ABSTRACT

Contributors to online crowdsourcing systems generally work independently on pieces of the product but in some cases, task interdependencies may require collaboration to develop a final product. These collaborations though take a distinctive form because of the nature of crowdsourced work. Collaboration may be implicit instead of explicit. Individuals engaged in a group conversation may not stay with the group for long, i.e., the group is an "occasional group." Occasional group interactions are often not well supported by systems, as they are not designed for team work. This dissertation examines the characteristics and work of occasional groups in the Gravity Spy citizen science project. Occasional groups in this system form to reach agreement about the description of novel categories of data that volunteers identify in the system.

The author first employed virtual ethnography over six months to investigate volunteers' interactions and to identify features of the occasional groups in this setting. Most groups were transient, interacting only for a short time to develop one product, but a few worked together repeatedly. To describe the overall process of finding new categories brings individuals to work together, the author interviewed nine active volunteers about their work practices. Volunteers individually or collectively use tools such as hashtags, collections and a search tool to identify examples of a new category and to agree on a name and description.

Finally, the author investigated the details of the processes of developing proposals for four new categories over three years. She employed virtual and trace ethnography to collect messages from several discussion threads and boards to identify the analytical moves made by occasional group members in developing a new category. Volunteers would speculate on

a new pattern and its causes, discuss how different categories are related and split or merge descriptions. They employed techniques such as detailed descriptions of data to create common ground, @-mention of other volunteers to increase the visibility of their work to each other and use of the category proposal as a vehicle to coordinate their actions.

Findings contribute to the group literature by recognizing that groups with no formal formation and work processes are capable of doing work that would not otherwise be possible. The results advance our understanding of group categorization literature by showing how the analytical moves are different when group members work occasionally. The thesis also provides some suggestions for better support of occasional groups in crowdsourcing platforms.

Occasional Groups in Crowdsourcing Platforms

by

Mahboobeh Harandi

B.S., University of Science and Culture, 2005
M.S., Norwegian University of Science and Technology, 2015

Dissertation
Submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Information Science and Technology

Syracuse University
July 2021

ACKNOWLEDGEMENT

With immense pleasure and a deep sense of gratitude, I would like to express my sincere thanks to my advisor, Kevin Crowston, for his support and guidance over my Ph.D. journey. Kevin led me to conduct research effectively and taught me how to become an independent researcher. I grew a lot as a researcher and learned how to pursue my goals patiently and consistently. I am also profoundly grateful to Carsten Østerlund, supporting me as a second advisor. Carsten helped me better understand my research's value and improve my work while conducting different studies.

I would like to thank my committee members Bryan Semaan, Joshua Introne, Caroline Haythornthwaite, Dan Cosley, and Michael Kalish. They have provided me invaluable feedback to enhance my thesis and helped me to enjoy the process. I am also deeply thankful to Steve Sawyer, who supported me through his wisdom and kindness over these years.

I would like to thank other faculty members I had the chance to work with over my Ph.D. journey. I learned a lot working with Jenny Stromer-Gally, Jeff Hemsley, Nancy McCracken, and Yatish Hedge in the Center for Computational and Data Science, Daniel Acuña in the Science of Science and Computational Discovery Lab, and Bei Yu.

I am thankful to faculty members who nurtured me through insightful conversations; Elizabeth Liddy, Barbara Kwasnik, Jeffrey Stanton, Ingrid Erickson, Jian Qin, Rachel Ivy Clarke, Jasmina Tacheva, Martha A. Garcia-Murillo, Brian McKernan, Radhika Garg, John Jordan, Arthur Thomas, Michelle L. Kaarst-Brown, Yang Wang and Deborah Nosky.

I would like to thank iSchool and SU staff who supported me over these years. I am so thankful to Jennifer Barclay, Sheila Clifford-Bova, Ellen Hobbs, Susan Neimer, Lois Elmore, Peggy Takach, Blythe Bennett, Sarah Weber, Maureen O'Connor Kicak, Jennifer Pulver, Kevin Shults, Jeffrey Fouts, Christopher Perrello, Kathleen Benjamin, Jose Tavarez, Simona Rosu, Daniel Olson-Bang, Cathy Mantor, Fariba Rahmanzadeh, and Mary Idzior.

