yang et al. (2016a), we used Random k -labelsets **RAKEL** method that randomly chooses l small subset with k categories from the overall set of categories. We set l as 26, twice the size of the categories, and set k as 3. **MLKNN** method that classifies edit intentions based on K ($K=10$) nearest neighbor method. We used C4.5 decision tree classifiers in **BR** and

⁷<https://ores.wmflabs.org/v2/scores/enwiki/goodfaith/71076450>

⁸<http://pythonhosted.org/mwreverts/api.html>

⁹<http://pythonhosted.org/revscoring/>

RAKEL, as recommended by prior work (Daxenberger and Gurevych, 2013; Potthast et al., 2013). Prior research shows that sophisticated neural network models for text-classification largely rely on factors such as dataset size (Zhang et al., 2015; Joulin et al., 2016). Due to the size of our corpus and the complexity of this task, we did not use them.

To evaluate the relative accuracy of the multi-label classifier, we compared it to several baselines. The random baseline, denoted as **Random** in Table 2, assigns labels randomly. The majority category baseline, denoted as **Majority**, assigns all edits the most frequent intention, elaboration. Since revision comments may be especially as informative in reflecting edit intentions, the comment baseline, denoted as **CMT**, is a Binary Relevance classifier that includes only the comments features from Set II. We also created a Binary Relevance classifier, denoted as **BR-**, which excludes comment features and only used features from Sets I, III and IV.

Table 5.2 shows the evaluation metrics for the baselines and our multi-label classifiers. The metrics include the Exact Match subset accuracy, which evaluates whether the predicted labels are the same as the actual labels. These classifiers are available upon request. Table 5.2 also shows example-based measures of Accuracy, Precision, Recall and F1 Score, weighting each edit equally. It also shows label-based measures of accuracy – the micro- and macro-averaged F1 scores– which weight each edit intention category equally. As a ranking based measure, we measured One Error, which evaluates how many times the top ranked predicted intention is not in the set of true labels of the instance.

Results show that the Binary Relevance (BR) and MLKNN classifiers, which used all our constructed features, outperformed Random and Majority baselines. Moreover, the BR and MLKNN methods show relatively similar best performances. Although multiple studies have utilized revisions’ comments as “groundtruth” to collect desired edits, the CMT method, which includes only comment features, is less accurate than either the BR or MLKNN models. Note that predicting 14-category semantic intentions is more challenging compared to classifying low-level syntactic actions, such as inserting an image (Daxenberger and Gurevych, 2013). The code for edit intention classifications is publicly available at here¹⁰.

¹⁰https://github.com/diyiy/Wiki_Semantic_Intention

5.5 Intentions, Survival and Quality

The automated measurement of edit intentions provides a general framework to analyze revisions and can facilitate a wide range of applications, such as collecting specific types of revisions (Yatskar et al., 2010; Recasens et al., 2013; Zanzotto and Pennacchiotti, 2010) and outlining the evolution of author roles (Arazy et al., 2015; Yang et al., 2016a). In this section, we demonstrate two examples of how this intention taxonomy can be applied to better understand the success of online collaboration communities (Kraut et al., 2010), specifically the process of these sites to retain new contributors and create innovative products. To this end, we first investigate what newcomers are intended for in their first sessions and whether their edit intentions can account for their survival in Wikipedia. We then examine how edits carrying on different intentions at distinct times in an article’s history influence changes in its quality.

5.5.1 How Edit Intentions Affect Survival

To explore newcomers’ intentions during their first experience editing articles, we focus on users’ first edit sessions in Wikipedia. Here, **Edit Session** is defined as a sequence of edits performed by a registered user with less than one hour’s time gap between two adjacent edits (Halfaker et al., 2013). We then compare edit intentions of newcomers who survive - **Survivors**, and newcomers who do not - **Non-survivors**. Here, newcomers are defined as surviving if they performed an edit at least two months after their first edit session.

Intention Comparison

Among 100,000 randomly sampled Wikipedia users, 21,096 made revisions in the main article namespace during their first editing session. Among these 4,407 were survivors (i.e., made an edit two months after registering) and 16,689 were non-survivors. We applied our edit intention model to 53,248 revisions in users’ first sessions, and compared the percentages of different types of edit intentions between survivors and non-survivors, as shown in *Intention Dist* column in Table 5.3. We also performed 1-way ANOVA to test whether survivors and non-survivors have the same mean for each edit intention. We observed that, survivors tend to do more copy-editing ($\Delta_+ = 2.3\%$) and more wikification ($\Delta_+ = 6.5\%$), while non-survivors seem to perform more simplification and vandalism, which might provide signals for detecting vandals.

Edit Intention	<i>Intention Dist</i>		<i>Revert Ratio</i>	
	NS	SS	NS	SS
clarification	0.2%	0.4%	0.1%	0.1%
copy editing	12.1%	14.4%	6.9%	3.8%
counter vandalism	0.1%	0.0%	0.1%	0.0%
disambiguation	0.0%	0.0%	0.0%	0.0%
elaboration	27.7%	26.5%	16.5%	6.9%
fact update	4.2%	3.8%	3.4%	1.7%
neutral point of view	0.1%	0.2%	0.0%	0.1%
process	2.0%	2.3%	1.9%	0.7%
refactoring	1.1%	1.3%	0.9%	0.5%
simplification	3.7%	3.1%	3.1%	1.4%
vandalism	13.8%	6.1%	16.0%	4.7%
verification	7.0%	7.4%	3.8%	2.7%
wikification	25.8%	32.3%	14.0%	6.9%

Table 5.3: The edit intention distribution in the first sessions (Intention Dist) and the revert ratio comparison (Revert Ratio), among non-survivors (NS) and survivors (SS). The numbers are bolded if 1-way ANOVA tests for difference between two groups are significant, with $p < 0.05$.

Revert Analysis

To explore the relationship between rejection of contributions and newcomer retention, we also visualized the revert ratios of different types of edit intentions for survivors and non-survivors in their first session. Here, **Revert** refers to whether an edit from the author was reverted or completely removed by another user, and we detect reverts using MediaWiki Reverts library¹¹. We then measured the revert ratio for each edit intention by calculating the percentage of revisions belonging to a specific edit intention, among all reverted revisions in users' first sessions. As shown in the *Revert Ratio* column in Table 5.3, in general, non-survivors get reverted more compared to survivors, across all edit intentions. Interestingly, non-survivors compared to survivors get reverted more when performing *Wikification*, *verification* and *Refactoring*, suggesting that sophisticated types of work might not be suitable for beginners.

¹¹<http://pythonhosted.org/mwreverts/>

Edit Intention	Survival	Quality Changes
clarification	0.029	0.001
copy editing	0.033	0.011 [†]
counter vandalism	0.004	-0.020 [†]
disambiguation	-0.003	-0.006 [†]
elaboration	-0.024	0.061 [†]
fact update	-0.001	0.002
point of view	0.041	-0.003
process	0.051 [†]	-0.024 [†]
refactoring	-0.013	0.011 [†]
simplification	-0.002	-0.008 [†]
vandalism	-0.211 [†]	-0.005 [†]
verification	0.047	0.068 [†]
wikification	0.099 [†]	-0.010 [†]

Table 5.4: Regression coefficients of different edit intentions for predicting Newcomer **Survival** and Article **Quality Changes**. Here, † means the coefficient is statistically significant ($p < 0.05$)

Newcomer Survival

As a further exploration of the relationship between edit intentions and newcomer survival, we performed a logistic regression using edits in survivors’ and non-survivors’ first sessions. To handle this imbalanced data (i.e., many more negative examples than positive examples in training), we performed majority-class under-sampling to make this dataset balanced. Similar to [Halfaker et al. \(2013\)](#), we controlled the number of revisions completed during the first session (a proxy for an editor’s initial investment), and the number of revisions reverted in their first sessions. This logistic model boosted the McFadden’s Pseudo R-squared from 0.025 (simply using the two control variables) to 0.051. We described the regression coefficients of statistically significant edit intentions in the **Survival** column of Table 5.4. This logistic model achieves an Accuracy of 60.98%, Recall of 58.30%, Precision of 78.08% and F1-score of 66.76%. Editing articles for the purposes of *Process*, *Verification* and *Wikification* significantly predict the survival of newcomers, while performing vandalism is a strong negative predictor for survival.

5.5.2 How Intentions Affect Article Quality

Although there are over 5.5 million articles in the English Wikipedia, fewer than 0.2% have been evaluated by Wikipedians as good articles and around 92% have been evaluated as start or stub class articles, Wikipedia’s two lowest quality categories. In this section, we examine how edits with different intentions at distinct times in an article’s history influence changes in its quality.

This task is framed as a prediction task, i.e. using edits’ intentions and a set of control variables to predict changes in article quality. We borrowed a Article Quality Prediction Dataset released in [Yang et al. \(2016a\)](#), which consists of the quality ratings collected in January and June, 2015 of 151,452 articles. We collected 1,623,446 revisions made to these articles between January and June 2015, by randomly sampling 10% revisions that were made to these articles during that time periods. Specifically, the outcome *article quality change* is calculated by subtracting the previous quality score from the end quality score. The control variables include the previous article quality score, the total number of edits, the total number of editors, the changed bytes to an article, and the total number of edits to the article talk page during the six months. To construct edit-intention predictors, we summed the number of edits for each edit intention during the six months divided by the total number of revisions in this article.

Results of the linear regression model, shown in **Quality Changes** column of Table 5.4, show that our constructed regression model is significantly predictive of article quality changes ($R^2 = 0.225$). The results show that, keeping all control variables fixed, more *Copy Editing*, *Elaboration*, *Refactoring* and *Verification* are positively associated with improvements in article quality; in contrast, *Vandalism*, *Counter Vandalism*, *Disambiguation*, *Process* and *Simplification* predict declines in article quality. The first four of these edits types often occur with reducing the article content, removing or redirecting pages. Improper use of them might be detrimental to article quality.

To determine if the effect of edit intentions on quality changes depends upon the initial quality of the article, we added the interaction terms between the previous quality score and edit percentages of different intentions (e.g., clarification x previous quality), and visualized interaction effects in Figure 5.2. When examining the interaction terms in more detail: the negative slope of *copy editing* (when prev=2) suggests that, as articles

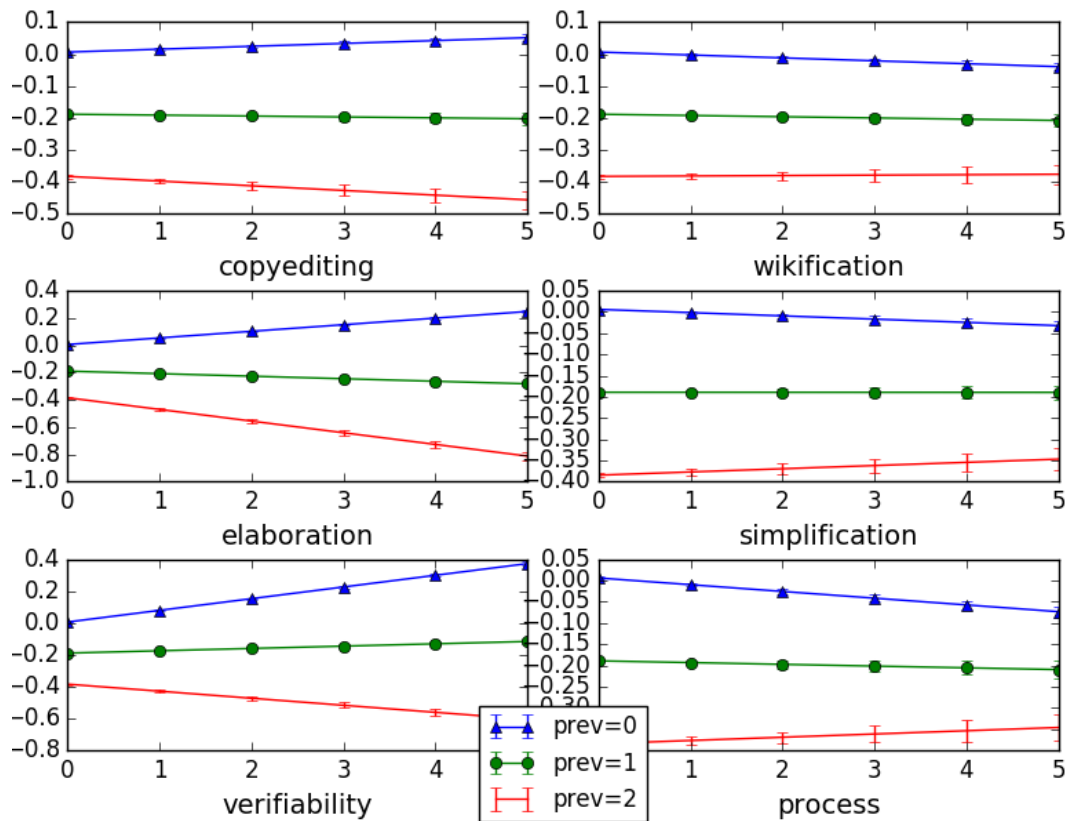


Figure 5.2: Interaction effect of different levels of edit intentions and different levels of previous article quality (**prev**) on article quality changes. All variables are standardized. The Y-axis measures the predictive margins and X-axis refers to different standardized levels of edit intention.

increase in quality, copy editing is needed less. We found similar trends for interactions between previous quality and *elaboration* and *verification*, which are essential for articles in the starting stages. In contrast, the positive slopes for *simplification*, *wikification* and *process* suggest that, as articles increase in quality, simplifying articles' content, adding proper links or reorganizing structures becomes more important. Overall, these results reveal that different types of intentions are needed at different quality stages of articles.

5.6 Discussion and Conclusion

In this work, we proposed 13 semantic intentions that motivate editors' revisions in English Wikipedia. Example edit intentions include copy editing, elaboration, simplification, etc. Based in a labeled corpus of revisions, we developed machine-learning

models to automatically identify these edit intentions. We then examine the relations between edit intentions, newcomers survival, and article quality improvement. We found that (1) survivors tend to do more copy editing and wikification; non-survivors seem to perform more vandalism and other sophisticated types of work, and the latter often gets reverted more; (2) Different types of contributions are needed by articles in different quality stages, with elaboration and verification are needed more for articles in the starting stages, and simplification and process become more important as article quality increases.

Our proposed edit intention taxonomy and the constructed corpus can facilitate a set of downstream NLP applications. First, classifiers based on this intention taxonomy can help retrieve large scale and high quality revisions around simplification, neutral point of view or copy editing, which provides amazing corpora for studying lexical simplification, language bias detection and paraphrases. Second, as we showed in Section 5.2, determining how different edit types influence changes in articles is of great use to better the causes of quality variance in collaborative writing, such as detecting quality flaws ([Anderka et al., 2012](#)) and providing insights on which specific aspects of an article needs improvement and what type of work should be performed. The ability to identify the need for editing, and specifically the types of editing work required, can greatly assist not only collaborative writing but also individual improvement of text. Moreover, even though our edit taxonomy is for English Wikipedia, it can be applied to other language versions of Wikipedia. We are now deploying the same edit intention taxonomy for Italian Wikipedia, and plan to apply it to other low resourced languages in Wikipedia. Finally, beyond the context of Wikipedia, similar taxonomies can be designed for analyzing the collaboration and interaction happened in other online contexts such as academic writing (e.g., Google Docs or ShareLatex, etc).

5.7 Reflection

This work investigates the facet of **Goal**, i.e., the intentions of why editors made their edits, which can potentially improve role definition to better facilitate role discovery. Note that we predicted editors' goals of editing via their observed behaviors (e.g., low-level syntactic actions) because goals are manifested in the core characteristic behaviors of role holders. In other words, it is impossible to tease apart goals from behavioral patterns. As we demonstrated in the present chapter, the same set of syntactic edits

may suggest a different set of editing goals, but one specific edit intention usually relates to a similar collection of behaviors. Explicitly inferring goals from behaviors can make the process of role identification more interpretable from a basic scientific role understanding perspective, compared to enumerating individuals' all kinds of behavioral regularities in a flat manner. An interesting research question for future work is to combine both behavioral patterns (i.e., edit type taxonomy) and goals (i.e., edit intention) for role identification, which may reduce too much subjective interpretation of role behaviors and paint a more nuanced picture of for what purposes individuals embody specific roles.

Part II

Role Identification on Cancer Survivors Network

Chapter 6

Identifying Roles on CSN

The self is not something ready-made, but something in continuous formation through choice of action.

– John Dewey

Participants in online communities often enact a variety of social roles in the process of helping their communities and the public at large. For example, in cancer support communities, some users specialize in providing information about a specific type of disease or and some specialize in socializing new members. While prior studies have described community trends in aggregate, the current work operationalizes behavioral patterns of users of a cancer support community into specific functional roles. This chapter introduces a systemic empirical method for automatically inferring members' functional roles when participating in online health communities. In contrast to the studies in Chapter 4 and Chapter 5, this work examines the problem of role identification in a new context - Cancer Survivors Network. In this chapter We will demonstrate how to utilize prior work on online health communities to provide guidance and supervision for role-related behaviors — **role postulation**, to operationalize multiple facets of our role framework (**Interaction, Goal, Context, Expectation**) — **role definition**, to extract a set of coherent roles that can well explain participants' behaviors — **role identification**, and to validate the derived roles comprehensively via four evaluation methods: quantitatively in terms of model fit on the held-out data, qualitatively in terms of domain experts' interpretation, directly in terms of correlations with input from role holders, and indirectly in terms of performance boost to downstream applications — **role evaluation**.

Building on a series of comprehensive validation, we identify eleven roles that members occupy such as *emotional support provider*, *welcomer*, and *story sharer*. We investigate member role dynamics interacting with long-term participation and dropout in the community, and describe how roles change as part of a member's life-cycle. We found that early assumption of certain roles such as *welcomer* are predictive of members' continued participation, and members frequently change their roles over time from seeking resources to roles that offer help to others; resource-seeking roles are typical of newcomers who are likely to leave the community, while help-offering roles are taken up over time and typical of long-time members. Our methodology is foundational to identifying members' roles in communities early, and facilitating better use of their skills and interests in support of community-building efforts.

6.1 Introduction

A wide body of literature studying online health communities has developed and tested hypotheses on how these communities differ from the internet at large, how users support each other, and how communities thrive over time. For example, [Wang et al. \(2012\)](#) studied how social support exchange in an online cancer support group affects the length of people's participation, and [Chancellor et al. \(2018\)](#) examined support exchange around behavior changes in online weight loss communities. Using descriptive statistical models, this research modeled characteristics of user behavior to show that early actions result in differential long-term membership trends. For instance, users self-disclose more personal information in online health communities than in parallel technical support communities, like Stack Overflow ([Balani and De Choudhury, 2015](#); [Mayfield et al., 2012](#)). Not all users display these behaviors, though: for instance, many users join when facing crucial healthcare events, like the start of chemotherapy, and are seeking information for decision-making rather than hoping to join a community ([Wen and Rosé, 2012](#)). Early actions and interactions can be predictive of commitment. Newcomers looking for informational support are significantly less likely to transition into long-term community membership, and those who receive support are more likely to continue than those who do not ([Wang et al., 2012](#); [Yang et al., 2017b](#)). Yet 10% of support-seeking messages get no replies, and many of the replies do not provide the support sought, as when long-time members provide emotional support when the new user was seeking information ([Wang et al., 2015](#)).

Interaction in health support communities is in part the products of the *roles* that members occupy (Stewart et al., 2005). For example, some members might specialize in seeking support, providing disease-related information or socializing new members. In contrast to roles in conventional organizations, where roles are often assigned and come with defined responsibilities, roles in most online communities are emergent. For example, a user can assume an “expert” role in the community without seeking permission from others. Researchers have clustered lower-level behavior to identify roles in some online communities like Wikipedia (Welser et al., 2011; Yang et al., 2016a). However, few studies have applied similar approaches to online health communities (Jones et al., 2011a).

The goal of the current paper is to study members’ participation and coordination in online health communities, and develop a taxonomy of the emergent roles that are observed in these communities, linking individual behaviors with community-level outcomes. Identifying emergent roles can be beneficial for sustaining communities. Understanding the roles that are important for a community and the roles particular people are likely to occupy can help to optimize user experiences. For example, information experts can be matched to information seekers, giving the expert fulfilling work to do while helping the seeker get timely responses; welcomers can be matched to newcomers to ensure they receive timely support that will help them become integrated into the community.

To this end, we follow the role framework proposed in Chapter 2 for defining social roles in online communities together with a general modeling methodology from Chapter 3. We use data from an online cancer support community to identify behavioral features associated with different facets of social roles. We then build an unsupervised Gaussian mixture model from the data to discover 11 roles that members occupy. We validate these roles through a series of quantitative robustness checks of the modeling procedure, followed by confirmatory interviews with domain experts in the community.

To demonstrate the utility of the role model, we examine how roles predict the stability of activities on the site and participation by users as they enter the community and evolve from being newcomers to old-timers. (1) We find that occupying socially positive roles, such as private communicator and story sharer, is associated with members staying in the community longer, while members occupying roles such as informational

support seeker are associated with lower long-term participation in the community. (2) While the distribution of roles in the community is relatively stable over time, members change their roles frequently across their participation. As members stay longer in the community, they are more likely to occupy the roles of emotional support provider and welcomer and less likely to occupy roles such as story sharer and informational support seeker. A closer look at members' role transitions suggests that they frequently change their roles from seeking resources to roles that offer help to others. (3) Both the tendency of certain roles' occupants to drop out of the community and the trajectory of roles in users' lifecycle in the community follow consistent patterns. These findings suggest the value of the role framework as the basis for intervention in online health communities, opening a new opportunity for socio-technical systems to support users and communities in their healthcare needs.

6.2 Research Question

The current research investigates members' emergent, behavioral roles when participating in online health communities independently of the demographics of the people who occupy them. For example, any member can assume the role of emotional support provider, no matter their gender, age or cancer type. Our goal is to design a model that can ultimately be deployed in online interventions, in environments where both technical constraints and user privacy dictate that demographics should not be a factor in the technical system. Thus, we do not model personal attributes of members in our research. Future studies in constrained, privacy-aware contexts may extend this work to directly cross the behavioral roles identified with some of members' personal attributes (e.g., informational support provider \times cancer type).

6.3 Research Site

Our research was conducted on the American Cancer Society's Cancer Survivor Network¹ (CSN), which is the largest online support community for people suffering from cancer and their caregivers. The CSN discussions boards are public places where registered members can participate by starting new threads or commenting on other members' existing threads. Registered members of CSN can also communicate directly with

¹<https://csn.cancer.org/>

each other using a function called “CSN Email”. Conversations between two people are recorded in a format like email or private chat messages and are only visible to individuals addressed in the message headers. We were provided access to all public posts and comments, private chats as well as the profile information for users registered between Dec 2003 and Mar 2018. During this period, there were a total of 66,246 registered users who exchanged 139,807 private messages, 1,080,260 comments and 141,122 threads. This work was approved by Carnegie Mellon University’s Institutional Review Board (IRB).

6.4 Generative Model for Role Identification

Our method of identifying emergent social roles in online communities is *a repeated cycle of role postulation, definition, automated processing and evaluation*. When participating in the community, a user takes on one or more implicit roles for their activities. In their future interactions, they may take on the same roles or shift roles. To model this, we define a Gaussian mixture model (McLachlan and Basford, 1988), a statistical model that clusters heterogeneous user-session representations into a set of coherent, discovered user roles. Unlike traditional unsupervised learning such as k -means clustering, in which an object can only be a member of a single cluster, a mixture model allows users to occupy multiple roles during a session (e.g., a welcomer and information provider).

The model assumes that user activities can be described by a set of observable behaviors X , and there exist k components per role $\{c_{i=1}^k\}$. Each component c_i has an associated vector μ_i of average values for each feature in X . A user’s activity is generated from a mixture of these components and a covariance matrix Σ_i , representing the likelihood of each role co-occurring with each other role. Formally, Gaussian Mixture models are a linear combination of Gaussians, with a probability density function as follows:

$$p(x) = \sum_{k=1}^K \pi_k \cdot N(x|\mu_k, \Sigma_k), \text{ where } \sum_k \pi_k = 1$$

Here, $\{\pi_{i=1}^K\}$ are called mixing coefficients, and each user will be assigned a coefficient π_i for each role c_i . The coefficient represents the proportion of a user that was associated with a particular role; each user unit is modeled as a mixture of roles, which enables us to capture participants’ versatility and dynamics in the online community. When building this model, we need to learn mixing parameters $\{\pi_1, \pi_2, \dots, \pi_K\}$, means

$\{\mu_1, \mu_2, \dots, \mu_K\}$ and covariances $\{\Sigma_1, \Sigma_2, \dots, \Sigma_K\}$ from data $\{x_i\}_{i=1}^N$. Here, each x_i is a heterogeneous vector of features extracted from each user, while N represents the total number of user units in our corpus. Given a large corpus of data, we can estimate the covariance matrices by positing that each component has its own general covariance matrix.

This model has three key parameters that need to be set by researchers: the behavior features X , the length of user representation l , and the number of implicit roles K . Each is an aspect of the model that is susceptible to over-fitting. In the following, we describe the procedures used to set each parameter and the steps taken to design robust models.

6.4.1 Operationalizing Behavioral Features

To extract the emergent roles that members take on when participating on CSN, we identified a set of behavioral features that operationalize the four components in the theory-driven framework of role definition described in Chapter 2, including *goal, interaction, expectation* and *context*.

Recently, deep learning based techniques have been proposed to learn user embeddings based on their interactions in an end-to-end manner (Hamilton et al., 2017; Ribeiro et al., 2017; Henderson et al., 2012). Although that approach requires less domain knowledge and manual feature construction, it suffers from lack of interpretability especially about the nature of discovered roles and the people who occupy them. In terms of techniques for identifying social roles online, most research employed clustering analysis or principal component analysis to cluster each user into one or more clusters (Welser et al., 2011; Yang et al., 2016a). To make the derived roles interpretable, we followed this common practice to construct explainable patterns to capture members' role-relevant behaviors.

Goal (11 features)

Many people with chronic illnesses, including cancer patients and survivors, participate in online health support groups. Ridings and Gefen found that 76% of people who joined online health groups were looking for two types of social support (Ridings and Gefen, 2004) - informational support and emotional support. Informational support contains information, advice, or knowledge, and emotional support refers to the pro-

vision of empathy, sympathy or encouragement. Building on prior studies on social support (Wang et al., 2012; Biyani et al., 2014), we operationalized a set of goal-oriented actions that members exchange in the context of support groups. This resulted in 4 features of linguistic behaviors: *seeking informational support*, *providing informational support*, *seeking emotional support*, and *providing emotional support*.

We observed from our data that people tend to employ very specific language strategies when providing emotional support to others. Some choose to show empathy, saying that they understand what the recipient is going through and identify with their emotional reactions and feelings. Some express encouragement and hope that others' situations will improve. Others show appreciation for others' accomplishments to increase others' senses of worth, value and competence. To capture these nuanced intentions, we differentiated three finer-grained sub-categories of providing emotional support: *providing empathy*, *providing encouragement*, and *providing appreciation*. In addition to exchanging social support, members also share their experiences and stories to help others understand who they are and to provide social comparison information (De Choudhury and De, 2014). Thus, we also considered the language people use to self-disclose via two additional features: *self-disclosing positively* and *self-disclosing negatively*.

We described the definitions and examples of those nine conversational acts in detail in Table 6.1. Automatic text analysis techniques can accurately measure the amount members' messages contain each of these nine features. Four trained nursing students rated a sample of 1,000 messages threads and their first responses for degree they represented these nine goal-oriented conversational acts. Using previously developed procedures (Biyani et al., 2014; Wang et al., 2012), we built machine learning models to predict the students' assessments of the nine conversational acts in messages. These machine learning models map a set of linguistic features, as described in (Wang et al., 2012; Yang et al., 2017c), to a set of continuous output values, indicating how much informational support, emotional support, positive self-disclosure, and negative self-disclosure a thread-starting message conveys as well as how much informational support, emotional support, empathy, encouragement, appreciation, positive self-disclosure, and negative self-disclosure responses provided. Human annotation agreement on a training dataset was high (mean ICC=.84), and the machine learning models achieved reasonable correlation with the average of the human judgments (mean Pearson correlation

Conversation Acts	Definition and Examples
seeking informational support	Seek information, advice, referrals or knowledge. <i>"I was wondering if anyone who has had whole brain radiation has had hair not grow, back on head ?"</i>
providing informational support	Provide informational support to the person starting the thread. <i>"It was explained to me that microcalcifications look like as if one were to throw rock salt on a blacktop driveway and they would 'cluster and fall' in many locations"</i>
seeking emotional support	Seek understanding, encouragement, sympathy or caring. <i>"So, much of the stuff I find on the web is 'doom and gloom'. Would love to hear from some long-term, survivors!!!! Mainly cuz I'm scared, out of my wits about all this - any thoughts?"</i>
providing emotional support	Provide emotional support <i>"I do understand the frustration and anger and sadness of having drugs fail you and then venturing forth on unknown territory yet again. This whole journey is fraught with crappy bumps and turns. wish you the best."</i>
providing empathy	Express empathy that he or she understands what the recipient is going through and identifies with his/her emotions and feelings. <i>"We're so very, very sorry you're now a member of the club that NO ONE wants to join."</i>
providing encouragement	Express hope that situations will improve or support someone in their efforts when facing challenges. <i>"I want to make sure you know that i am with you. Keep the faith. We're all pulling for you."</i>
providing appreciation	Express appreciation and provide support for someone's a sense of worth, value, and competence. <i>"You have had such a difficult road, but yet manage to do well in school. I'm truly inspired by you."</i>
self-disclosing positively	Discuss positive thoughts or emotions, such as gratitude and love. <i>"My family is so supportive and makes me feel like such a loved person."</i>
self-disclosing negatively	Discuss negative thoughts or emotions, such as worry or anger. <i>"I am freaked out after reading my mammogram report"</i>

Table 6.1: Definitions and examples of nine goal-oriented conversational acts.

Goal-oriented conversational acts	ICC	Correlation
seeking informational support	0.91	0.73
providing informational support	0.92	0.79
seeking emotional support	0.83	0.64
providing emotional support	0.92	0.75
providing empathy	0.74	0.72
providing encouragement	0.68	0.64
providing appreciation	0.73	0.67
self-disclosing positively	0.90	0.72
self-disclosing negatively	0.90	0.71

Table 6.2: The intra-class correlation and correlations between human decisions and predictions for 9 conversational acts

$r=.71$; see Table 6.2). We then applied these models to estimate the nine conversational acts in all messages in our corpus.

Separate from these automatic annotations, we also extracted 2 features measuring raw activity count for users - the number of threads initialized, and the number of comments.

Interaction (53 features)

The actions members take toward achieving their goals are essential for understanding the roles they occupy. In this part we use two methodologies to extract interaction features: *linguistic* and *network-based*.

We developed linguistic indicators of members’ topical interests by comparing each person’s word usage with semantic categories provided by the psycho-linguistic lexicon LIWC (Pennebaker et al., 2015). The presences of affective expressions such as anxiety, sadness, or anger related words, were used as indicators of members’ emotional orientation. To figure out whether members talked about their personal relationships, we counted their usage of words related to family and friends via corresponding dictionaries in LIWC. Similarly, members’ religious orientations and emphasis on themselves vs others (interpersonal pronouns) were calculated via related dictionaries. In total, 16 features were extracted via using corresponding LIWC categories. Topic modeling (Blei

et al., 2003) was conducted to derive topics that members discuss with others on CSN, resulting in 25 topics including prayer, surgery, radiation, clinical trials, and chemotherapy side effects. One feature is included for each topic. We also incorporated domain knowledge from Freebase to capture 4 features counting members' use of words related to disease, medicine, ingredients, and symptoms in their messages when providing information to others. To identify potentially knowledgeable members, we extracted 2 features: the number of external links and the number of words in messages.

We then looked at interaction patterns that emerge from users' social networks in the online community. Previous studies demonstrated methods for revealing network structure and people's relationships with other users (Fisher et al., 2006; Welser et al., 2011, 2007). For this purpose, we constructed a user-reply network and extracted features through network analysis, where the vertices represent members who have participated in at least one messages, and edges represent replies. For example, an edge from user u to user v means that u replied to v 's messages. From this graph, we extracted six network-based features: (1) To capture the centrality of members' role in the social structure, we calculated their (1) in-degree and (2) out-degrees. To capture tenure effects we measured (3) members' ratio of talking to newcomers and (4) being talked to by old-timers. Here, newcomers are defined as people who have stayed at CSN for less than a month. (5) To measure whether users talk mainly to several specific users or broader audiences, we calculated the entropy of the user-user interaction distribution. Here, a higher entropy means users talking to broader audiences. Finally, to measure a user's breadth of interests, we measured the number of sub-forums a person has posted in, where each sub-forum represents one cancer type.

Expectation (2 features)

Emergent roles may be associated with informal implicit "negotiated understandings" among individuals about what persons should do if they seem to occupy such roles. Members on CSN might indicate such positive or negative evaluations of others via their language choices such as complaining to administrators or telling others what to do. To this end, we extracted two features: (1) the number of messages members exchanged with moderators and (2) their usage of modal words such as "should", "could", and "must". Here, modality in members' messages may convey their suggestions, request or advice to others.

Context (17 features)

The context of communication matters. For the purposes of this study, we focused on public vs private conversations as the context. Members may talk to others in private chats to protect their personal information or interact with them on the public discussion board. To capture members' potential concerns of privacy, we differentiated all 9 *Goal* features and their 6 network-based *Interaction* features into separate values for communication in private chats and in the public forum. For example, *seek informational support* will have two features: *seek informational support in private chats*² and *seek informational support in the forum*. Similarly, *being talked to by oldtimers* becomes *being talked to by oldtimers in private chats* and *being talked to by oldtimers in the forum*. Note that this domain differentiation is a common practice in text representation for statistical modeling (McCallumzy et al., 1999) as well as in social computing research (Bazarova et al., 2015, 2013). Finally, we calculated 1 feature that measures the ratio of members' private communication to all their private and public activities to capture their preferences for different contexts.

This will result in a total vector of 83 $(9+2) + (16+25+4+2+6) + (2) + (9+2+6+1)$ features representing each user unit. These features are held constant for all further analyses in this work.

6.4.2 Determining the Granularity of User Activity

Determining the unit of analysis for appropriately representing members' activity is key decision in modeling social roles. Treating users as an aggregation of all their historical actions on CSN prevents one from examining the evolution of roles or transitions between them. On the other hand, employing very small time intervals, such as a single user action, might miss important larger constructs like a cluster of actions needed to achieve a goal.

In this analysis we use aggregated data from each **user session**, which is defined as a time interval in which the time gap between any two adjacent actions is less than a

²Note that for privacy concerns, annotators are not allowed to view and annotate private messages. In these cases, we applied the trained regression models from public forum posts to predict 9 conversational acts in private messages. Accuracy may be lower in these contexts, as this prediction requires transferring the model to a slightly different domain.

threshold (24 hours). Within sessions, users' behaviors were regarded as consistent. We operationalized the 83 features described above to capture members' behaviors within each session.

To test the robustness of the role models, we explored the degree to which they varied across different temporal units—all activity within each calendar day, week, or month. We found that frequently-occurring roles were consistent across different settings. The roles that emerged using a calendar day as the unit of analysis were very similar model to those emerging from session-level modeling, likely due to the similar time-scale. As the temporal unit increased from a day to a week to a month, the derived roles became harder to interpret. This suggests that unlike assigned roles in offline organizations (e.g., professor in a university), emergent roles in this community are more variable over time. This variability led us to examine transitions between roles, described in more detail below.

Role theory also states that role are based on multiple interactions (Turner, 1990), suggesting that detection of roles based on only one observed action is impossible. To address this, we conducted a sensitivity analysis removing sessions that had fewer than t actions ($t \in \{1, 2, 3\}$). We did not observe any significant changes in the derived roles. For all analyses below, we follow the 24-hour inactivity threshold to define sessions and include all sessions, without removing ones with few actions. In total, this resulted in 517,272 user-sessions from 66,246 users.

Quantitatively, in Figure 6.1, we describe the number of members in a log scale who have a certain number of user-sessions under our 24-hour inactivity threshold (left), the number of members in a log scale who occupy a certain number of roles (middle), and the number of user-sessions that involve in a specific number of roles (right). We find that members tend to occupy no more than three roles within each session, probably because user sessions are relatively short and cleanly-defined time intervals that are subject to the appropriate amount of variation.

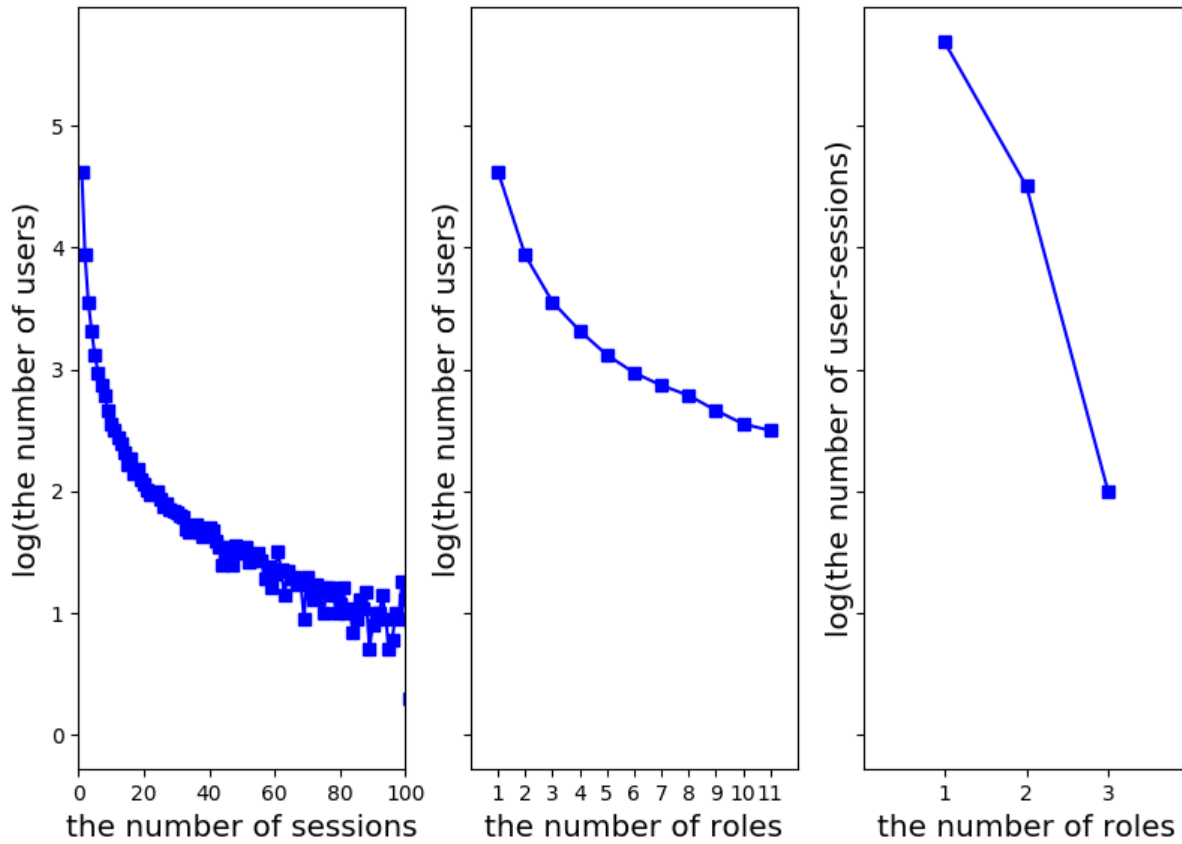


Figure 6.1: Statistics about users' participated sessions (left), the number of distinct roles they occupied throughout their lifetime (middle), and their role occupation per user-session (right).

Role Name	Dist (%)	Typical Behaviors
Emotional support provider	33.3	Provide emo support, provide appreciation, and provide encouragement in the forum, # subforums a user participated, provide empathy and provide info support in the forum
Welcomer	15.9	out-degree in forum, # replies in the forum, the ratio of talking to newcomers in the forum provide encouragement, provide empathy, the entropy of user-user interaction in the forum
Informational support provider	13.3	Provide info support, provide empathy, and provide encouragement in the forum, mention symptom related words, mention drug and anxiety related words
Story sharer	10.2	# threads, self-disclose positively, and seek emo support in the forum, self-disclose negatively, seek info support in the forum, use interpersonal pronouns
Informational support seeker	8.9	# threads in the forum, seek info support, self-disclose negatively in the forum, seek emo support in the forum, mention disease and symptom related words
Private support provider	5.3	Provide emo support, provide appreciation and empathy in private chats, self-disclose positively, provide info support and encouragement in private chats,
Private communicator	5.3	Preference for private chats, provide encouragement and info support in private chats, provide emo support and empathy, seek info support in private chats
All-round expert	2.5	# messages in private chats, provide appreciation and emo support in private chats, provide encouragement, # replies in the forum, self-disclose positively in the forum
Newcomer member	2.4	# threads, seek info support in the forum, self-disclose positively in the forum, self-disclose negatively, seek emo support in the forum, mention diagnose related words
Knowledge promoter	2.2	# urls per message, mention ingredient related words, provide info support in the forum, mention drug, anxiety and death related words, read behavior
Private networker	0.8	The entropy of user-user interaction, out-degree, in-degree in private chats, # messages in private chats, the ratio of being talked to by oldtimers, # private conversations

Table 6.3: Derived roles and their representative behaviors ranked by their importance in descending order.

6.4.3 Determining the Number of Roles

Quantitative Setting of Upper and Lower Bounds

The number of roles K in this model is a free parameter and is the element most susceptible to over-tuning (Schlattmann, 2003). We used the Bayesian Information Criterion (BIC) to select the number of components in the Gaussian mixture model (GMM). We trained Gaussian mixture models on the user-session corpus and experimented with K ranging from 2 to 20 to determine the optimal number of components/roles, as described in Figure 6.2. We found that models with $K \in [10, 15]$ seemed to be a good fit.