This thesis focuses on volunteers who contribute to an online citizen science project, Gravity Spy. I am so thankful to thousands of volunteers whose contributions made my research possible.

I had the privilege to do an internship over the summer of 2020 at Facebook under the supervision of Sanya Attari. I sincerely appreciate her guidance and support. She thoughtfully guided me on how to conduct research in an industry setting. I also had a chance to learn from

other scholars, Charley Scull, Bud Lambert, Jefferey Javed, for whom I am thankful.

I am thankful to Women in Science and Engineering and graduate school programs to provide valuable resources and insightful events to help me grow as a researcher and know my value as a woman of color. Special thanks to Glenn Wright, Sharon W. Alestalo, Stephanie Wyatt, and Amanda Latreille.

I had the chance to meet and work with amazing friends at the iSchool. Corey Jackson, my academic brother, supported me over these years with his intelligence, caring, and kindness. I learned a lot from Jerry Robinson, Natã Barbosa, Bryan Dosono, and Sikana Tanupabrungsun, supporting me through my Ph.D. study and internship experience. I am so thankful for their help and caring.

I am so grateful to my friends at the iSchool, Maxwell, the college of art and science, and Falk college. Specifically, I am thankful to Sarah Brett, Yingya Li, Yaxing Yao, Ehsan Sabaghian, Alain Shema, Qunfang Wu, Jennifer Sonne, Erin Bartolo, Ivan Shamshurin, Jean-Philippe Rancy, Han Zhuang, Alexander Smith, Brian Dobreski, Sarika Sharma, Olga Boichak, Alex Hanson, Jason Markins, Noah Wilson, Brett Keegan, and DeAnna Bay.

I would like to extend my appreciation to my professors in my former school, the Norwegian University of Science and Technology. They encouraged me to follow my dream and pursue my Ph.D. degree, specifically Pinar Øzturk, Audrey van der Meer, and Jon Atle Gulla.

I have been fortunate to meet wonderful people over these years with whom I made hopefully lifelong friendships. I am so thankful to Azam Noori, Reyhaneh Abdolazimi, Rafaela Tillmann Barbosa, Hooman Abootorabi, Faezeh Tabatabie, Sam Azghandi, Shirin Jouzdani, Hajar Faal, Paria Peyravi, Mahdi Nejatbaksh, Reza Hamrah, Atrisa Samadi, Azin Semsar, Nadine Fernande, Marie Williams, Sandy Bargainnier, Haleh Seraji, Shahriar Seraji, Maysam Seraji, Hanna Seraji, Audrey Shinn, Mohammadhossein Mamaghani, Maeve Maloney, Saman Modiri, Amanda Almeida Leavitt, Muge Olgun Baytas, Nadya Rodriguesm, Asal Andarzipour, and Mahnaz Valipour.

I have been blessed to have many incredible friends and I would like to express my gratitude to some of them who supported me in different ways over my Ph.D. study. Special thanks to Mina Khosravifard, Maryam Kashani, Sorour Zerehgar, Sahar Taherpour, Mahtab Javadi, Bahar Toosie, Edna Molhempour, Lilian Betkolia, Minoush Tavasoli, Nazanin Pershang.

Last but not least, I would like to thank my family and relatives. Due to the travel ban, we could not visit each other as I expected before starting my Ph.D. However, they supported

me virtually and encouraged me to do my best as usual. I am deeply thankful to my mother, Giti Moradi, my father, Mohammad Harandi, my brother, Hamidreze Harandi, and my sister-in-law, Saba Mousavi. I would also like to extend my gratitude to my grandmother, uncles, aunts, aunts-in-law, and cousins for their support and positive thoughts over these years.

I dedicate this thesis to my parents for their lifelong support, caring, and kindness.