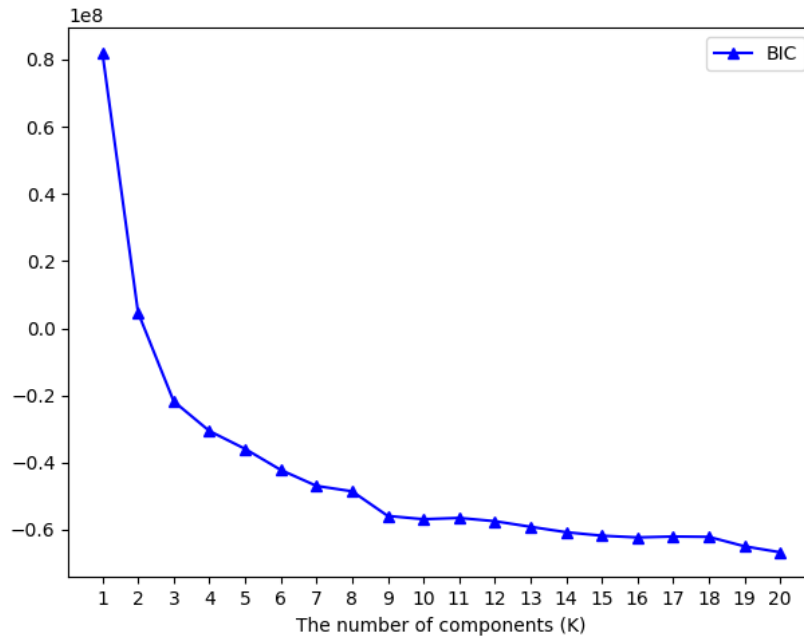


Figure 6.2: BIC scores for GMM model with different number of K

Qualitative Validation of Final Setting

Validating these behavioral role components inferred from unsupervised methods is challenging. Existing work on similar tasks such as LDA topic modeling has tried to validate the derived components by asking people to provide summary labels for each component (Blei et al., 2003; Nguyen et al., 2015; Xie et al., 2015) or by measuring the purity of the clusters or components (Chang et al., 2009; Mehrotra et al., 2013). However, interpreting topics or components by researchers themselves might introduce biases, and defining the purity of components that consist of member behaviors rather than

simpler features, like bag-of-words representations of topics, is hard to operationalize.

To overcome these problems, we followed a qualitative protocol to finalize the number for user roles and their names. We ran the Gaussian mixture model with our behavior features and user-session length for different values of K . We then discussed the extracted components with 6 domain experts (5 moderators from CSN and a senior researcher familiar with the site). We used their input to help interpret the latent components. We showed the domain experts the top ranked features associated with each role as well as three users who were most representative of each role (i.e., the three users from each role component whose behaviors were closest to the centroid representation of that component). The details about our semi-structured interview with domain experts is described in Appendix B. Based on their input, we set $K=11$.

6.5 Discovered Roles in Online Health Communities

After final parameter tuning and validation from discussions with domain experts, we have evidence that the model is effective in identifying latent roles that members occupy. Once these parameters were set, we worked with the 6 domain experts to co-develop short names and interpretable descriptions of each component in the model, describing the roles that emerged. These roles, their frequency in the corpus, and highest-probability features are described in Table 6.3.

1. **Emotional Support Provider:** people who respond to others with empathy, encouragement and emotional support. These active forum members participate in a number of sub-forums, in contrast to most users on CSN who only participate in one sub-forum most relevant to their cancer type.
2. **Welcomer:** people who respond to newcomers after they first post on CSN. These higher-tenured members interact with newcomers frequently and provide supportive empathy and encouragement.
3. **Informational Support Provider:** people who offer information and advice to others in the discussion board. This group of members discusses cancer-specific issues by mentioning symptoms and ingredient-related words, and provides in-

formation to others on the public forum.

4. **Story Sharer:** people who disclose personal information and emotions in order to receive support. They share their own experiences and stories in an introspective and verbose manner, which might help similar users and/or inform potential support providers about their situations.
5. **Informational Support Seeker:** people who ask questions and seek information from others in public forums. Members with this role initialize more threads, and seek around 1.7 standard deviations more informational and emotional support than average. They also talk more frequently about metastasis and other aspects of their disease.
6. **Private Support Provider:** people who use private chats to provide social support to others. People in this role provide emotional support, encouragement, appreciation and information to others in private chats, as well as self-disclose in a positive manner to encourage others.
7. **Private Communicator:** people who are protective of their personal details and only choose to participate in private chats. They seek and provide different types of support such as informational support, empathy and encouragement, and have strong tendency to communicate privately (3.7 standard deviations more frequently than the average level).
8. **All-round Expert:** people who engage in a large set of support exchange behaviors in both public discussion board and private chats. This group of members active engages and performs various kinds of actions such as providing appreciation in private chats, replying to others and self-disclosing positively in the forums.
9. **Newcomer Member:** people who ask questions and seek support shortly after joining CSN. Most members in this group stay at CSN for less than one month. They use the discussion board to ask for both informational and emotional support, and emphasize the uncertainty associated with cancer diagnosis results (0.8 standard deviation more than average).

10. **Knowledge Promoter:** users who post links and information from outside CSN. Those users present themselves as knowledgeable about what they are talking about and recommend external research pointers to members in need of help. Compared to regular members, knowledge promoters share two standard deviations more links in their replies to others.
11. **Private Networker:** people who seem to be network hubs in private chats. Although they participate in the discussion forum and exchange social support in private chats from time to time, they talk to a larger set of members in private chats and exchange more messages compared to other members.

After discussion with domain experts, we obtained agreement on the name and characteristics of 10 of the 11 derived roles. However, we failed to achieve consensus for *all-round expert*³. Despite this, domain experts agreed that the set of behavioral roles we identified were comprehensive:

“It seems very comprehensive and there are so many different examples, so I feel like it is covered very well with your different roles and labels.”

“I feel more comfortable to look at the three typical user messages than the descriptions of the features, which seem quite abstract.”

Domain experts did point out roles that our model did not capture. For instance, they identified “Guardian” or “Defender” role - people who fight with spammers or violate norms on CSN, trying to regulate others’ behaviors. One of the domain experts described the defender role this way:

“The one that I think did not emerge is the policeman, these people complain to moderators when some people are doing things wrong or tell other people that they are violating norms. They shouldn’t be diagnosing the way that they are diagnosing or other sorts of problems.”

“there are not a lot of them, but they kind of stick in your memories since they are telling others what to do.”

³We urge readers to interpret our follow up analyses about *all-round expert* with caution.

The defender role likely does exist on CSN, but our model did not capture it, either because the behaviors that characterize the defender role occur infrequently or the features we used to characterize user-sessions did not reflect these behaviors.

6.6 Evaluating Roles

Evaluation is an important issue. The unsupervised nature of role identification methodology makes model selection and the specification of role number challenging. In the sections above, we have demonstrated the evaluation of our derived roles via quantitative measures of model fit in terms of BIC scores and via qualitative semi-structured interviews with domain experts. The present section introduces another two types of role evaluations – *evaluation via downstream applications* and *validation with role holders*, which we introduced in detail in Chapter 3.1.4.

6.6.1 Recommender System with Roles

As a natural follow-up, in this part, we utilize the roles identified to help improve recommender systems to match CSN members with others who are likely to meet their needs (e.g., matching Greeters to Newcomers, Caregivers to Support Seekers). We expect that this role-based recommender can not only help boost the recommendation performance, but also provide interpretability and explainability to users about why such recommendations are made. Before diving into the details of our role-based recommender, we begin with an introduction to the basic concepts in recommender systems.

Classical recommender systems predict users' preferences over items such as movies or products and proactively recommend to users items that they might be interested in. The field of recommendation can be categorized into two basic architectures (Bobadilla et al., 2013). The first is *Content-Based* system, where the focus is on the properties of items and recommendations are made based primarily on the similarity between users' and items' auxiliary information (Ferman et al., 2002). The second is *Collaborative Filtering* (CF), where the systems focus on the relationship between users and items and recommendations are made based on finding similar users and recommending what similar users like. Latent factor models like matrix factorization, and neighborhood models are two canonical approaches in CF to capture users' interests (Koren et al., 2009; Koren, 2008; Yang et al., 2014a,b). With the recent advances in deep learning, there

are also various neural extensions of traditional recommendation methods (Zhang et al., 2017; Sedhain et al., 2015). For example, it is straightforward to construct a dual neural network to model the interaction between users preferences and items features, similar to the decomposition of matrix factorization (Dziugaite and Roy, 2015; He et al., 2017).

Role-based Recommender System

In the discussion forum of CSN, our goal is to direct participants to useful and informative threads that they might be interested in. This section presents our recommendation prediction model for this context, which can be fit into a class of popular matrix factorization models (Rendle, 2010). The relevance matrix between participants and threads is denoted as R with entry $r_{u,t}$ representing the preference of user u towards thread t . Here, we proposed two ways to define the **preference** between a user and a thread. In `Setting 1`, if a user u posted a comment to a thread t , then $r_{u,t}$ increases by 1. However, the content in the comment differs a lot; some comments only contained a few words such as “*hang in there*”, while others might provide concrete details or personal stories about their cancer journeys. Thus, we also introduced `Setting 2` - if a user u posted a comment to a thread t , then $r_{u,t}$ increases by a score that relates to the length (word count) of the comment.

Formally, for each user u , thread t , and u 's preference towards t , the predicted score $\tilde{r}_{u,t}$ is defined as follows. Here, p_u and q_t are latent vectors associated with users and threads. μ is the overall average preference, and b_u, b_t are user/thread biases.

$$\tilde{r}_{u,t} = \mu + b_u + b_t + p_u^\top q_t \quad (6.1)$$

We further incorporate our identified social roles into this framework. The underlying assumption is that participants with similar roles might share similar interests towards threads. Here, we averaged the role occupations across u 's historical participated sessions and used this $R(u)$ to assist the preference prediction. Our recommender system with role information can be characterized as follows. $\phi(v)$ models the influence ability of the role v on participants.

$$\tilde{r}_{u,t} = \mu + b_u + b_t + \left(p_u + \frac{\sum_{v \in R(u)} \phi(v)}{\sqrt{|R(u)|}} \right)^\top q_t \quad (6.2)$$

Model	Setting 1		Setting 2	
	RMSE	Correlation	RMSE	Correlation
Popularity	0.593	0.564	1.703	0.518
Classical MF	0.514	0.598	1.625	0.578
Role-based Recommender	0.475	0.664	1.604	0.592

Table 6.4: Performance comparison for different types of recommender systems.

Offline Recommendation Performance

We conducted our experiments on the public discussion board of CSN. It has 48, 317 registered users who have exchanged 1,073,020 messages belonging to 131,237 threads. Root Mean Square Error (RMSE) and Correlation are our evaluation metrics. We also compared our role-based recommender with **Popularity** that conducts thread recommendation based on thread popularity and recency, and with **Classical MF**. Note that our goal here is not to develop the state-of-the-art recommender; instead, we are interested in whether incorporating social role information can increase the performances of some downstream applications such as recommendation.

The recommendation results are summarized in Table 6.4. As we can see, **classical MF** significantly outperforms popularity based approach. Introducing the eleven social roles we identified can boost the correlations to 0.664 from 0.598 in *Setting 1*, with an 10% increase. Similar results were found in *Setting 2*. Overall, this demonstrates that behavioral roles that members occupy make an important contribution in capturing the latent matching between interest of participants and the topics involved in threads.

6.6.2 Deployment Studies on Recommendation

As an initial step, we have deployed a basic version of this recommender in the live site of CSN. We recruited participants by simply posting an opt-in link for existing CSN users and an opt-out link for new registrants on the CSN website. Clicking the link automatically changed the user interface to the site to incorporate our interventions and experiment. When a user logged into CSN, recommendations about useful and informative threads as well as similar members will be made, as shown in Figure 6.3. When a user is browsing a specific thread, we will also recommend other relevant threads on the right side bar (see Figure 6.4).

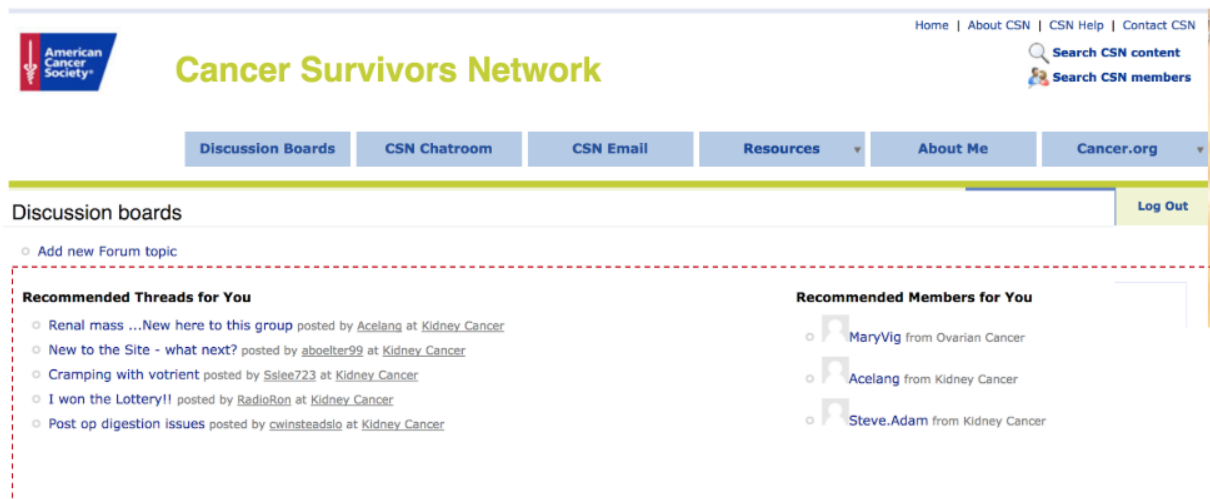


Figure 6.3: Recommendations when a user logged into CSN

As of Sep 28, 2018, over 10,000 people are currently using the new recommender-interface to the ACS support groups, including 75% of those who registered on the ACS site since Dec 15. In addition, we have over 450 long-term members using our new version of the site. We did a within-participant experiment to test whether people are more likely to read the content using our interface. Based on current received data, our deployed system almost doubled people’s hit rate of reading threads. That is, when we use our system to match people with information and appropriate helpers, members are twice more likely to click the recommended content compared to the default recency or popular-based recommendations.

This system is running on the live site of CSN, and is producing real-world benefits. Our next step for this intervention is to better encourage members to view and participate in discussions that match their illness-related needs based on members’ social roles. That is, we can use social roles to explain our recommendations to users. For example, when a user logs in, instead of saying “*Recommended Members for You*”, we can explain the set of members we recommend to him/her as “*Here are some newcomers you might want to say hi*” or “*Here are some information experts you could reach out*”.

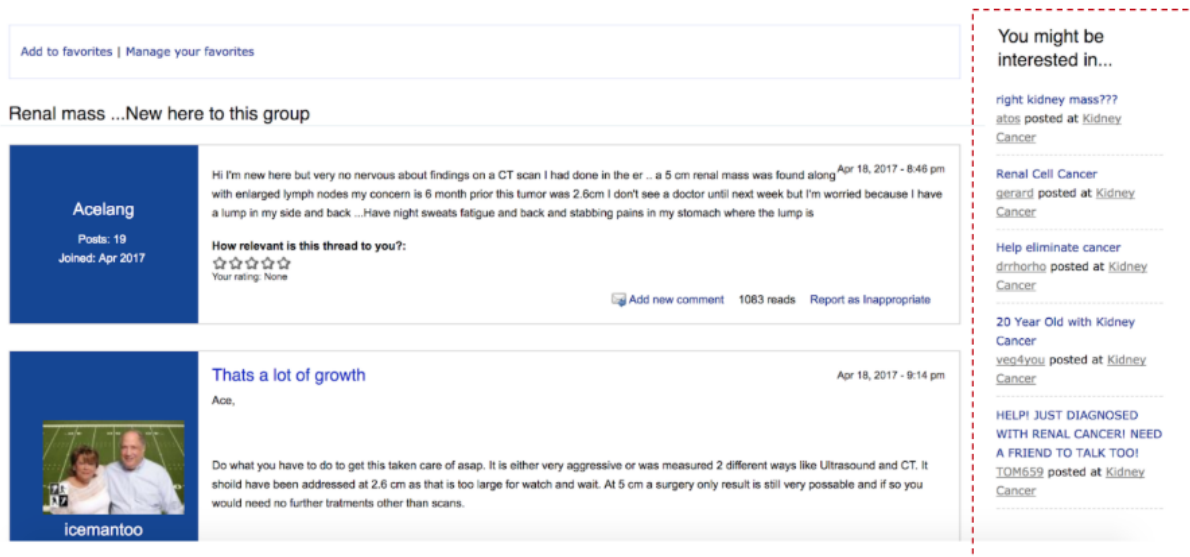


Figure 6.4: Recommendations when a user was browsing a thread on CSN

6.6.3 Survey on Roles

To further validate whether our identified roles are consistent with what members do, we designed a survey to ask people how they self-identify with regard to each of the roles that we have identified. Directly asking what roles people think they occupy might suffer from social desirability bias. For example, people may answer our role question in a way that makes them look more favorable by claiming they are “support givers” rather than “support seekers”. Therefore, we chose to ask how people behave on CSN across a set of role-typical behaviors.

Survey Design

We used the most characteristic behavior (in most cases the behavior feature that has the largest weight on the role representation) associated with each role to design the question measure for that role. For example, we framed the role of *informational support seeker* as the behavior of *ask questions and seek information from others on the Discussion Boards*. *Knowledge promoter* is represented via the question of *post links and information from outside CSN*. Other roles are similarly interpreted as their most typical behaviors. We asked participants to judge to what extent they perform each of the role behaviors in a 1-5 Likert Scale, ranging from “Not at all”, “A little”, to “Very much”. Beyond questions that relate to our identified functioning roles, we also added three extra questions

to capture disengaged members (people who rarely visit CSN to read or post), lurkers (people who visit CSN to read, but not post), and defenders (people who help enforce CSN norms by communicating with members and moderators about inappropriate behavior). These role questions are described in detail in Table 6.5, and were integrated into a large-scale behavioral survey on CSN.

How much does each of the following describe how you typically behave when visiting CSN?

	Not at all	A little	Somewhat	Quite a bit	Very much
(S1) Rarely visit CSN to read or post	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
(S2) Visit CSN to read, but not to post	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
(S3) Ask questions and seek information from others on the Discussion Boards	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
(S4) Provide information to others on the Discussion Boards	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
(S5) Respond to others on CSN with emotional support, empathy, or encouragement	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
(S6) Use CSN email to provide support privately to others	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
(S7) Tell my cancer story and disclose personal information about myself	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
(S8) Comment on posts from a wide variety of people on CSN	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
(S9) Ask questions and sought support shortly after joining CSN	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
(S10) Respond to newcomers after they first post in CSN	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
(S11) Maintain personal relationships with others	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
(S12) Post links and information from outside CSN	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
(S13) Communicate about technical aspects of cancer treatment, such as radiation, clinical trials, side effects, metastasis, hair loss	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
(S14) Help enforce CSN norms by communicating with members and moderators about inappropriate behavior	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Table 6.5: Question that examines how much people self-identify themselves with different role behaviors.

Self-reported Role Occupation

This survey was sent out to active CSN users who had logged into CSN at least once since Sep 1st, 2017 and are US residents according to CSN databases. In total, 474 participants answered our survey questions, among which 243 users had ever made posts either in the public discussion board or via private message. We predicted the roles that those 243 members enacted via our social role models during their participation sessions prior to when they answered our survey, and used the averaged role probabilities as their final role occupations.

The Pearson correlations between members' self-reported roles and our predicted roles that members occupy was reported in Table 6.6. Note that due to sample size concerns we removed the role of *Private Communicator* from our predicted roles because less than ten people had occupied it. In Table 6.6, each row represents one self-reported role behavior, the descriptions of which can be found in Table 6.5 using the corresponding indicators such as **S1**.

We observed that *Information Seeker* was found to be weakly associated with *S9* ($r=0.09$) - asking questions. *All-round Expert* correlated well with *S11* (maintain relationship, $r=0.20$), *S14* (report inappropriate behaviors, $r=0.11$), *S7* (share stories, $r=0.15$) and *S6* (provide support in private, $r=0.11$). People who self-reported sharing stories (*S7*) are very likely to be a *Story Sharer*, with an r of 0.11. The correlation between behavioral *Knowledge Promoter* and *S12* (post link and information) was 0.10, and *Private Networker*, *Private Support Provider*, and *Newcomer Member* have reasonable correlations with their corresponding survey questions.

In contrast, we did not find expected correlations among *Welcomer* and *S10*, and among behavioral *Emotional Support Provider* and *S1* - rarely visit CSN. The correlation between *Informational Support Providers* and *S4* (provide information) was also negligible. The lack of correlation and significance may occur because a large portion of respondents were lower tenured members (around half of the population had stayed in CSN for less than 6 months) who might not have developed enough expertise for performing well on their self-reported roles. Our session level modeling of social roles may also partially account for why self-reports from the survey seems not map well on roles defined at the session level, since respondents may be presumably providing

person level information about how they typically behave. The *Valid* row in Table 6.6 indicates whether people's self-reported roles are consistent with our predicted roles in each category.

Corr.	emotional support provider	welcomer	information provider	information seeker	all-round expert	story sharer	knowledge promoter	private networker	private support provider	new member
(S1)	-0.12	-0.10	0.07	0.01	0.05	-0.10	0.04	0.07	0.05	0.09
(S2)	-0.06	-0.02	-0.04	0.12	-0.01	0.07	0.03	-0.08	-0.04	0.02
(S3)	0.00	0.06	-0.05	0.06	0.03	0.10	-0.02	0.05	-0.08	-0.15
(S4)	-0.04	0.05	0.01	0.03	-0.02	0.09	0.01	-0.01	-0.05	-0.06
(S5)	-0.01	0.03	-0.01	-0.04	0.10	0.14	-0.05	0.09	0.02	-0.13
(S6)	0.02	-0.04	-0.15	0.06	0.11	0.01	-0.01	0.22	0.10	-0.12
(S7)	0.05	-0.03	-0.09	-0.01	0.15	0.11	-0.01	0.08	0.01	-0.07
(S8)	0.09	0.07	-0.10	-0.03	0.07	0.13	0.04	0.01	0.08	-0.16
(S9)	-0.01	0.05	-0.10	0.09	0.06	0.14	-0.01	0.05	-0.03	-0.15
(S10)	0.08	0.05	-0.17	0.01	0.08	0.19	-0.05	0.06	0.03	-0.06
(S11)	0.02	0.05	-0.13	-0.01	0.19	0.02	-0.02	0.21	0.03	-0.12
(S12)	0.01	0.03	-0.05	0.02	0.03	-0.01	0.10	0.09	-0.02	-0.18
(S13)	-0.02	0.06	-0.04	-0.02	0.01	0.11	0.03	0.11	-0.02	-0.07
S(14)	-0.01	-0.01	-0.02	-0.01	0.11	0.01	0.06	0.05	-0.03	-0.05
Valid	✗	✗	✗	✓	✓	✓	✓	✓	✓	✓

Table 6.6: Pearson correlations between members' self-reported roles and our predicted roles that members occupy.

To figure out the potential causes for those insignificant findings, we calculated the correlations among users' self-reported role behaviors and their actual behaviors on CSN. Note that differences exist between this correlation check and Table 6.6 — the predicted roles (column) in Table 6.6 come from the aggregations of all sorts of users' behaviors, while the actual behaviors in the present correlation check are specific types such as the number of their comments or the average amount of emotional support they provided per message.

Table 6.7 describes the correlations between members' self-reported role related behaviors and their actual behaviors on CSN. Here, *# threads* refers to the number of thread-starting messages they have initialized on CSN prior to when they answered our survey. *# comments* denotes the total number of comments they posted, and *# private messages* are the amount of private emails they sent to others. The measures of *seek informational support*, *seek emotional support*, *provide informational support*, and *provide emotional support* refer to the amount of two types of social support that each user sought per thread-starting message and provided per comment, which are predicted via machine learning models described in Table 6.2 in Section 6.4.1. We found that the average informational support a user provided per message has a negative correlation ($r=-0.033$) with users self-reporting they provide information on CSN, partially confirming the social desirability bias Fisher (1993). In contrast, people who self-reported that they welcome new members actually have more thread-starting messages ($r=0.268$) which are generally thought to be oriented more towards seeking support. Moreover, the emotional support they expressed seems less evident, compared to the amount of informational support ($r=0.143$) and emotional support ($r=0.223$) they sought in their messages. This suggests certain degree of inconsistencies in members' self-reported measures, which may help explain why we failed to validate the role of *welcomers* in Table 6.6.

Correlations	# threads	# comments	# private messages	seek info support	seek emo support	provide info support	provide emo support
(S1) Rarely visit CSN to read or post	0.067	0.106	0.138	-0.092	-0.087	-0.034	-0.048
(S2) Visit CSN to read	0.073	0.003	-0.041	0.109	0.067	-0.075	-0.010
(S3) Ask questions & seek information	0.171	0.036	-0.020	0.157	0.115	-0.136	0.013
(S4) Provide information to others	0.235	0.171	0.114	0.155	0.141	-0.033	0.094
(S5) Respond to others with emotional support, empathy, or encouragement	0.223	0.150	0.100	0.011	0.069	-0.136	0.164
(S6) Use CSN email to provide support privately	0.297	0.197	0.150	-0.015	-0.016	-0.066	0.007
(S7) Tell my cancer story and disclose personal information about myself	0.254	0.178	0.124	0.070	0.142	-0.017	0.181
(S8) Comment on posts from a wide variety of people on CSN	0.263	0.153	0.082	0.090	0.145	-0.049	0.156
(S9) Ask questions and sought support shortly after joining CSN	0.183	0.091	0.028	0.248	0.246	-0.113	0.055
(S10) Respond to newcomers after they first post	0.268	0.158	0.083	0.143	0.223	-0.156	0.158
(S11) Maintain personal relationships with others	0.364	0.268	0.204	0.037	0.102	-0.139	0.130
(S12) Post links and information from outside	0.184	0.211	0.141	0.029	0.051	-0.026	-0.002
(S13) Communicate about technical aspects of cancer treatment, such as radiation, clinical trials, side effects, metastasis, hair loss	0.213	0.171	0.117	0.111	0.081	0.012	0.077
(S14) Help enforce norms by communicating with members & moderators about inappropriate behavior	0.145	0.212	0.145	0.067	0.045	-0.064	0.007

Table 6.7: Pearson correlations between members' self-reported role behaviors and their actual behaviors

Role	HR	Std.Err
Emotional support provider	0.984	0.027
Welcomer	0.883***	0.028
Informational support provider	1.060	0.034
Story sharer	0.872***	0.034
Informational support seeker	1.324***	0.023
Private support provider	0.842***	0.033
Private communicator	1.031	0.022
All-round expert	0.869***	0.028
Newcomer member	1.054***	0.025
Knowledge promoter	1.091***	0.028
Private networker	0.916*	0.035

Table 6.8: Survival Analysis predicting how long members continue to participate in the community. $p < 0.001$: ***; $p < 0.01$ **: **; $p < 0.05$ *: *. Number of users = 66,246. Number of user-session records = 522,429

6.7 Influence of Emergent Roles on Commitment

Members' patterns of activities and roles can influence their contribution and commitment to the community. Although previous research has investigated members' commitment to both offline and online organizations ([Bateman and Strasser, 1984](#); [Kim, 2000](#); [Ren et al., 2007](#); [Yang et al., 2017b](#)), no computational research has examined how members' assumption of emergent roles relates to commitment in online health communities. This section examines how emergent roles help predict continued participation of members on CSN. Doing so will allow us to better understand members' engagement, as well as demonstrate the utility of our derived roles.

We use survival analysis to investigate how members' occupation of social roles correlates with the length of their participation on CSN. Survival analysis is a type of regression analysis for estimating influences on the time to an event of interest, especially for censored data. In our context, the event is defined as members dropping out of CSN. We used Stata survival command with a Weibull distribution of survival times in order to perform this analysis ([StataCorp et al., 2007](#)), with the unit of analysis being the user-session.

Control variables included the member's gender, whether the member had cancer, and his/her tenure (i.e., how many months they have stayed at CSN). Since the continuous explanatory variables were standardized, the Hazard Ratio (HR) is the predicted change in the probability of dropout from CSN for a standard deviation increase in the predictor. A hazard ratio greater than one means the role is associated with a higher than average likelihood of dropping out, while a hazard ratio less than one means a lower than average likelihood of dropping out. Because of the correlations between different roles, and correlations among roles and tenure, we built separate survival models for each role, resulting in 11 models.

Results of the survival analyses are shown in Table 6.8. The analyses show that members occupying certain roles - *knowledge promoter*, *informational support seeker* and *newcomer member* - are less likely to continue in CSN (i.e., lower survival rates). Specifically, members who were one standard deviation more likely to occupy *informational support seeker* roles were 32.4% more likely to leave the community after that session. Similarly, members who were one standard deviation more likely to be *newcomer-seekers* were 5.4% more likely to drop out from the community, while members who share external knowledge with others on CSN (*knowledge promoters*) were 9.1% less likely to continue their participation. These results suggest that roles related specifically to information-sharing are associated with higher rates of drop-out, possibly because researching disease or treatment relevant information is a distinct, time-consuming use of online resources, separate from community-building goals. These members may see CSN as a more transactional resource, either giving or receiving information, and represent a less committed user.

In contrast, occupying roles such as private networker, private support provider, newcomer welcomer, and story sharer are associated with members staying at CSN longer. This may be because being support-providers to others encourages members to interact with other members time after time, developing stronger relationships. People who respond to newly registered members with support were 12% more likely to stay on CSN; members who were willing to self-disclose their experiences to seek support or benefit others had a 13% higher survival rate.

6.8 Stability and Dynamics of Roles

As members go through their life cycles, they might choose to drop out or stay on CSN. The roles of those who stay might change over time. For example, as previously described by the Reader-Leader framework (Preece and Shneiderman, 2009), people may change from being peripheral to core members of the community. In this section, we examine whether members' emergent roles vary over their tenure at CSN, and we test the stability of users' emergent roles at both individual- and community- levels.

6.8.1 Community Level Stability

We first investigated the mixture of roles in the forum overall over a thirteen years period (see Figure 6.5). The frequency of the majority of the behavioral roles on CSN did not change substantially over time. This demonstrates that although new members join and old members leave, organization-level compositions in terms of emergent role behaviors remain stable. A closer look at the year-by-year role composition revealed that informational support provider increased to 25.5% in 2017 from 11%-13% in earlier years (2004-2015). We also observed a weak increase for newcomer seekers, likely due to large increase in active forum users after 2015. In contrast, the percentage of welcomers in the community decreased to 4% in recent years, perhaps suggesting that old-timers, who dominate the welcomer role, are becoming less welcoming to newcomers or less polite over time.

6.8.2 Individual Level Dynamics

Changes in Role Occupation Over the User Lifecycle

When members first join CSN, they may have high uncertainty about the type of people who are members and the group's norms (Bauer et al., 2007). Over time those who stay may accumulate experience in terms of both domain knowledge related to their diseases and the group and its norms. This knowledge may increase people's ability to give back to the community. To investigate whether higher tenured members occupy a different set of roles than newcomers, we compared role associated with members' tenure in CSN, as described in Figure 6.6. Specifically, we looked at members' role occupation in their first month - (0, 1], from their second month to six months - (1, 6], from six months to a year - (6, 12], and after one year - (12, +]. Among 66,246 members, 93% of users

participated in CSN in their first month after registering.

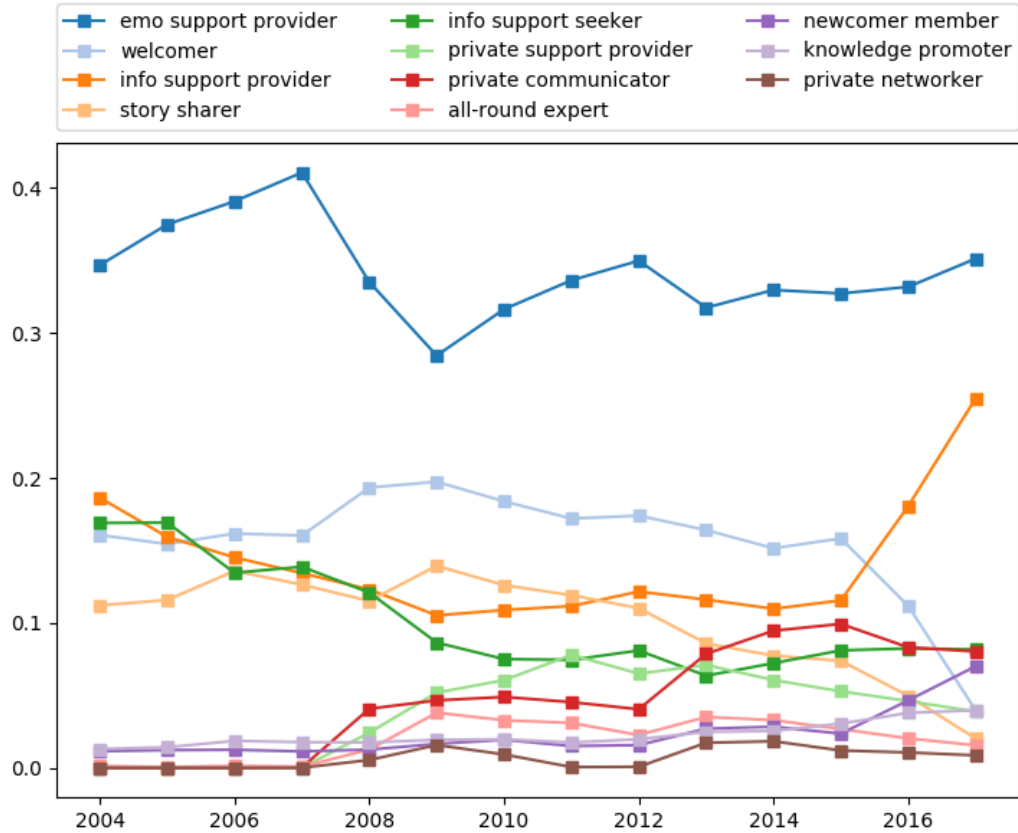


Figure 6.5: The percentage of different role occupation from 2004 to 2017.

Figure 6.6 shows that emotional support providers, welcomers, informational support providers, story sharers and informational support seekers were the most common roles. During members' first month on CSN, roughly 20% of them occupied the role of information support seeker, and 15% choose to share their experiences and stories to start their conversations. As tenure increases, members were more likely to occupy the role of emotional support provider, private support provider and private networker. In contrast, members are less likely to occupy the story sharer and information support seeker roles the longer they stayed on CSN, while they were more likely to be newcomer welcomers after their first month. Although Figure 6.6 includes only users who have been at CSN for a year, similarity results obtain for users with who have been at CSN for less than 12 months or less than 6 months.

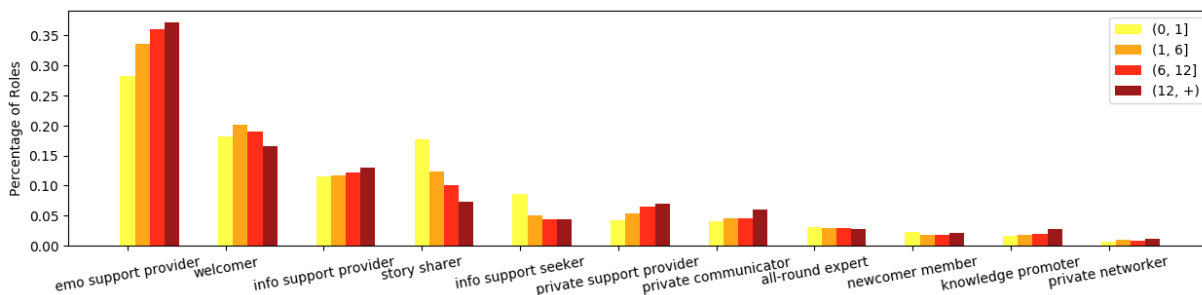
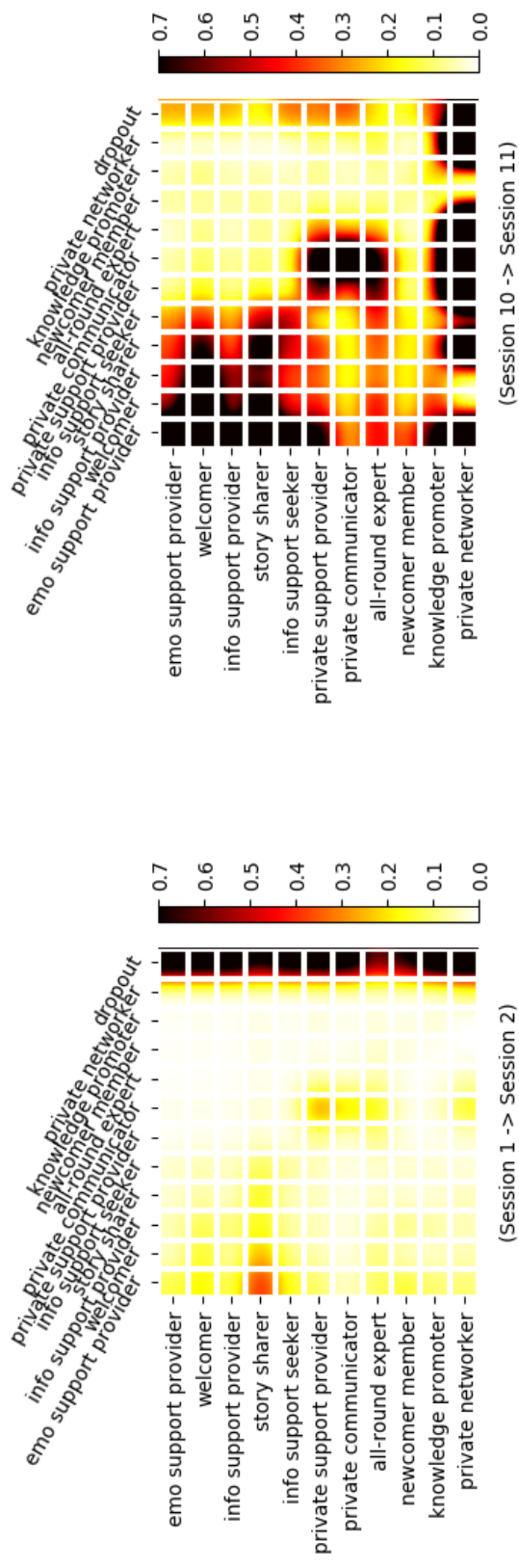


Figure 6.6: The percentage of role occupation for users who participated in CSN for at least 12 months in their different tenure. (0, 1] refers to members role occupation in their first month, with (1, 6] as their second months till six months. Similarly, (6, 12] denotes role percentages from their six months to one year and (12, +) means after one year.

Role transition pattern	Prob
private communicator → private communicator	0.413
info support provider → emo support provider	0.362
emo support provider → emo support provider	0.336
welcomer → emo support provider	0.335
newcomer member → emo support provider	0.330
info support seeker → emo support provider	0.326
private networker → private communicator	0.315
story sharer → emo support provider	0.312
story sharer → welcomer	0.207

Table 6.9: The top 9 most frequent role transition patterns.



120 Figure 6.7: Conditional probability of role transitions from one session (row) to another (column) after the first (left) and tenth (right) session.

Role Transition Processes

These results suggest that members assume different roles in different stages of participation. To further investigate role evolution, we examined the process of members' moving from one role to another across sessions. Specifically, we model users' role transitions as a Markov process, i.e., if a user assumed a particular role during session i , what is the probability that he or she would take on any specific one of the eleven roles in session $i + 1$? We calculated the presence of each role transition pattern by looking at members' roles in any adjacent sessions. Here, a user is said to occupy a role in a session if that role had the largest weight across the 11 roles. We also model a user's likelihood of dropping out (i.e., discontinuing participation in CSN) after occupying a role. This produces 132 total possible transitions (11 x 12, where the one added transition probability leads to dropout).

We described the most common transitions overall in Table 6.9. Since 70% members dropped out of CSN after 30 days, we calculated this transition pattern only for members who stay on CSN longer than that. We found that private communicators are the most stable role, at 41.3% carryover from session to session; users who take on this role are more likely to maintain it in their next session compared to any other role. Not only do users who provide emotional support in one session tend to continue in that role in the next session, but it is the most common role for users to transition into from other roles - 33.5% of welcomers, 36.2% of informational support providers, 32.6% of information support seekers and 31.2% of story sharers. The conditional probability of transiting from informational support seekers to emotional support providers is 0.326, confirming the typical transitions from outside observers into core members of the community (Preece and Shneiderman, 2009). This also reflects the rule of reciprocity that members who seek resources eventually give back to their communities. This showed that members transit from roles that seek for resources to roles that offer help to others.

The *emotional support provider* role derives its stability partially from being a role associated with longer-term users, rather than newcomers. We show this by next deriving transition matrices *conditioned on session*. Figure 6.7 shows the results for two particular session transitions: from session 1 to session 2 (left side), indicating the first step of users from newcomers to group membership; and from session 10 to session 11 (right side), as

an example of the more stable matrix that emerges as users become long-term members.

We found three distinct groups of newcomers. The first group does not follow any of the public roles that engage in broader discussion forum, but instead use the site primarily as a vehicle for private conversations, such as from *private communicator* to *private communicator* (25.4%). The second group is primarily *information seekers*, who then transition into providers (of both informational and emotional support) and *welcomers* in their follow-up sessions. The third common group, *story sharers*, are notable for their very low dropout - 64.2% of story sharers return for a second session on CSN, compared to 35.5% of first-time users that assume all other roles combined.