Syracuse, NY

07/01/2021 **Mahboobeh Harandi**

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

Introduction

1.1 Background and Motivation

Individuals join crowdsourcing platforms such as Wikipedia or Citizen Science projects to contribute to developing a product (i.e., writing a high-quality article or data analysis). Some crowdsourcing projects allow teamwork as the complexity of work is beyond individual contributions and requires teams to create a product. An example of such platforms is FoldIt, a citizen science project that invites teams to fold the best structures of selected proteins, and the research team decides which team product is the best to apply in real-world proteins.

Some other crowdsourcing platforms focus on data processing and support individual tasks to perform data analysis. However, while individuals often work on pieces of a product independently, they may also need to occasionally work and collaborate, either implicitly or explicitly, depending on the project's requirements. Individuals form a group based on task interdependencies rather than sharing specific characteristics or a particular social class of individuals (Lewin 1948) and (Wilder and Simon 1998), forming what the author calls an occasional group. They may join groups of individuals to gain or share required information to fulfill a task or a project.

We can see an example of the occasional group phenomenon in online citizen science projects, crowdsourced science projects, in which scientists invite the general public to participate in a science project. Most citizen science projects ask volunteers to either collect or categorize data along with providing training resources about the science and required tasks. However, a few projects ask volunteers to go beyond the primary classification and identify new categories of data to create knowledge.

A study of this process (Jackson et al. 2018) found that while volunteers in the Gravity Spy project start by identifying new categories independently, they also need to discuss and converge their choices and adopt one name, and consequently create a taxonomy collectively. All volunteers label data at different times to define new categories and need to have a shared understanding of new categories that leads them to create shared terminologies.

On the one hand, each individual is allowed to name a new pattern, mainly using a hashtag that makes sense. On the other hand, they need to make sense collectively as they collaborate to expand their knowledge and organize a massive heterogeneous dataset. Consequently, task interdependencies increase among volunteers and bound them as a group. They engage in various community deliberations such as brainstorming, question and answer conversation, and sharing information to create a new category collectively. However, tedious processes of retrieving relevant information from what they have done individually or collectively in the past and lack of awareness about concurrent relevant discussions hinder the work processes and coordination.

Like any other online community, only some individuals in Gravity Spy join the community deliberation, and some of them have a transactive discussion about a specific task that moves them towards developing a shared product. Transactive discussion or transactive exchange is when individuals continue a discussion based on what has been discussed by other members of a community (Berkowitz and Gibbs 1983). Individuals can refine, extend, or argue against what has been discussed, and, consequently, interdependent work among some individuals increases that would result in developing a shared product. A study by (Wen et al. 2017) explored how community deliberation affects success on team learning.

Occasional groups, however, do not form a team to develop further shared products. The emergence of occasional groups is similar to emergent teams under crisis, fulfilling non-regular tasks studied by (Dynes and Aguirre 1979), but they do not have an identity as a team, and their existence is not officially recognized within a system.

Primary observations of some sorts of group behaviors among individuals and lack of recognition of their group work by the system motivated the author to study their collective activities to elaborate the characteristics of these groups. Also, to understand how they perform catego-

rization as a group and manage current challenges, the author identified their analytical moves developing a new category. This foundational research should help scholars better understand how individuals occasionally work as a group and still achieve goals despite the lack of group identity. Also, it should inform platform organizers for better support and eliminate extra labor for coordination among individuals and increase the visibility of their work.

1.2 Problem Statement

There are different studies on groups and their activities within an organization in various disciplines such as organizational theories (e.g., (Weick 1979)), social psychology (e.g., Allport 1962, Campbell 1975, McGrath 1984) and psychology (e.g., Levi 2017, Hertel and Hüffmeier 2011). Scholars employed different theories from a field that suits their research goals in understanding virtual teams. For instance, theories in social psychology have been adopted to understand how teams work in virtual settings, such as (e.g., Woolley et al. 2010, Powell et al. 2004). Other scholars explored how different aspect of a team such as compositional features and social capital affect their success (e.g., Dascalu et al. 2014, Joyce et al. 2021).