As tenure increases in the 10th session transition matrix, members are likely to transition out of the role of *information support seeker* and *story sharer*, and more likely to transition into the role of *emotional support providers* and *welcomers*. These roles are common and “sticky” - users have high probabilities of maintaining that role from session to session. *Private support providers* and *private networkers* were present at high rates among longer-term users, and maintain their roles over time. While support providers transition into their roles over time, *private networkers* were more likely to have taken on this role early in their tenure.

Note that for role transition analyses, we used a heuristic rule and treated each user in a session as occupying a single role - the role with the highest weight - to model the process of role transition. Since users can occupy hybrid roles, it is possible that co-occurring roles might affect our role transition results. For example, users transit from one set of roles to another set of roles in their next sessions or dropout if they did not have a next session. Future work could address this multiple role transition by modeling the mapping from 2^K roles to 2^K roles and dropout, resulting in a $2^K \times (2^K + 1)$ matrix compared to a $K \times (K + 1)$ matrix in Figure 6.7. For example, how do people transit from informational support seeker, newcomer seeker to emotional support provider, welcomer. However, such a complete approach might run into challenges with data sparsity, so the right course of action will likely be to investigate the tradeoffs in representation.

6.9 Discussion

This research investigated the functional roles that members occupy in an online cancer support community, and how such role occupation influences their engagement within their communities. We utilized the generic framework introduced in Chapter 2 to define emergent roles in online communities with four components - goal, interaction, expectation and context. We operationalized a set of behavioral features to represent each component and then employed unsupervised models to extract the functioning roles that members occupy, which discovered 11 interpretable roles in online cancer support groups.

Among the few studies that investigated emergent roles in online communities, most have paid attention to platforms such as Wikipedia (Yang et al., 2016a; Arazy et al., 2015, 2016). Previous research in online health communities suggested that there are distinct subsets of users with different “roles” (Yang et al., 2017b), but had no formal methods of modeling what those subsets were. We extend this line of work into another type of community - to the best of our knowledge, the first work to use data-driven methods to identify behavioral roles in online health communities. Some of the prototypical behaviors associated with the roles we derived correspond to roles in conceptual frameworks; for instance, our “informational support seeker” and “informational support provider” correspond to “information seeker” and “information giver” (Benne and Sheats, 1948). The role of “emotional support provider” seems to reflect the role of “encourager” (Mumford et al., 2006, 2008), which involves showing understanding and acceptance of others’ ideas and suggestions.

In addition to helping define these roles, this generative model to describe subsets of users can both identify a user’s assumption of a role in real time, and model how an individual member is likely to transition across roles over time. Most earlier research on role identification used limited metrics in evaluating roles, and statistical models more well-suited to analysis of static datasets, rather than real-time prediction in a machine learning architecture. These models also required metrics of success such as model fit or manual labeling, suffering from potential biases and lack of domain knowledge. To overcome such issues, in addition to quantitative validation of model fit, we followed through with in-depth interviews with 6 domain experts who have a deep understanding of CSN. The results of these interviews support the validity and quality

of our derived roles. We believe that most existing empirical methods for identifying roles in other domains (Yang et al., 2016a; Arazy et al., 2016) can be abstracted into this generic methodology, which can be applied to any other types of community, both online and offline.

Our studies on how roles influence members' survival revealed that socially positive roles such as support providers and newcomer welcomers were associated with staying longer at CSN. It may be that to take on these socially positive roles, members have to stay in the group for a while to be familiar with the group norms and other members; occupying such roles may also indicate that members already have relationships with and attachment to others or the group as a whole. The role transition analyses illustrate that members on CSN enact emergent roles and frequently transit to other roles, confirming prior work that such roles are transient (Arazy et al., 2016).

6.9.1 Implication

Our research sheds light on how to build more successful online communities from both practical and theoretical perspectives. Theoretically, our work contributes to the understandings of emergent roles by validating the general multi-faceted role framework that we proposed in Chapter 2. The iterative role identification process described here is reproducible broadly within the HCI community, as are our mixed-methods (quantitative/qualitative) criteria for evaluating the quality of derived roles. Practically, our role modeling methods can be employed to develop tools that detect members' needs, track their activities, and offer them help and task of interests. Such identified roles can better help patients know themselves and others. Future work should focus on incorporating this information into profile pages and other interface affordances. The derived roles can be incorporated as additional features for connecting users to other users, content and tasks based on their roles along with other information about them (e.g., their disease, expertise or, emotional support needs).

In addition to the potentials in boosting the recommendation performance, members' functioning behavioral roles can also be used as explanations to users about why such recommendations are made. For example, instead of "You might be interested in ...," the recommendations can be explained like "This is an information expert who can help you with breast cancer." Online communities could also introduce some of these derived

roles as badges to encourage users to assume these roles and reward those who do.

6.9.2 Limitations

This research has significant limitations. While it is an initial step towards understanding emergent roles in online support groups, we do not have self-reported evaluations from CSN members about their perceived role occupations. Although we validate our derived roles with a set of domain experts, future work surveying members who tend to occupy such roles will allow us to compare model predictions with user-perceived role occupation.

Second, while we make correlative descriptions of members' role occupation and their engagement on CSN, our work is not causal. Thus occupying socially positive roles may motivate users to stay longer, but alternatively, new users who were more likely to maintain membership may be more likely to perform such roles, reversing the causal link. While this research looks at one online cancer support group, we cannot necessarily generalize findings to other online communities without further work.

Finally, the opportunity to use role predictions to alter user experiences and make recommendations has important ethical considerations. We have developed a model with the potential to predict users' future behaviors in online communities, and adjust their user experience based on those predictions. However, such models have the potential to become a self-fulfilling prophecy, shepherding users into a particular activity path without giving them the full breadth of opportunity to explore other roles. As this research evolves into interventions, a crucial element for analysis will be interviews with members, observation of changes in their behaviors compared to baseline conditions, and an interdisciplinary analysis on the changed outcomes for users - particularly vulnerable, healthcare-seeking users - in these and similar communities.

6.10 Reflection

This chapter introduces systemic empirical work for automatically inferring members' functioning roles when participating in online cancer support groups. Different from the studies in Chapter 4 and Chapter 5, this work examines the problem of role identification in a new context - Cancer Survivors Network. Overall, this work successfully

validates the social role framework and our generic role identification methodology. First, we employed prior work on social support in online health communities to provide guidance and supervision for role-related behaviors — role postulation, and then operationalized multiple facets (**Interaction, Goal, Context, Expectation**) of social roles simultaneously, which leveraged various types of computational techniques including linguistic modeling of members’ messages, social network analysis of members’ interaction structures, and machine learning based estimation of members’ goals. Second, by incorporating members’ actions and attributes from role definition above into a statistical generative model, we are able to learn a set of coherent roles that can well explain individuals’ regularities. Third, such derived roles were validated comprehensively via four evaluation manners: quantitatively in terms of the BIC score on the held-out data, qualitatively in terms of six domain experts’ interpretation, directly in terms of correlations with self-reported roles from role holders via a large-scale behavioral surveys, and indirectly in terms of performance boost to a recommender system. In addition to those careful considerations, we also conducted sensitivity analyses to select the appropriate unit of analyses (i.e., session) and the number of functioning roles. To sum up, the present work demonstrates that our social role framework and our iterative role identification process are reproducible broadly, as are our mixed-methods criteria for evaluating the quality of derived roles. It is worth mentioning that our operationalization of **Expectation** is too simplistic, by only measuring the number of messages that members sent to moderators and their usage of modal words. This partially explained our failure in capturing roles such as defenders or vandal fighters. We also did not examine how our derived emergent roles deviate from their expected role behaviors. In terms of context, we simply looked at where members’ interactions were happening, i.e., their communication channels. While the role facet of *agent* has not been necessary for the intended roles modeled here, it may provide helpful guidance for more complex role models. Future studies may extend this work to directly cross the behavioral roles identified with some of the members’ personal attributes (e.g., informational support provider × cancer type).

In addition to identifying the behavioral roles that member enact on CSN, we also studied the mobility and stability of roles by looking at how members change their roles and found that (1) members assume different roles in different stages of their participation, and (2) there are a set of common and stable transitions across sessions. Note that here we used two time-related units for our role transition analysis — at a

person level as in Figure 6.6 and at a session level as in Table 6.9 and Figure 6.7. Similar trends exist, for instance, members change from roles seeking resources to ones offering help. However, session level analysis reveals more nuanced findings such as the private communicator are the most stable role and there are three distinct groups of newcomers. In contrast, person level modeling provides an overall picture of how roles gradually change over users' life-cycles. One potential explanation for different transition patterns is that the session-level transitions may be a result of a context shift, while the person-level transitions may be the result of more fundamental changes to the individual. Having said that, the two time-related units are not in conflict with each other. As the basis for modeling social roles, *user session* was introduced as an appropriate temporal unit to represent role holders' activity, which is relatively short and cleanly-defined time intervals. We also tested the robustness of the role models with other different temporal units (e.g., day, week, or month) and chose the granularity of the session because of its interpretability. To understand users' role occupation over a long time period, we can organize users' role occupations from all sessions by chronological order, and examine any potential roles changes at a long time window, a person or community level in an aggregate manner. To sum up, special attention should be paid to determine the micro-level unit of role analysis, and examine how time frame correlates with people's roles at a macro level.

Part III

Envoi

Chapter 7

Conclusion

There is no real ending. It's just the place where you stop the story.

– Frank Herbert

Millions of people participate in online communities, exchange expertise and ideas, and collaborate to produce complex artifacts such as Wikipedia, the world's most comprehensive encyclopedia or open source software projects that run the Internet. This thesis investigates social roles that members enact in online communities when helping their communities and the public at large. This work makes a significant contribution to the theoretical framework of social roles by defining what are social roles and proposing five generic and measurable components, and to the generic computational methodology for role identification. Throughout Chapter 4 to Chapter 6, this thesis has demonstrated how to operationalize each role facet to model social roles in two socially important contexts - Wikipedia and Cancer Survivor Network. Via combining theories about social roles and computational models for identifying roles on those two large-scale platforms, this research reveals details about emergent, behavioral, functioning roles in two different environments, and a set of computational techniques to identify such roles via fine-grained operationalization of role holders' behaviors. This work fills the longstanding gap in role theory and empirical modeling about emergent roles in online communities, which lays the foundation for future work to identify and analyze various roles that people actually enacted in group processes both online and offline. This present chapter summarizes the findings from previous chapters and presents the limitations of this thesis, as well as directions for future research.

7.1 Summary of Findings & Contributions

This thesis lays the foundation for computational modeling of social roles by introducing a five-facet role framework and investigating each of those facets in two socially important contexts. Detailed contributions are summarized as follows:

The Facet of Interaction & Role Identification on Wikipedia

Chapter 4 focuses on developing new techniques to identify roles that editors enact when editing Wikipedia articles and investigate how work contributed by people from different roles affect the article quality. From a theoretical perspective, it looks at the facet of **Interaction** in our five-facet role framework to represent editors, and **strictly** follow the generic role identification method to postulate roles, identify roles and evaluate roles. Specifically, we developed 24 edit categories to understand how different users perform the editing task collaboratively, and developed machine learning methods for the automated measurement of these edit categories revealed in users' edits. In our taxonomy, edits (editors' interactions with articles) are distinguished contextually in terms of the object being edited (e.g. information, template, reference, etc.) and functionally, in terms of the edit operation (e.g. insert, delete, modify, etc.). Building on this automated measurement of edit types, we use a graphical model that treats an editor as comprising multiple roles at the same time to identify the latent roles editors occupy. This work revealed eight functional roles such as Face Checker and Copy Editor — some of which are not directly reflected by prior methods, and demonstrated that different sets of roles are needed in the different quality stages of the article. Overall, this work paves a way for future research to automatically identify fine granularity edit types for Wikipedia editors, to extract a mixture of editor roles and to encourage specific role setting to improve the quality of articles. The present work also helps in how to develop intelligent task routing systems to recommend users to tasks that match their expertise.

The Facet of Goal on Wikipedia

Chapter 5 explicitly models the facet of **Goal** in our role framework to improve role representation (postulation and definition), in order to potentially help the identification of roles occupied by editors on Wikipedia. As an effort to differentiate editors who occupy different editing roles, we introduced a generic and fine-grained taxonomy

of the reasons why an author in Wikipedia made an edit. Example edit intentions include copyediting, elaboration, verification, and simplification. This clean higher-level semantic categorization enables us to easily identify textual meaning changes and to connect revisions to “what happens in the mind of the revising author during the revision” (Fitzgerald, 1987; Daxenberger, 2016). We contributed both research data resources and computational models to identify these edit intentions from differences between revisions of Wikipedia articles. The automated measurement of edit intentions provides a general framework to analyze revisions. We demonstrate two examples of how this intention taxonomy can be applied to better understand the success of online collaboration communities (Kraut et al., 2010), specifically the process of these sites to retain new contributors and create innovative products. These findings showed that specific types of editing work were positively correlated with newcomer survival and articles in different stages of development benefited differently from different types of edits. This work on edit intention taxonomy can facilitate a set of downstream NLP applications, such as collecting specific types of revisions (Yatskar et al., 2010; Recasens et al., 2013; Zanzotto and Pennacchiotti, 2010), outlining the evolution of roles (Arazy et al., 2015; Yang et al., 2016a), detecting quality flaws (Anderka et al., 2012) and providing insights on which specific aspects of an article needs improvement and what type of work should be performed. The ability to identify the need for editing, and specifically the types of editing work required, can greatly assist not only collaborative writing but also individual improvement of text. Beyond the context of Wikipedia, this work can inform the design of goal facet for better analyzing the collaboration and interaction happened in other online contexts such as academic writing (e.g., Google Docs or ShareLatex, etc).

Multiple Facets and Role Identification on CSN

Different from the studies in Chapter 4 and Chapter 5, Chapter 6 examines the problem of role identification in a new context - Cancer Survivors Network. We demonstrate how to utilize prior work on online health communities to provide guidance and supervision for role-related behaviors — **role postulation**, to operationalize multiple facets of our role framework (**Interaction, Goal, Context, Expectation**) — **role definition**, to extract a set of coherent roles that can well explain participants’ behaviors — **role identification**, and to validate the derived roles comprehensively via four evaluation manners: quantitatively in terms of model fit on the held-out data, qualitatively in terms of domain experts’ interpretation, directly in terms of correlations with input from role

holders, and indirectly in terms of performance boost to downstream applications — **role evaluation**. A set of behavioral features were operationalized to represent each component and then unsupervised models were employed to extract the functioning roles that members occupy, which discovered 11 interpretable roles in online cancer support groups. Follow-up analyses on roles showed that occupying socially positive roles, such as private communicator and story sharer, is associated with members staying in the community longer, while members occupying roles such as informational support seeker are associated with lower long-term participation in the community. While the distribution of roles in the community is relatively stable over time, members change their roles frequently across their participation. A closer look at members' role transitions suggests that they frequently change their roles from seeking resources to roles that offer help to others. These findings suggest the value of the role framework as the basis for intervention in online health communities, opening a new opportunity for socio-technical systems to support users and communities in their healthcare needs. Chapter 6 also showed potentials that most existing empirical methods for identifying roles in other domains (Yang et al., 2016a; Arazy et al., 2016) can be abstracted into this generic methodology, which can be applied to any other types of community, both online and offline. Practically, our role modeling methods can be employed to develop tools that detect members' needs, track their activities, and connect users to other users, content, and tasks.

7.2 Limitation

Expectation

This thesis does not look at **Expectation** in the process of social role identification or understanding. As discussed earlier, roles are thought to be associated with shared expectations among role holders and others, i.e., role behaviors are predicted or regulated by the expectation. For example, the audience is expected to cheer in a football match and would be inappropriate if they did it in the church. In most cases, such expectations are implicit, informal, and not written in online communities, making it challenging for any empirical investigation. For example, our work in the context of online cancer support groups revealed that there are a set of functioning roles, such as *welcomers*, *support providers*, however, such self-selected roles are usually not explicitly visible to or recognized by other community members, suggesting that there will not be any penalty

or loss when *welcomers* do not welcome newcomers, or when *support providers* decline to provide care or encouragement in a context where emotional support is highly needed. On CSN, there is no explicit code of conducts for how to enact a role at a platform level. As a result, we are not able to examine such **deviations** of expected role behaviors due to the limitations of our research context. Similarly, standards for emergent functioning roles such as *copy editors*, or *fact checkers* are also not available on Wikipedia. As a result, individuals may have their own understandings of whether they should obey the rules and regulations, and to what extent. This introduces challenges for both utilizing expectations for role identification and understanding whether role holders behave well on their roles.

Context

Roles are generally associated with specific contexts. For example, team roles like *critic* and *note-taker* exist in the context of teamwork but do not occur for a party context. Our two role identification studies do not go deep into the facet of context. When identifying editors' roles on Wikipedia, we assumed the context of editing is already provided and did not take into account editors' other type of behaviors such as defending others' edits in the talk pages. Similarly, topic-based contexts (e.g., Wikipedia articles on sports, biology or music) are also not utilized for identifying editors' roles. Although we chose English Wikipedia, similar methodologies can be adjusted to other language versions of Wikipedia. Note that, switching the context from English Wikipedia to Italian or Chinese Wikipedia may produce different sets of roles or roles with different percentages, because of the population bias or demographic differences. When modeling the behavioral roles of members in online cancer support groups, we simply divided context via members' communication channels into a private context and a public context, which facilitates the identification of context-aware roles such as *private support provider* and *public support provider*. However, we did not further differentiate the contexts to take care of finer-granularity of thread themes or expertise areas. For instance, members may be *informational support providers* in subforums about their cancers, but can be *emotional support providers* for patients or caregivers from other cancer sub-forums.

Methodology for Role Identification

Our methodology for role identification is in a pipeline manner, which begins with operationalizing features from different facets of roles in a specific context followed

by clustering such heterogeneous representations of users. However, if there are any inaccuracies or misrepresentations in those constructed user features, such cascade of errors may largely influence the derived roles. For instance, for role identification on CSN, the features we used to characterize users may not well capture their defending behaviors, thus we failed to capture the defender role. Despite our reasonable feature operationalization of different facets from role framework, we acknowledge this as a limitation and urge future work to design end-to-end techniques for role identification.

7.3 Future Work

This thesis opens up several research directions that deserve further pursuit.

7.3.1 Cross-community Role

This thesis has demonstrated the success of building computational models to identify social roles. For example, in the context of Wikipedia, we computationally identified roles such as substantive expert, fact checker, copy editor and markup maven, based on an empirically derived taxonomy of 24 edit types (e.g., inserting a reference, deleting a sentence, fixing grammar) and 13 edit intentions (e.g., simplification, vandalism, elaboration) (Yang et al., 2017a, 2016a). In the context of health support groups, we computationally identified roles such as emotional support provider, story sharer, information support seeker, based on empirically derived common conversational acts, such as providing informational support, self-disclosing negative experience or seeking emotional support (Yang et al., 2019b). However, these roles are primarily community-specific ones. Future work can advance this thesis by developing social-role models to distinguish trans-community roles from community-specific roles in online production communities, which may require connecting the low-level actions in a community to well-defined roles that transcend the community. The intuition is that volunteers enacting specific roles often follow similar patterns of behavior to communicate their goals to others in different environments. For example, those in leadership roles will exhibit executive management behaviors such as delegating tasks to others, contributing to community building, providing feedback to others, and influencing others to adopt their vision for the group. Although the language used to accomplish these goals may differ across communities, the core actions of community building, providing feedback or influencing others are common across them.

7.3.2 Role and Expectation

This thesis points several directions towards the modeling of expectation and its association with roles and role performance in different online communities. Overall, the facet of expectation can be operationalized in diverse ways, including, but not limited to, whether members are complaining to administrators or moderators about others' inappropriate behaviors or whether role holders are telling others what to do, whether role holders are receiving positive or negative feedback in terms of their behaviors, and whether role holders are behaving correspondingly, following the standards and requirements. Each manifestation of expectation can be translated into a set of features such as role holders' usage of modal words, their message exchange with moderators, the number of received "likes" or upvotes, etc.

Most online communities do not have an explicit code of conducts for emergent, behavioral roles, thus it is often unclear for members on which roles to occupy and how to perform such roles as demonstrated in our case studies. However, certain standards for a small set of assigned roles may exist, which can be utilized to better model expectation. For instance, on Wikipedia, there exist guidelines and policies about how editors should normally follow for several assigned roles such as *administrators*, which we have not yet examined. However, these explicit standards could be used to assess what editors' roles are and what roles they are supposed to occupy — behavioral roles and assigned roles. For instance, [Burke and Kraut \(2008\)](#) conducted empirical studies to analyze editors' behavioral criteria that correspond to the role of administrator, and their correlations with editors' promotion success to administrator status in Wikipedia. These expectation measures also shed lights on how to learn expectations, and how to utilize such aggregated expectations to assess how well the role holders perform their roles and to provide guidance to other role holders on how they should behave. In addition to written standards, on Wikipedia, there is other auxiliary information available that can be used to identify roles such as the ways of how people react to the role-holders' actions. For instance, Wikipedia reverts¹ reverse or undo one's prior edits, carrying implication for vandalism, disruptive edits or disagreement. In contrast, the Thanks Notification² and Barnstars³ offer ways to give positive feedback to users who make useful edits. Such external reactions rather than actions that the role holder

¹<https://en.wikipedia.org/wiki/Wikipedia:Reverting>

²<https://en.wikipedia.org/wiki/Wikipedia:Notifications/Thanks>

³<https://en.wikipedia.org/wiki/Wikipedia:Barnstars>

performs can act as implicit feedback for role identification and assessment.

7.3.3 Role and Supervision

Most role identification studies have been conducted in an unsupervised manner (Welser et al., 2011; Fazeen et al., 2011; Agarwal et al., 2008) where the goal is to learn hidden structures in data. Those derived roles are not typically defined in terms of what they are meant to accomplish, although they may be associated with kinds of things they do. Role identification guided by supervisions or human expectations may be more useful to practitioners, such as extracting roles that can facilitate group success or accomplish group task. Several empirical studies have started to use such supervisions to weakly guide the identification of roles (Maki et al., 2017; Yang et al., 2015; Ferschke et al., 2015). Our role framework defined roles in terms of what the role holders are meant to accomplish by incorporating the facet of **goal**, such as *newcomer welcomers* and *knowledge promoters*, but “supervision” is indirectly modeled in the phase of feature construction. At a team or group level, different anthologies of roles from theories in social science and psychology already exist (Benne and Sheats, 1948). How can role identification models be adjusted to reveal such roles and to utilize human insights? Furthermore, in other cases, the facet of goals may shape roles. For instance, editors volunteer to edit articles on Wikipedia by enacting a set of editing roles. Such role distribution and occupation are likely to be affected if editors got paid for their edits, especially if certain tasks were rewarded more. Therefore, individuals’ choices of roles may be shaped by those group and individual level goals. Future work could focus on deciphering the interdependence between the roles, tasks, and goals.

Note that our five facet role framework and computational role methodology can also be generalized to offline organizations, for instance, to figure out what types of roles employees play in companies based on their working history. Most organizations keep records of their employee’s background, working logs, emails, and performances, which can serve as the basis for facilitating and supervising the identification of users’ roles, and the interdependence between their roles, tasks, and goals.

7.3.4 Role Generation

This thesis mainly focused on the identification or discovery of social roles. One natural follow up is towards generation with roles or personas. Specifically, this refers to

imposing roles to conversational agents — software systems that can interact or chat with users in natural language. Such roles may include production roles that help users complete specific tasks such as copy-editing articles or booking flights, or personality roles that communicate with users like humans such as ELIZA (Weizenbaum, 1966), Jabberwacky and ALICE (Robin, 2007). For instance, the movie *Her* personified an artificially intelligent virtual assistant “Samantha” through a female voice, who reveals a sensitive, playful personality. Such role-specific behaviors and language styles can be learned from exponentially increasing human-human and human-computer interaction data, and can be incorporated into the generation process for better user experiences. User modeling has received extensive attention in the field of conversational agents (Kobsa and Wahlster, 1989; Lin and Walker, 2011). Recently, there has been a number of research studies that endow data-driven systems with “persona”, character styles (Li et al., 2016), or task completion abilities. For instance, Zhang et al. (2018) endowed chit-chat conversational agents with configurable persona, which can produce more personal, specific and engaging responses than persona-free models. Guru⁴ is a conversational agent with built-in grammar corrector for helping copy-editing as an educational tool for English, and Chatbot Lucy serves as a language tutor to conduct conversations with learners to help them with language learning (Wang and Petrina, 2013). This type of work sheds lights on practically encoding roles and role-related actions to facilitate the generation of role-specific language or behaviors in downstream applications.

7.3.5 Role and Stereotype

Certain behavioral characteristics of roles might be associated with stereotypes, especially in terms of basic roles such as gender (e.g., female vs. male), race (e.g., black vs. white), job roles such as a doctor, nurse, professor, and team roles such as leader, critic. For example, gender stereotypes (Basow, 1992; Rudman and Glick, 2001) about what is appropriate for females and males “limit their societal roles, thereby affecting their participation in the labor force and their contributions to their families” (Berkery et al., 2014). Hurwitz and Peffley (1997) found a strong relationship between whites’ images of African-Americans and their judgments of crime and punishment for black criminals who commit violent crimes. Recent advances in machine learning especially word embedding also revealed gender biases in text data and computational models (Bolukbasi et al., 2016b,a; Zhao et al., 2017). Garg et al. (2018) used word embeddings to

⁴https://www.chatbots.org/conversational_agent/grammar_guru/

characterize how different groups of ethnic minorities are viewed during the 20th and 21st centuries starting from 1910 and found that both gender and ethnic occupation biases in the embeddings significantly track with the actual occupation participation rates. These stereotypes become even more severe when a large amount of user-generated data is used for role identification or other tasks such as semantic role labeling (Zhao et al., 2017), word embedding (Garg et al., 2018). To what extent can we measure how much the derived roles correlated with biases, and how are roles perceived by other individuals? From a longer term, there is much more to be done in investigating how roles can reveal insights about human cognitive biases and cultural stereotypes.

7.3.6 Role Transitions

As people go through their participation in either online or offline communities, they may move from one role to another, or change their orientations toward a role already assumed. Chapter 6 explicitly models the stability and dynamics of roles and confirmed prior research that members frequently transit to other roles throughout their life cycles. While understandings of these emergent social roles are beginning to form, how and why individuals transit from one role to another still remains unclear. This paves ways for the computational modeling of individuals' roles from a temporal perspective. In addition to individuals' temporal role changes, can we model the associations between individuals' transitions of roles and their switches of contexts, audiences, goals, or expectations, such as a mother role vs. a professor role when sending messages to her children vs. her students? The more we can incorporate this temporal and context into our role identification process, the deeper we can understand role transitions, and the better they can help sustain user engagement and coordinate individuals to contribute to the goals of the communities. Note that practically, this line of research on role transitions requires rigorous modeling of time scale, including, but not limited to, defining an appropriate temporal unit to represent role behavior and tracking the transition of members' roles over time.

7.3.7 Micro Level Role Identification

This thesis focused on macro-level role identification, namely, identifying the roles that people play via role identification algorithms and validating it with different evaluation measures. Although it provides an overall aggregation of roles exists in a context, it does not explicitly explain each action done by a role from a micro level. One natural follow

up is to zoom into the process of how people perform their roles — what actions they do and under what conditions. This allows us to define two major semantic types: *action* which refers to words, behavior types, and *state* that denotes the current context where a specific action happens. For instance, example states could include goals or expectations in a context, the present of other roles, or other contextual factors. With such semantic types, we can define a list of action-state pairs with strengths for each role — {ROLE: <ACTION | STATE>} that describe at a state *s*, role *r* does action *a*. In this denotation system, the facet of goal and context can be represented as a list of states, while the expectation facet will become a list of {ROLE: <DESIRED ACTION | STATE>} pairs. This micro-level representation will enable us to (1) model expectations by comparing what a role occupant actually does with his/her desired behaviors, (2) assess whether role occupants are performing well on their roles or not, and (3) track how they change their role-specific actions over time.

For production-related roles like copy editor or substantive experts, this micro-level role framework requires an operationalization of people’s behaviors first. Take the context of editor role identification on Wikipedia for example, a structured way of representing editors’ edits such as our edit type taxonomy in Figure 4.1 or edit intention taxonomy in Table 5.1, is needed to create a list of *actions*. Example *states* include different namespaces on Wikipedia such as *Main/Article* or *Template*, different topical areas such as articles on *arts* or *politics*. When it comes to expectation around roles, we can translate standards associated with a role into such pairs. For instance, *administrators* on Wikipedia are required to demonstrate strong edit history, varied experience, user interaction, etc. (Burke and Kraut, 2008), whose expected role behaviors may be reflected in a larger set of action-state pairs, a more diverse set of states in terms of namespaces, a higher frequency of *User*, *User Talk*, *Talk* states. Such nuanced pairs of actions and states can result in finer-grained modeling of social roles, and enable us to examine the discrepancies between expected and actual actions.

Other scenarios of role identification such as occupation roles can also be modeled both at a macro level via our generic role framework, and at a micro level. There has been micro level work examining fine-grained social roles. For instance, Beller et al. (2014) identified a set of verbs (actions) that preferentially select a particular social role as an argument. In their example, descriptions like “*an artist drew ...*” or “*an artist designed ...*” were utilized to count all verbs (e.g., “*drew*”, “*designed*”) that appear with

a role (e.g., “*artist*”), and were further refined by looking at their point-wise mutual information with a specific role. Such modeling can be directly abstracted into our micro-level role representation with action-state pairs.

7.3.8 Role Balance in Groups

Although the performance of a team is determined by the team members’ roles, teams with a wider or larger set of social roles are not guaranteed to be linked with higher performances. Specifically, Belbin suggested that roles should not be duplicated, and balanced teams perform better than non-balanced teams (Belbin, 1993), and Ten Haaf et al. (2002) elaborated on the assertion that scarcity of one of the team roles is detrimental for the team performance. However, team balance is a complex subject and has mixed results (Senior, 1997). For example, Partington and Harris (1999) did not find any strong relationship between “team balance” and team performance. One possible explanation, suggested by van de Water et al. (2008), is that the the definition for *balanced teams* of (Ten Haaf et al., 2002) is not isomorphic with the qualitative notions of (Belbin, 1993). Therefore, the question of whether balanced teams perform better or not cannot be simply answered unless there is a well and uniquely definition for what constitutes balanced teams, and what are the methods used to determine the degree of balance.

7.3.9 Role Configuration in Groups

Many open questions exist about what roles and in what balance would make the ideal group composition (Neuman et al., 1999), and how those role configurations interact with other contextual factors (Senior, 1997; Meredith Belbin, 1981). Possible reasons include the difficulty of finding an adequate number of teams, an aggregated team-level performance measure that links well with individuals’ traits, the complete and detailed data dump of team interactions, and the operationalization of various composition variables related to individuals’ skills and abilities and teams’ social and task conditions. Thus, research resources and modeling approaches that can be used to understand role compositions that are particularly predictive for any given context would potentially have both highly practical and theoretical value.

On the other hand, an improper role composition may lead to the dysfunctions of groups. How can we perform role composition analysis to diagnose whether groups suffer from a missing role, an ill-functioning role, or other problematic role-relevant

issues? This altogether sheds lights on different role-related mechanisms behind group formation, evolution, performance, and success.

7.3.10 Role with Malicious Intent

Most roles that have been examined are about positive roles or roles beneficial to tasks or groups. Regarding bad roles in online communities, [Kumar et al. \(2017\)](#) found that some users create multiple sock-puppets and engage in malicious and deceptive behavior by deceiving others or manipulating discussions. There has been considerable work on identifying vandals in Wikipedia ([Adler et al., 2011](#)). For instance, [Kumar et al. \(2015\)](#) studied the problem of detecting vandals on Wikipedia, and tools such as ClueBot NG⁵, STiki⁶, and Snuggle⁷ were used heuristic rules and machine learning algorithms to flag acts of vandalism. In collaborative environments, there are also “bad roles”, such as social loafers in teamwork ([Kouliavtsev, 2012](#)). The study of [Jones et al. \(2011b\)](#) examined emergent user roles in asynchronous distributed collaborative idea generation, and revealed five user roles, two out of which are negative roles - Social Loafer and Absentee. Do bad roles in different environments demonstrate the same set of intermediate behaviors, and how can we extract them simultaneously?

Explicitly modeling roles with malicious intent may better help the understanding of the massive spread of fake news and misinformation, which have been alleged to influence elections, threaten democracies, and disrupt societies. Prior research found that social bots and certain members play a key role in the diffusion of misinformation ([Shao et al., 2017](#)). Thus, studying the roles of viral misinformation creators, propagators and tracking how those role holders spread information, where and when can help develop solutions to detect and reduce misinformation, which will make a meaningful impact on both defense and attack research. Although there have been extensive studies on (mis) information propagation or diffusion, few of them have actually modeled the roles of these members.

7.3.11 Privacy, Ethics, and Role

Identification of social roles usually involves inferring users’ non-behavioral attributes such as their age or gender, which requires special ethical attention ([Nov and Wattal,](#)

⁵https://en.wikipedia.org/wiki/User:ClueBot_NG

⁶<http://en.wikipedia.org/wiki/Wikipedia:STiki>

⁷<https://en.wikipedia.org/wiki/Wikipedia:Snuggle>

2009). In addition to personal sensitive information, the privacy type of platforms also matters. Our prior studies reveal that members often use private channels to develop close relationships with others, or to report to moderators about inappropriate behaviors (Yang et al., 2019c), suggesting the existence of various social roles in private. Our two role identification studies have adopted rigorous steps to protect users' privacy, such as (1) avoiding asking annotators to view or annotate private messages but to directly apply machine-learning models trained on the public discussion board, (2) paraphrasing all example quotes to make them less search-able via the Internet (Bruckman, 2002). However, how to understand users' functioning roles ethically in sensitive and private environments such as in mental health communities (De Choudhury and Kiciman, 2017) or very personal social sites (Rudder, 2014) remains challenging.

Special attention should also be paid to the scenarios of generation, simulation, and manipulation of roles for ethical considerations. For instance, experimental simulations have been conducted to seek optimal role configuration for groups, which may involve the assignment of real users to roles that do not fit their expertise or intentions. Similarly, roles may be introduced to manipulate the public and spread misinformation to disrupt societies, which deserves more attention. In addition to those cases, social role studies also need to consider fairness or bias in their models and outputs. For instance, a sensitive or protected attribute may not be used as a feature for role identification (Grgic-Hlaca et al., 2016), and the identified roles may need to generalize equally on specific groups. Future work should design guidelines and regulations for better informing and building secure, fair, and trust-able role related systems and algorithms.

Appendix A

Edit Intention Annotation Training

To expose annotators to more working knowledge of Wikipedia, we provided two-hour training session where five annotators were asked to label a small set of revisions and to discuss their disagreement until consensus. We describe the overall structure of this training session together with the time duration for each step as below. The detailed procedures are described in Section A.1 to Section A.3.

- **Edit Types and Task Explanation [20min]**
 - Explain the Task and what edit intentions are
- **Sample Annotation Task 1 [15min]**
 - Annotate 3 worksets (30 revisions)
- **Task 1 Agreement Check [25min]**
 - Assess the agreement between 5 annotators
 - Discuss the inconsistent annotation
- **Sample Annotation Task 2 [20min]**
 - Annotate 4 worksets (40 revisions)
- **Task 2 Agreement Check [25min]**
 - Assess the agreement between 5 annotators
 - Discuss the inconsistent annotation
- **Summary and Questions [10min]**
 - Summary and recap of two rounds annotation
 - Questions discussion

A.1 Annotation Task

The interface we designed for this annotation task allows you to (1) visualize a revision in diff format, (2) select a “semantic intention” for the edit, and (3) add an optional comment. For each revision, your task is to annotate why (intention) the editor made this edit: please select one or more of the possible values in the Semantic Intention dropdown menu. Tooltips will appear when you hover on each intention. A list of available Semantic intentions is below. Also, please judge whether this edit added/modified/removed information in this article. If information is not added, not modified, and not removed, please leave it as it is.

The semantic intention is describe in detail in https://en.wikipedia.org/wiki/Wikipedia:Labels/Edit_types/Taxonomy

A.2 Frequently Q & A

1. Can I select multiple intentions?
 - Yes, each revision can have multiple semantic intentions. (An editor could do multiple things at the same time)
2. What should I do if no intention applies?
 - If you think none of the semantic intentions accurately describes a revision, please label it as Other and leave comments in the Notes field.
3. Where does “adding a category” belong?
 - Adding Categories usually belongs to Wikification
4. Where does “adding an image” belong?
 - Adding images or files usually belongs to Elaboration
5. What is the difference between Copy Editing and Wikification?
 - Copy Editing refers to fixing generic grammar or spelling errors; Wikification refers to formatting the text to comply with Wikipedia’s manual of style, adding links, etc.
6. When should I not use Verifiability?
 - If an edit only changes the syntax of a reference or citation, it does not belong

to Verifiability but it should be labeled as Wikification.

7. What should I do for operations happened in the tables?

- If an edit only adds one row or column into the table, it belongs to Elaboration; if it adds or updates the cell value(data value), it belongs to Fact Update.

8. How should I label judge whether information is modified?

- Formatting or copy-editing often does not involve information changes; adding new sentences, figures, files, tables (rows) might change the information.

A.3 The Annotation Interface

To use the annotation interface, please follow the below instructions.

- Step 1: Please send your Wikipedia username to Diyi Yang
- Step 2: Go to this page <https://en.wikipedia.org/wiki/Wikipedia:Labels>, and install the gadget. The installation instructions are here https://meta.wikimedia.org/wiki/Wiki_labels#Installation.

In order to perform difficult analyses and train intelligent wiki-tools (e.g. for detecting vandalism and assessing the quality of articles), we need **labeled data** and lots of it. **Wikipedia:Labels** is the project on the English Wikipedia for getting the labeling work done. We use the **m:Wiki labels** tool to collaboratively label wiki artifacts quickly and easily.

2016-02-16 (10/10)	review
2016-02-17 (10/10)	review
2016-02-17 (10/10)	review
2016-02-18 (10/10)	review
2016-02-18 (10/10)	review
2016-02-18 (10/10)	review
2016-02-19 (10/10)	review
2016-03-20 (0/10)	open

request workset

Resources [edit source]

- Documentation
- Repository [↗](#) (MIT License)
- File a bug report [↗](#)

Categories: [Wikipedia administration](#)

Figure A.1: The annotation interface on Wikipedia

Please go to the below page to post your concerns, questions or suggestions about the annotation. https://en.wikipedia.org/wiki/Wikipedia_talk:Labels/Edit_types

Appendix B

Interview for Role Validation

To validate and provide names for (i.e., label) the roles, we conducted structured feedback sessions with six moderators split across two sessions and conducted over Skype. Each session contained three elements.

It began with a 5-min introduction to the task of naming roles. In addition to our verbal explanation, we also provided detailed instructions in text. Specifically, we used plain language to describe typical behaviors for each role (i.e., top ranked features in Table 2) and example messages from three representative users occupying that role.

We then gave moderators around 40 min to read the typical behaviors and representative messages associated with each role and use keywords to label it. We asked them to use the role's typical behaviors as the main basis for labeling roles and use the messages to help facilitate their decisions. We provided information about the 11 roles and a free text-labeling interface in a separate Google slide for each moderator. Thus moderators shared instructions and information about the roles, but could label them independently without being influenced by others' role names.

At the end of the session, the first author summarized the names the three moderators in a session gave and led a discussion about why they made their choices, whether they agreed with each other, and what would be an acceptable name for each role. This section lasted for about 35 minutes. The Google slide used for this annotation can be found here: <https://goo.gl/WslPdD>.

Bibliography

- B Thomas Adler, Luca De Alfaro, Santiago M Mola-Velasco, Paolo Rosso, and Andrew G West. Wikipedia vandalism detection: Combining natural language, metadata, and reputation features. In *International Conference on Intelligent Text Processing and Computational Linguistics*, pages 277–288, 2011.
- Nitin Agarwal, Huan Liu, Lei Tang, and Philip S Yu. Identifying the influential bloggers in a community. In *Proceedings of the 2008 International Conference on Web Search and Data Mining*, pages 207–218. ACM, 2008.
- Edoardo M. Airoldi, David M. Blei, Stephen E. Fienberg, and Eric P. Xing. Mixed membership stochastic blockmodels. volume 9, pages 1981–2014. JMLR.org, June 2008.
- Ablimit Aji, Yu Wang, Eugene Agichtein, and Evgeniy Gabrilovich. Using the past to score the present: Extending term weighting models through revision history analysis. In *Proceedings of the 19th ACM international conference on Information and knowledge management*, pages 629–638. ACM, 2010.
- George A. Akerlof and Rachel E. Kranton. Economics and identity. *Quarterly Journal of Economics*, 115(3):715–753, August 2000.
- Elie Aljalbout, Vladimir Golkov, Yawar Siddiqui, Maximilian Strobel, and Daniel Cremers. Clustering with deep learning: Taxonomy and new methods. *arXiv preprint arXiv:1801.07648*, 2018.
- Maik Anderka, Benno Stein, and Nedim Lipka. Predicting quality flaws in user-generated content: The case of wikipedia. In *Proceedings of the 35th International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR '12*, pages 981–990, New York, NY, USA, 2012. ACM.
- O. Arazy, H. Lifshitz, O. Nov, J. Daxenberg, M. Balestra, and C. Cheshite. On the how and why of emergent role behaviors in wikipedia. In *Proceedings of the ACM SIGCHI*

- Conference on Computer Supported Cooperative Work*. ACM, 2017.
- Ofer Arazy, Felipe Ortega, Oded Nov, Lisa Yeo, and Adam Balila. Functional roles and career paths in wikipedia. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing, CSCW '15*, pages 1092–1105, New York, NY, USA, 2015. ACM.
- Ofer Arazy, Johannes Daxenberger, Hila Lifshitz-Assaf, Oded Nov, and Iryna Gurevych. Turbulent stability of emergent roles: The dualistic nature of self-organizing knowledge coproduction. *Information Systems Research*, 27(4):792–812, 2016.
- Blake Ashforth. *Role transitions in organizational life: An identity-based perspective*. Routledge, 2000.
- Sairam Balani and Munmun De Choudhury. Detecting and characterizing mental health related self-disclosure in social media. In *Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems*, pages 1373–1378. ACM, 2015.
- David Bamman, Brendan O'Connor, and Noah A. Smith. Learning latent personas of film characters. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 352–361, Sofia, Bulgaria, August 2013. Association for Computational Linguistics.
- David Bamman, Ted Underwood, and Noah A Smith. A bayesian mixed effects model of literary character. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics*, volume 1, pages 370–379, 2014.
- Susan A Basow. *Gender: Stereotypes and roles*. Thomson Brooks/Cole Publishing Co, 1992.
- Bernard M Bass and Ralph Melvin Stogdill. *Bass & Stogdill's handbook of leadership: Theory, research, and managerial applications*. Simon and Schuster, 1990.
- Thomas S Bateman and Stephen Strasser. A longitudinal analysis of the antecedents of organizational commitment. *Academy of management journal*, 27(1):95–112, 1984.
- Talya N Bauer, Todd Bodner, Berrin Erdogan, Donald M Truxillo, and Jennifer S Tucker. Newcomer adjustment during organizational socialization: a meta-analytic review of antecedents, outcomes, and methods. *Journal of applied psychology*, 92(3):707, 2007.
- Natalya N Bazarova, Jessie G Taft, Yoon Hyung Choi, and Dan Cosley. Managing impressions and relationships on facebook: Self-presentational and relational con-

- cerns revealed through the analysis of language style. *Journal of Language and Social Psychology*, 32(2):121–141, 2013.
- Natalya N Bazarova, Yoon Hyung Choi, Victoria Schwanda Sosik, Dan Cosley, and Janis Whitlock. Social sharing of emotions on facebook: Channel differences, satisfaction, and replies. In *Proceedings of the 18th ACM conference on computer supported cooperative work & social computing*, pages 154–164. ACM, 2015.
- R Meredith Belbin. *Team roles at work*. Routledge, 1993.
- Charley Beller, Craig Harman, and Benjamin Van Durme. Predicting fine-grained social roles with selectional preferences. In *Proceedings of the ACL 2014 Workshop on Language Technologies and Computational Social Science*, pages 50–55, 2014.
- Kenneth D Benne and Paul Sheats. Functional roles of group members. *Journal of social issues*, 4(2):41–49, 1948.
- Elaine Berkery, Siobhan Tiernan, and Michael Morley. The relationship between gender role stereotypes and requisite managerial characteristics: the case of nursing and midwifery professionals. *Journal of nursing management*, 22(6):707–719, 2014.
- Bruce J Biddle. *Role theory: Expectations, identities, and behaviors*. Academic Press New York, 1979.
- Prakhar Biyani, Cornelia Caragea, Prasenjit Mitra, and John Yen. Identifying emotional and informational support in online health communities. In *COLING*, pages 827–836, 2014.
- David M Blei, Andrew Y Ng, and Michael I Jordan. Latent dirichlet allocation. *the Journal of machine Learning research*, 3:993–1022, 2003.
- Herbert Blumer. *Symbolic interactionism: Perspective and method*. Univ of California Press, 1986.
- Jesús Bobadilla, Fernando Ortega, Antonio Hernando, and Abraham Gutiérrez. Recommender systems survey. *Knowledge-based systems*, 46:109–132, 2013.
- Freimut Bodendorf and Carolin Kaiser. Detecting opinion leaders and trends in online social networks. In *Proceedings of the 2nd ACM workshop on Social web search and mining*, pages 65–68. ACM, 2009.
- Tolga Bolukbasi, Kai-Wei Chang, James Zou, Venkatesh Saligrama, and Adam Kalai. Quantifying and reducing stereotypes in word embeddings. *arXiv preprint arXiv:1606.06121*, 2016a.

- Tolga Bolukbasi, Kai-Wei Chang, James Y Zou, Venkatesh Saligrama, and Adam T Kalai. Man is to computer programmer as woman is to homemaker? debiasing word embeddings. In *Advances in Neural Information Processing Systems*, pages 4349–4357, 2016b.
- Philip Bramsen, Martha Escobar-Molano, Ami Patel, and Rafael Alonso. Extracting social power relationships from natural language. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies-Volume 1*, pages 773–782. Association for Computational Linguistics, 2011.
- Amit Bronner and Christof Monz. User edits classification using document revision histories. In *Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics, EACL'12*, pages 356–366, 2012. ISBN 978-1-937284-19-0.
- Amy Bruckman. Studying the amateur artist: A perspective on disguising data collected in human subjects research on the internet. *Ethics and Information Technology*, 4(3):217–231, 2002.
- C Shawn Burke, Kevin C Stagl, Cameron Klein, Gerald F Goodwin, Eduardo Salas, and Stanley M Halpin. What type of leadership behaviors are functional in teams? a meta-analysis. *The Leadership Quarterly*, 17(3):288–307, 2006.
- Moira Burke and Robert E. Kraut. Mopping up: Modeling wikipedia promotion processes. In *CSCW 2008: Proceedings of the ACM Conference on Computer-Supported Cooperative Work*, New York, 2008. ACM Press.
- Patricia J Bush. The male nurse: A challenge to traditional role identities. In *Nursing Forum*, volume 15, pages 390–405. Wiley Online Library, 1976.
- Stevie Chancellor, Andrea Hu, and Munmun De Choudhury. Norms matter: Contrasting social support around behavior change in online weight loss communities. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, page 666. ACM, 2018.
- Jonathan Chang, Sean Gerrish, Chong Wang, Jordan L Boyd-Graber, and David M Blei. Reading tea leaves: How humans interpret topic models. In *Advances in neural information processing systems*, pages 288–296, 2009.
- R.B. Cialdini, C.A. Kallgren, and R.R. Reno. A focus theory of normative conduct: A theoretical refinement and reevaluation of the role of norms in human behavior. *Advances in experimental social psychology*, 24:201–234, 1991.
- Jose M Cortina. What is coefficient alpha? an examination of theory and applications.

Journal of applied psychology, 78(1):98, 1993.

Dan Cosley, Dan Frankowski, Loren Terveen, and John Riedl. Suggestbot: Using intelligent task routing to help people find work in wikipedia. In *Proceedings of the 12th International Conference on Intelligent User Interfaces, IUI '07*, pages 32–41, New York, NY, USA, 2007. ACM. ISBN 1-59593-481-2. doi: 10.1145/1216295.1216309.

Cristian Danescu-Niculescu-Mizil, Michael Gamon, and Susan Dumais. Mark my words!: linguistic style accommodation in social media. In *Proceedings of the 20th international conference on World wide web*, pages 745–754. ACM, 2011.

Johannes Daxenberger. *The Writing Process in Online Mass Collaboration: NLP-Supported Approaches to Analyzing Collaborative Revision and User Interaction*. PhD thesis, Technische Universität, 2016.

Johannes Daxenberger and Iryna Gurevych. A corpus-based study of edit categories in featured and non-featured Wikipedia articles. In *Proceedings of COLING 2012*, pages 711–726, Mumbai, India, December 2012. The COLING 2012 Organizing Committee.

Johannes Daxenberger and Iryna Gurevych. Automatically classifying edit categories in Wikipedia revisions. In *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*, pages 578–589, Seattle, Washington, USA, October 2013. Association for Computational Linguistics.

Munmun De Choudhury and Sushovan De. Mental health discourse on reddit: Self-disclosure, social support, and anonymity. In *ICWSM*, 2014.

Munmun De Choudhury and Emre Kiciman. The language of social support in social media and its effect on suicidal ideation risk. In *Proceedings of the... International AAAI Conference on Weblogs and Social Media. International AAAI Conference on Weblogs and Social Media*, volume 2017, page 32. NIH Public Access, 2017.

Gabriel De la Calzada and Alex Dekhtyar. On measuring the quality of wikipedia articles. In *Proceedings of the 4th Workshop on Information Credibility, WICOW '10*, pages 11–18, New York, NY, USA, 2010. ACM. ISBN 978-1-60558-940-4.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*, 2018.

Gintare Karolina Dziugaite and Daniel M Roy. Neural network matrix factorization. *arXiv preprint arXiv:1511.06443*, 2015.

- Helen Rose Fuchs Ebaugh. *Becoming an ex: The process of role exit*. University of Chicago Press, 1988.
- Martin Ester, Hans-Peter Kriegel, Jörg Sander, Xiaowei Xu, et al. A density-based algorithm for discovering clusters in large spatial databases with noise. In *Kdd*, volume 96, pages 226–231, 1996.
- Lester Faigley and Stephen Witte. Analyzing revision. *College composition and communication*, 32(4):400–414, 1981.
- Mohamed Fazeen, Ram Dantu, and Parthasarathy Guturu. Identification of leaders, lurkers, associates and spammers in a social network: context-dependent and context-independent approaches. *Social Network Analysis and Mining*, 1(3):241–254, 2011.
- Ronen Feldman, James Sanger, et al. *The text mining handbook: advanced approaches in analyzing unstructured data*. Cambridge university press, 2007.
- A Mufit Ferman, James H Errico, Peter van Beek, and M Ibrahim Sezan. Content-based filtering and personalization using structured metadata. In *Proceedings of the 2nd ACM/IEEE-CS joint conference on Digital libraries*, pages 393–393. ACM, 2002.
- Oliver Ferschke, Johannes Daxenberger, and Iryna Gurevych. A survey of nlp methods and resources for analyzing the collaborative writing process in wikipedia. In *The People’s Web Meets NLP*, pages 121–160. Springer, 2013.
- Oliver Ferschke, Diyi Yang, and Carolyn P Rosé. A lightly supervised approach to role identification in wikipedia talk page discussions. In *Proceedings of the International AAAI Conference on Weblogs and Social Media. Workshop on Wikipedia, a Social Pedia: Research Challenges and Opportunities*, pages 43–47, 2015.
- Samantha Finkelstein, Evelyn Yarzebinski, Callie Vaughn, Amy Ogan, and Justine Cassell. The effects of culturally congruent educational technologies on student achievement. In *International Conference on Artificial Intelligence in Education*, pages 493–502. Springer, 2013.
- Danyel Fisher, Marc Smith, and Howard T. Welser. You are who you talk to: Detecting roles in usenet newsgroups. In *Proceedings of the 39th Annual Hawaii International Conference on System Sciences - Volume 03, HICSS ’06*, pages 59.2–, Washington, DC, USA, 2006. IEEE Computer Society.
- Robert J Fisher. Social desirability bias and the validity of indirect questioning. *Journal of consumer research*, 20(2):303–315, 1993.

- Jill Fitzgerald. Research on revision in writing. *Review of educational research*, 57(4):481–506, 1987.
- Peter Kin-Fong Fong and Robert P. Biuk-Aghai. What did they do? deriving high-level edit histories in wikis. In *Proceedings of the 6th International Symposium on Wikis and Open Collaboration*, WikiSym '10, pages 2:1–2:10, New York, NY, USA, 2010. ACM. ISBN 978-1-4503-0056-8.
- Denae Ford, Justin Smith, Philip J Guo, and Chris Parnin. Paradise unplugged: Identifying barriers for female participation on stack overflow. In *Proceedings of the 2016 24th ACM SIGSOFT International Symposium on Foundations of Software Engineering*, pages 846–857. ACM, 2016.
- Wikimedia Foundation. Wikipedians, Aug 17 2017.
- Viola Ganter and Michael Strube. Finding hedges by chasing weasels: Hedge detection using wikipedia tags and shallow linguistic features. In *Proceedings of the ACL-IJCNLP 2009 Conference Short Papers*, ACL Short '09, pages 173–176, 2009.
- Nikhil Garg, Londa Schiebinger, Dan Jurafsky, and James Zou. Word embeddings quantify 100 years of gender and ethnic stereotypes. *Proceedings of the National Academy of Sciences*, 115(16):E3635–E3644, 2018.
- R. Stuart Geiger and Aaron Halfaker. Using edit sessions to measure participation in wikipedia. In *Proceedings of the 2013 Conference on Computer Supported Cooperative Work*, CSCW '13, pages 861–870, New York, NY, USA, 2013. ACM. ISBN 978-1-4503-1331-5.
- Felix A Gers, Jürgen Schmidhuber, and Fred Cummins. Learning to forget: Continual prediction with lstm. 1999.
- Erving Goffman. Presentation of self in everyday life. *American Journal of Sociology*, 55: 6–7, 1949.
- Erving Goffman. Presentation of self in everyday life. 1959.
- Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. In *Advances in neural information processing systems*, pages 2672–2680, 2014.
- Nina Grgic-Hlaca, Muhammad Bilal Zafar, Krishna P Gummadi, and Adrian Weller. The case for process fairness in learning: Feature selection for fair decision making. In *NIPS Symposium on Machine Learning and the Law*, volume 1, page 2, 2016.

- Aditya Grover and Jure Leskovec. node2vec: Scalable feature learning for networks. In *Proceedings of the 22nd ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 855–864. ACM, 2016.
- Scott A. Hale. Multilinguals and wikipedia editing. In *Proceedings of the 2014 ACM Conference on Web Science, WebSci '14*, pages 99–108, New York, NY, USA, 2014. ACM. ISBN 978-1-4503-2622-3.
- Aaron Halfaker, R. Stuart Geiger, Jonathan Morgan, and John Riedl. The rise and decline of an open collaboration system: How wikipedia’s reaction to sudden popularity is causing its decline. *American Behavioral Scientist*, 57(5):664–688, May 2013.
- Will Hamilton, Zhitao Ying, and Jure Leskovec. Inductive representation learning on large graphs. In *Advances in Neural Information Processing Systems*, pages 1024–1034, 2017.
- Manoj Harpalani, Michael Hart, Sandesh Singh, Rob Johnson, and Yejin Choi. Language of vandalism: Improving wikipedia vandalism detection via stylometric analysis. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, pages 83–88, Portland, Oregon, USA, June 2011. Association for Computational Linguistics.
- Xiangnan He, Lizi Liao, Hanwang Zhang, Liqiang Nie, Xia Hu, and Tat-Seng Chua. Neural collaborative filtering. In *Proceedings of the 26th International Conference on World Wide Web*, pages 173–182. International World Wide Web Conferences Steering Committee, 2017.
- Jerold Heiss. Social roles. 1990.
- Keith Henderson, Brian Gallagher, Tina Eliassi-Rad, Hanghang Tong, Sugato Basu, Leman Akoglu, Danai Koutra, Christos Faloutsos, and Lei Li. Rolx: structural role extraction & mining in large graphs. In *Proceedings of the 18th ACM SIGKDD international conference on knowledge discovery and data mining*, pages 1231–1239. ACM, 2012.
- Malcolm Higgs, Ulrich Plewnia, and Jorg Ploch. Influence of team composition and task complexity on team performance. *Team Performance Management: An International Journal*, 11(7/8):227–250, 2005.
- Charles G Hill, Maren Haag, Alannah Oleson, Chris Mendez, Nicola Marsden, Anita Sarma, and Margaret Burnett. Gender-inclusiveness personas vs. stereotyping: Can we have it both ways? In *Proceedings of the 2017 CHI Conference on Human Factors in*

- Computing Systems*, pages 6658–6671. ACM, 2017.
- Meiqun Hu, Ee-Peng Lim, Aixin Sun, Hady Wirawan Lauw, and Ba-Quy Vuong. Measuring article quality in wikipedia: Models and evaluation. In *Proceedings of the Sixteenth ACM Conference on Conference on Information and Knowledge Management, CIKM '07*, pages 243–252, New York, NY, USA, 2007. ACM.
- Xia Hu and Huan Liu. Social status and role analysis of palin’s email network. In *Proceedings of the 21st International Conference Companion on World Wide Web, WWW '12 Companion*, pages 531–532, New York, NY, USA, 2012. ACM.
- David Huffaker. Dimensions of leadership and social influence in online communities. *Human Communication Research*, 36(4):593–617, 2010.
- Jon Hurwitz and Mark Peffley. Public perceptions of race and crime: The role of racial stereotypes. *American journal of political science*, pages 375–401, 1997.
- Seiji Isotani, Akiko Inaba, Mitsuru Ikeda, and Riichiro Mizoguchi. An ontology engineering approach to the realization of theory-driven group formation. *International Journal of Computer-Supported Collaborative Learning*, 4(4):445–478, 2009.
- Jagadeesh Jagarlamudi, Hal Daumé III, and Raghavendra Udupa. Incorporating lexical priors into topic models. In *Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics*, pages 204–213. Association for Computational Linguistics, 2012.
- Isa Jahnke. Knowledge sharing through interactive social technologies: Development of social structures in internet-based systems over time. In *Building the knowledge society on the Internet: Sharing and exchanging knowledge in networked environments*, pages 195–218. IGI Global, 2008.
- Ian Jolliffe. Principal component analysis. In *International encyclopedia of statistical science*, pages 1094–1096. Springer, 2011.
- John Jones. Patterns of revision in online writing a study of wikipedia’s featured articles. *Written Communication*, 25(2):262–289, 2008.
- Ray Jones, Siobhan Sharkey, Janet Smithson, Tamsin Ford, Tobit Emmens, Elaine Hewis, Bryony Sheaves, and Christabel Owens. Using metrics to describe the participative stances of members within discussion forums. *Journal of medical Internet research*, 13(1), 2011a.
- Sara Jones, Anni Poulsen, Neil Maiden, and Konstantinos Zachos. User roles in

- asynchronous distributed collaborative idea generation. In *Proceedings of the 8th ACM Conference on Creativity and Cognition, C&C'11*, pages 349–350, New York, NY, USA, 2011b. ACM. ISBN 978-1-4503-0820-5.
- Armand Joulin, Edouard Grave, Piotr Bojanowski, and Tomas Mikolov. Bag of tricks for efficient text classification. *arXiv preprint arXiv:1607.01759*, 2016.
- Amy Jo Kim. *Community building on the web: Secret strategies for successful online communities*. Addison-Wesley Longman Publishing Co., Inc., 2000.
- Yoon Kim. Convolutional neural networks for sentence classification. *arXiv preprint arXiv:1408.5882*, 2014.
- Aniket Kittur and Robert E. Kraut. Harnessing the wisdom of crowds in wikipedia: Quality through coordination. In *Proceedings of the 2008 ACM Conference on Computer Supported Cooperative Work, CSCW '08*, pages 37–46, New York, NY, USA, 2008. ACM. ISBN 978-1-60558-007-4. doi: 10.1145/1460563.1460572.
- Aniket Kittur and Robert E. Kraut. Beyond wikipedia: Coordination and conflict in online production groups. In *Proceedings of the 2010 ACM Conference on Computer Supported Cooperative Work, CSCW '10*, pages 215–224, New York, NY, USA, 2010. ACM.
- Aniket Kittur, Bongwon Suh, Bryan A. Pendleton, and Ed H. Chi. He says, she says: Conflict and coordination in wikipedia. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '07*, pages 453–462, New York, NY, USA, 2007. ACM.
- Aniket Kittur, Bryan Pendleton, and Robert E. Kraut. Herding the cats: The influence of groups in coordinating peer production. In *Proceedings of the 5th International Symposium on Wikis and Open Collaboration, WikiSym '09*, pages 7:1–7:9, New York, NY, USA, 2009. ACM.
- Alfred Kobsa and Wolfgang Wahlster. *User models in dialog systems*. Springer, 1989.
- Yehuda Koren. Factorization meets the neighborhood: a multifaceted collaborative filtering model. In *Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 426–434. ACM, 2008.
- Yehuda Koren, Robert Bell, and Chris Volinsky. Matrix factorization techniques for recommender systems. *Computer*, (8):30–37, 2009.
- Mikhail Kouliavtsev. Social loafers, free-riders, or diligent isolates: Self-perceptions in

- teamwork. *Atlantic Economic Journal*, 40(4):437–438, 2012.
- Steve WJ Kozlowski and Katherine J Klein. *A multilevel approach to theory and research in organizations: Contextual, temporal, and emergent processes*, pages 3–90. Jossey-Bass., San Francisco., 2000.
- Robert Kraut, Moira Burke, John Riedl, and P Resnick. Dealing with newcomers. *Evidencebased Social Design Mining the Social Sciences to Build Online Communities*, 1: 42, 2010.
- Travis Kriplean, Ivan Beschastnikh, and David W. McDonald. Articulations of wikiwork: Uncovering valued work in wikipedia through barnstars. In *Proceedings of the 2008 ACM Conference on Computer Supported Cooperative Work, CSCW '08*, pages 47–56, New York, NY, USA, 2008. ACM. ISBN 978-1-60558-007-4.
- Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pages 1097–1105, 2012.
- Srijan Kumar, Francesca Spezzano, and VS Subrahmanian. Vews: A wikipedia vandal early warning system. In *Proceedings of the 21th ACM SIGKDD international conference on knowledge discovery and data mining*, pages 607–616. ACM, 2015.
- Srijan Kumar, Justin Cheng, Jure Leskovec, and VS Subrahmanian. An army of me: Sockpuppets in online discussion communities. In *Proceedings of the 26th International Conference on World Wide Web*, pages 857–866. International World Wide Web Conferences Steering Committee, 2017.
- J Richard Landis and Gary G Koch. The measurement of observer agreement for categorical data. *biometrics*, pages 159–174, 1977.
- Jiwei Li, Michel Galley, Chris Brockett, Georgios Spithourakis, Jianfeng Gao, and Bill Dolan. A persona-based neural conversation model. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, volume 1, pages 994–1003, 2016.
- Grace Lin and Marilyn Walker. All the world’s a stage: Learning character models from film. In *Seventh Artificial Intelligence and Interactive Digital Entertainment Conference*, 2011.
- Nedim Lipka and Benno Stein. Identifying featured articles in wikipedia: writing style matters. In *Proceedings of the 19th international conference on World wide web*, pages 1147–1148. ACM, 2010.

- Jun Liu and Sudha Ram. Who does what: Collaboration patterns in the wikipedia and their impact on data quality. In *19th Workshop on Information Technologies and Systems*, pages 175–180, 2009.
- Thomas F Lyons. Role clarity, need for clarity, satisfaction, tension, and withdrawal. 1971.
- Keith Maki, Michael Yoder, Yohan Jo, and Carolyn Rosé. Roles and success in wikipedia talk pages: Identifying latent patterns of behavior. In *Proceedings of the Eighth International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, volume 1, pages 1026–1035, 2017.
- Charles J Margerison and Dick McCann. *Team management: Practical new approaches*. Mercury Books, 1990.
- Aurelien Max and Guillaume Wisniewski. Mining naturally-occurring corrections and paraphrases from wikipedia’s revision history. In Nicoletta Calzolari (Conference Chair), Khalid Choukri, Bente Maegaard, Joseph Mariani, Jan Odijk, Stelios Piperidis, Mike Rosner, and Daniel Tapias, editors, *Proceedings of the Seventh conference on International Language Resources and Evaluation (LREC’10)*, Valletta, Malta, may 2010. European Language Resources Association (ELRA). ISBN 2-9517408-6-7.
- Elijah Mayfield, Miaomiao Wen, Mitch Golant, and Carolyn Penstein Rosé. Discovering habits of effective online support group chatrooms. In *Proceedings of the 17th ACM international conference on Supporting group work*, pages 263–272. ACM, 2012.
- Andrew McCallum, Xuerui Wang, and Andrés Corrada-Emmanuel. Topic and role discovery in social networks with experiments on enron and academic email. *J. Artif. Int. Res.*, 30(1):249–272, October 2007. ISSN 1076-9757.
- Andrew McCallumzy, Kamal Nigamy, Jason Renniey, and Kristie Seymorey. Building domain-specific search engines with machine learning techniques. In *Proceedings of the AAAI Spring Symposium on Intelligent Agents in Cyberspace*, pages 28–39. Citeseer, 1999.
- Geoffrey J McLachlan and Kaye E Basford. *Mixture models: Inference and applications to clustering*, volume 84. Marcel Dekker, 1988.
- George Herbert Mead. *Mind, self and society*, volume 111. Chicago University of Chicago Press., 1934.
- Rishabh Mehrotra, Scott Sanner, Wray Buntine, and Lexing Xie. Improving lda topic models for microblogs via tweet pooling and automatic labeling. In *Proceedings of*

- the 36th international ACM SIGIR conference on Research and development in information retrieval*, pages 889–892. ACM, 2013.
- R Meredith Belbin. Management teams: Why they succeed or fail. *Human Resource Management International Digest*, 19(3), 1981.
- Tomáš Mikolov, Martin Karafiát, Lukáš Burget, Jan Černocký, and Sanjeev Khudanpur. Recurrent neural network based language model. In *Eleventh Annual Conference of the International Speech Communication Association*, 2010.
- Alan Mislove, Bimal Viswanath, Krishna P. Gummadi, and Peter Druschel. You are who you know: Inferring user profiles in online social networks. In *Proceedings of the Third ACM International Conference on Web Search and Data Mining, WSDM '10*, pages 251–260, New York, NY, USA, 2010. ACM. ISBN 978-1-60558-889-6.
- Troy V Mumford, Michael A Campion, and Frederick P Morgeson. Situational judgment in work teams: A team role typology. *Situational judgment tests: Theory, measurement, and application*, pages 319–343, 2006.
- Troy V Mumford, Chad H Van Iddekinge, Frederick P Morgeson, and Michael A Campion. The team role test: Development and validation of a team role knowledge situational judgment test. *Journal of Applied Psychology*, 93(2):250, 2008.
- Rani Nelken and Elif Yamangil. Mining wikipedia’s article revision history for training computational linguistics algorithms. In *Proceedings of the AAAI Workshop on Wikipedia and Artificial Intelligence: An Evolving Synergy*, pages 31–36, 2008.
- George A Neuman, Stephen H Wagner, and Neil D Christiansen. The relationship between work-team personality composition and the job performance of teams. *Group & Organization Management*, 24(1):28–45, 1999.
- David Newman, Jey Han Lau, Karl Grieser, and Timothy Baldwin. Automatic evaluation of topic coherence. In *Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, pages 100–108. Association for Computational Linguistics, 2010.
- Dat Quoc Nguyen, Richard Billingsley, Lan Du, and Mark Johnson. Improving topic models with latent feature word representations. *Transactions of the Association for Computational Linguistics*, 3:299–313, 2015.
- Oded Nov and Sunil Wattal. Social computing privacy concerns: antecedents and effects. In *Proceedings of the SIGCHI conference on human factors in computing systems*, pages 333–336. ACM, 2009.

- Wanda J. Orlikowski. Using technology and constituting structures: A practice lens for studying technology in organizations. *Organization Science*, 11(4):404–428, July 2000.
- Bo Pang, Lillian Lee, et al. Opinion mining and sentiment analysis. *Foundations and Trends® in Information Retrieval*, 2(1–2):1–135, 2008.
- Glenn M Parker. *Team players and teamwork*. Citeseer, 1990.
- David Partington and Hilary Harris. Team role balance and team performance: an empirical study. *Journal of Management Development*, 18(8):694–705, 1999.
- James W Pennebaker, Ryan L Boyd, Kayla Jordan, and Kate Blackburn. The development and psychometric properties of liwc2015. *UT Faculty/Researcher Works*, 2015.
- Bryan Perozzi, Rami Al-Rfou, and Steven Skiena. Deepwalk: Online learning of social representations. In *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 701–710. ACM, 2014.
- Ulrike Pfeil, Panayiotis Zaphiris, and Chee Siang Ang. Cultural differences in collaborative authoring of wikipedia. *Journal of Computer-Mediated Communication*, 12(1):88–113, 2006.
- Martin Potthast, Matthias Hagen, Tim Gollub, Martin Tippmann, Johannes Kiesel, Paolo Rosso, Efstathios Stamatatos, and Benno Stein. Overview of the 5th international competition on plagiarism detection. In *CLEF Conference on Multilingual and Multimodal Information Access Evaluation*, pages 301–331. CELCT, 2013.
- J Preece and B Shneiderman. The reader-to-leader framework: Motivating technology-mediated social participation. *AIS Transactions on Human-Computer Interaction*, 1(1): 13–32, 2009.
- Xiangju Qin, Derek Greene, and Pádraig Cunningham. A latent space analysis of editor lifecycles in wikipedia. *arXiv preprint arXiv:1407.7736*, 2014.
- Marta Recasens, Cristian Danescu-Niculescu-Mizil, and Dan Jurafsky. Linguistic models for analyzing and detecting biased language. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1650–1659, Sofia, Bulgaria, August 2013. Association for Computational Linguistics.
- Yuqing Ren, Robert Kraut, and Sara Kiesler. Applying common identity and bond theory to design of online communities. *Organization studies*, 28(3):377–408, 2007.
- Steffen Rendle. Factorization machines. In *Data Mining (ICDM), 2010 IEEE 10th*

- International Conference on*, pages 995–1000. IEEE, 2010.
- Leonardo FR Ribeiro, Pedro HP Saverese, and Daniel R Figueiredo. struc2vec: Learning node representations from structural identity. In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 385–394. ACM, 2017.
- Catherine M Ridings and David Gefen. Virtual community attraction: Why people hang out online. *Journal of Computer-Mediated Communication*, 10(1):00–00, 2004.
- John R Rizzo, Robert J House, and Sidney I Lirtzman. Role conflict and ambiguity in complex organizations. *Administrative science quarterly*, pages 150–163, 1970.
- Richard Robin. Commentary: Learner-based listening and technological authenticity. *Language learning & technology*, 11(1):109–115, 2007.
- Christian Rudder. *Dataclysm: Love, Sex, Race, and Identity—What Our Online Lives Tell Us about Our Offline Selves*. Crown, 2014.
- Laurie A Rudman and Peter Glick. Prescriptive gender stereotypes and backlash toward agentic women. *Journal of social issues*, 57(4):743–762, 2001.
- Peter Schlattmann. Estimating the number of components in a finite mixture model: the special case of homogeneity. *Computational statistics & data analysis*, 41(3-4):441–451, 2003.
- Suvash Sedhain, Aditya Krishna Menon, Scott Sanner, and Lexing Xie. Autorec: Autoencoders meet collaborative filtering. In *Proceedings of the 24th International Conference on World Wide Web*, pages 111–112. ACM, 2015.
- Joseph Seering, Tony Wang, Jina Yoon, and Geoff Kaufman. Moderator engagement and community development in the age of algorithms. In *New Media & Society*, pages 1–28, 2018.
- Barbara Senior. Team roles and team performance: is there a “really” link? *Journal of occupational and organizational psychology*, 70(3):241–258, 1997.
- Chengcheng Shao, Giovanni Luca Ciampaglia, Onur Varol, Alessandro Flammini, and Filippo Menczer. The spread of fake news by social bots. *arXiv preprint arXiv:1707.07592*, pages 96–104, 2017.
- Manas Somaiya, Christopher Jermaine, and Sanjay Ranka. Mixture models for learning low-dimensional roles in high-dimensional data. In *Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 909–918.

- ACM, 2010.
- Chunfeng Song, Yongzhen Huang, Feng Liu, Zhenyu Wang, and Liang Wang. Deep auto-encoder based clustering. *Intell. Data Anal.*, 18(6S):S65–S76, January 2014. ISSN 1088-467X.
- LP StataCorp et al. Stata data analysis and statistical software. *Special Edition Release*, 10, 2007.
- Greg L Stewart, Ingrid S Fulmer, and Murray R Barrick. An exploration of member roles as a multilevel linking mechanism for individual traits and team outcomes. *Personnel Psychology*, 58(2):343–365, 2005.
- Chenhao Tan and Lillian Lee. A corpus of sentence-level revisions in academic writing: A step towards understanding statement strength in communication. In *Proceedings of ACL (short paper)*, 2014.
- Yee W Teh, Michael I Jordan, Matthew J Beal, and David M Blei. Sharing clusters among related groups: Hierarchical dirichlet processes. In *Advances in neural information processing systems*, pages 1385–1392, 2005.
- W Ten Haaf, H Bikker, and DJ Adriaanse. *Fundamentals of business engineering and management*. Delft University Press Delft, 2002.
- Grigorios Tsoumakas and Ioannis Vlahavas. Random k-labelsets: An ensemble method for multilabel classification. In *European conference on machine learning*, pages 406–417. Springer, 2007.
- Grigorios Tsoumakas, Ioannis Katakis, and Ioannis Vlahavas. Mining multi-label data. In *In Data Mining and Knowledge Discovery Handbook*, pages 667–685, 2010.
- Grigorios Tsoumakas, Eleftherios Spyromitros-Xioufis, Jozef Vilcek, and Ioannis Vlahavas. Mulan: A java library for multi-label learning. *Journal of Machine Learning Research*, 12:2411–2414, 2011.
- Ralph H Turner. Role change. *Annual review of Sociology*, 16(1):87–110, 1990.
- Henny van de Water, Kees Ahaus, and Robert Rozier. Team roles, team balance and performance. *Journal of Management Development*, 27(5):499–512, 2008.
- Fernanda B. Viégas, Martin Wattenberg, and Matthew M. McKeon. The hidden order of wikipedia. In *Proceedings of the 2Nd International Conference on Online Communities and Social Computing, OCSC’07*, pages 445–454, Berlin, Heidelberg, 2007. Springer-Verlag.
- Patrick Wagstrom, Corey Jergensen, and Anita Sarma. Roles in a networked software

- development ecosystem: A case study in github. *Department of Computer Science & Engineering, University of Nebraska-Lincoln, Technical Report*, 2012.
- Yi-Chia Wang, Robert Kraut, and John M. Levine. To stay or leave?: The relationship of emotional and informational support to commitment in online health support groups. In *Proceedings of the ACM 2012 Conference on Computer Supported Cooperative Work, CSCW '12*, pages 833–842, New York, NY, USA, 2012. ACM. ISBN 978-1-4503-1086-4.
- Yi-Chia Wang, Robert E Kraut, and John M Levine. Eliciting and receiving online support: using computer-aided content analysis to examine the dynamics of online social support. *Journal of medical Internet research*, 17(4), 2015.
- Yi Fei Wang and Stephen Petrina. Using learning analytics to understand the design of an intelligent language tutor–chatbot lucy. *Editorial Preface*, 4(11), 2013.
- Morten Warncke-Wang, Dan Cosley, and John Riedl. Tell me more: An actionable quality model for wikipedia. In *Proceedings of the 9th International Symposium on Open Collaboration, WikiSym '13*, pages 8:1–8:10, New York, NY, USA, 2013. ACM. ISBN 978-1-4503-1852-5.
- Stanley Wasserman and Katherine Faust. *Social network analysis: Methods and applications*, volume 8. Cambridge university press, 1994.
- Joseph Weizenbaum. Eliza - a computer program for the study of natural language communication between man and machine. *Communications of the ACM*, 9(1):36–45, 1966.
- Howard T Welser, Eric Gleave, Danyel Fisher, and Marc Smith. Visualizing the signatures of social roles in online discussion groups. *Journal of social structure*, 8(2): 1–32, 2007.
- Howard T. Welser, Dan Cosley, Gueorgi Kossinets, Austin Lin, Fedor Dokshin, Geri Gay, and Marc Smith. Finding social roles in wikipedia. In *Proceedings of the 2011 iConference, iConference '11*, pages 122–129, New York, NY, USA, 2011. ACM.
- Miaomiao Wen and Carolyn Penstein Rosé. Understanding participant behavior trajectories in online health support groups using automatic extraction methods. In *Proceedings of the 17th ACM international conference on Supporting group work*, pages 179–188. ACM, 2012.
- Miaomiao Wen, Diyi Yang, and Carolyn Penstein Rosé. Virtual teams in massive open online courses. In *International Conference on Artificial Intelligence in Education*, pages 820–824. Springer, 2015.

- Junyuan Xie, Ross Girshick, and Ali Farhadi. Unsupervised deep embedding for clustering analysis. In *International conference on machine learning*, pages 478–487, 2016.
- Pengtao Xie, Diyi Yang, and Eric Xing. Incorporating word correlation knowledge into topic modeling. In *Proceedings of the 2015 conference of the north American chapter of the association for computational linguistics: human language technologies*, pages 725–734, 2015.
- Eric P Xing, Michael I Jordan, Stuart J Russell, and Andrew Y Ng. Distance metric learning with application to clustering with side-information. In *Advances in neural information processing systems*, pages 521–528, 2003.
- Elif Yamangil and Rani Nelken. Mining wikipedia revision histories for improving sentence compression. In *Proceedings of the 46th Annual Meeting of the Association for Computational Linguistics on Human Language Technologies: Short Papers*, pages 137–140. Association for Computational Linguistics, 2008.
- Diyi Yang, David Adamson, and Carolyn Penstein Rosé. Question recommendation with constraints for massive open online courses. In *Proceedings of the 8th ACM Conference on Recommender Systems, RecSys '14*, pages 49–56, New York, NY, USA, 2014a. ACM.
- Diyi Yang, Jingbo Shang, and Carolyn Penstein Rosé. Constrained question recommendation in moocs via submodularity. In *Proceedings of the 23rd ACM International Conference on Conference on Information and Knowledge Management*, pages 1987–1990. ACM, 2014b.
- Diyi Yang, Miaomiao Wen, and Carolyn Rose. Weakly supervised role identification in teamwork interactions. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 1671–1680, Beijing, China, July 2015. Association for Computational Linguistics.
- Diyi Yang, Aaron Halfaker, Robert Kraut, and Eduard Hovy. Who did what: Editor role identification in wikipedia. In *Tenth International AAAI Conference on Web and Social Media*, 2016a.
- Diyi Yang, Aaron Halfaker, Robert E Kraut, and Eduard H Hovy. Edit categories and editor role identification in wikipedia. In *LREC*, 2016b.
- Diyi Yang, Aaron Halfaker, Eduard Hovy, and Robert Kraut. Identifying semantic edit intentions from revisions in wikipedia. In *EMNLP*, September 2017a.

- Diyi Yang, Robert Kraut, and John M Levine. Commitment of newcomers and old-timers to online health support communities. In *Proceedings of the 2017 CHI conference on human factors in computing systems*, pages 6363–6375. ACM, 2017b.
- Diyi Yang, Zheng Yao, and Robert E Kraut. Self-disclosure and channel difference in online health support groups. In *ICWSM*, pages 704–707, 2017c.
- Diyi Yang, Shao Junru, Ma Xuezhe, and Eduard H Hovy. An energy network based profile machine for cancer support groups. In *Under Review*, 2019a.
- Diyi Yang, Robert E. Kraut, Tenbroeck Smith, Elijah Mayfield, and Dan Jurafsky. Seekers, providers, welcomers, and storytellers: Modeling social roles in online health communities. *Proceedings of the ACM Conference on Human-Computer Interaction*, 2019b.
- Diyi Yang, Zheng Yao, Joseph Seering, and Robert E. Kraut. The channel matters: Self-disclosure, reciprocity and social support in online cancer support groups. *Proceedings of the ACM Conference on Human-Computer Interaction*, 2019c.
- Mark Yatskar, Bo Pang, Cristian Danescu-Niculescu-Mizil, and Lillian Lee. For the sake of simplicity: Unsupervised extraction of lexical simplifications from wikipedia. In *Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, pages 365–368. Association for Computational Linguistics, 2010.
- Fabio Massimo Zanzotto and Marco Pennacchiotti. Expanding textual entailment corpora from wikipedia using co-training. In *Proceedings of the COLING-Workshop on The People’s Web Meets NLP: Collaboratively Constructed Semantic Resources*, volume 128, 2010.
- Torsten Zesch. Measuring contextual fitness using error contexts extracted from the wikipedia revision history. In *Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics*, pages 529–538, Avignon, France, April 2012. Association for Computational Linguistics.
- Fan Zhang and Diane Litman. Using context to predict the purpose of argumentative writing revisions. In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 1424–1430, San Diego, California, June 2016. Association for Computational Linguistics.
- Jing Zhang, Jie Tang, and Juanzi Li. Expert finding in a social network. In *Advances in Databases: Concepts, Systems and Applications*, pages 1066–1069. Springer, 2007.
- Saizheng Zhang, Emily Dinan, Jack Urbanek, Arthur Szlam, Douwe Kiela, and Jason

- Weston. Personalizing dialogue agents: I have a dog, do you have pets too? *arXiv preprint arXiv:1801.07243*, 2018.
- Shuai Zhang, Lina Yao, and Aixin Sun. Deep learning based recommender system: A survey and new perspectives. *arXiv preprint arXiv:1707.07435*, 2017.
- Xiang Zhang, Junbo Zhao, and Yann LeCun. Character-level convolutional networks for text classification. In *Advances in neural information processing systems*, pages 649–657, 2015.
- Jieyu Zhao, Tianlu Wang, Mark Yatskar, Vicente Ordonez, and Kai-Wei Chang. Men also like shopping: Reducing gender bias amplification using corpus-level constraints. *arXiv preprint arXiv:1707.09457*, 2017.
- Yuchen Zhao, Guan Wang, Philip S. Yu, Shaobo Liu, and Simon Zhang. Inferring social roles and statuses in social networks. In *Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '13*, pages 695–703, New York, NY, USA, 2013. ACM.
- Lei Zheng, Vahid Noroozi, and Philip S Yu. Joint deep modeling of users and items using reviews for recommendation. In *Proceedings of the Tenth ACM International Conference on Web Search and Data Mining*, pages 425–434. ACM, 2017.
- Haiyi Zhu, Robert E. Kraut, Yi-Chia Wang, and Aniket Kittur. Identifying shared leadership in wikipedia. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '11*, pages 3431–3434, New York, NY, USA, 2011. ACM.
- Haiyi Zhu, Robert Kraut, and Aniket Kittur. Effectiveness of shared leadership in online communities. In *Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work*, pages 407–416. ACM, 2012.