Given different types of teams or groups, scholars also studied the emergence of virtual ad-hoc teams within online organizations (e.g., Zellmer-Bruhn et al. 2004, Janicik and Bartel 2003, Okhuysen and Waller 2002). These teams resemble ad-hoc teams such as film crews and disaster response teams (Bechky 2006). More recently, scholars introduced a new phenomenon, flash organizations, a new form of organization that a system computes how to form such teams and assign roles and tasks to each individual within a team (Valentine et al. 2017).

Different online platforms support teamwork depending on system goals and team requirements. An example of such platforms is an online crowdsourcing system. Online crowdsourcing platforms support individuals work depending on the requirements that a crowdsourcer initiate in the project. Some projects require diverse responses from individuals, and many studies explored how to aggregate and evaluate individuals' responses (e.g., Lin et al. 2012, Dai et al. 2013, Salehi et al. 2015). There are also crowdsourcing platforms that support teamwork and mostly encourage competition among teams to receive the best result (e.g., O'Rourke 2011, Dissanayake et al. 2015).

While online crowdsourcing platforms either promote individuals or teamwork depending on task type and complexity, a system that only supports individual work would not help if individual-based works shift towards teamwork due to the complexity of work. Task interdependencies evolve, and consequently, the need for collaboration rises. The system should provide a different support for various forms of collaboration, such as different groups of individuals who work occasionally in different innovative projects within a system. Overlooking occasional groups and their characteristics limit our understanding of their collaborations and consequently hinders integrating required support into the system to facilitate their work.

This thesis examines the emergence of such groups in a citizen science project where individuals are invited to perform data analysis and categorize items into classes already defined in the system. However, as the task requirement changes, individuals also develop a new class of data. Identifying characteristics of such groups helped to understand how they develop new classes and suggest how the system would better support their work in the future.

1.3 Research Questions

Considering the focus of my research on group processes for occasional groups, The author answers the following questions:

RQ1. What are characteristics of occasional groups in a crowdsourcing platform?

RQ2. How do occasional groups do perform analytical moves required for categorizing a possible new class of data?

To answer the first question, the author conducted virtual ethnography by (Hine 2000) to identify characteristics of occasional groups. Then to answer the second question, she conducted virtual and trace ethnography, and interviews (Geiger and Ribes 2011) to understand how occasional groups perform categorization and how they overcome challenges while the system provides minimum support for collaboration among individuals.

1.4 Thesis Overview

The author situates her research on a citizen science project, Gravity Spy, as an example of crowdsourcing platform to study occasional groups and identify their characteristics while co-

ordinating their actions in pursuing a shared goal. Chapter 2 is the literature review of groups from three disciplines including organizational studies and social psychology to understand which theory is appropriate to use as a theoretical framework to identify characteristics of occasional groups. Also, categorization theory is studied to understand what theoretical lens is suitable to study processes of categorization in occasional groups.

Chapter 3 is the methodology of thesis to explain the research setting, data elicitation and data analysis. The author explains work settings of volunteers and a typical shared outcomes by a group of volunteers. Also, she explains how she collected data and analyzed to answers research questions. Chapter 4 reports results and summary of each study. Chapter 5 discuss the results, and reports conclusion, limitation and future work to better identify and support occasional groups.

CHAPTER 2

Literature Review

2.1 Introduction

This chapter presents the background on group studies in organizational theories and social psychology disciplines. Different disciplines have studied groups and explored their formation and processes to capture what makes a group successful. Each discipline brings a different insight to group studies, and understanding their approaches help to choose an appropriate framework for the current research.

Since the terms group and team entail specific criteria, the first section explains the differences. Then group beginning and processes are studied by elaborating on main models in each discipline to understand which model better explains the processes of occasional groups. As the focus of the current thesis is on occasional groups in crowdsourcing platforms, the author explains the platform and examines studies of virtual teams in such platforms. Then categorization theory is examined to provide a theoretical foundation for investigating how occasional groups categorize considering their characteristics. The discussion section explains why each theory has been chosen to study characteristics and work process of occasional groups. The summary of this chapter is reported in the conclusion.

2.2 Crowds, Groups or Teams

This section first explains how scholars differentiated teams from groups. Then the difference between groups and crowds is explained.

Park and Sanna (1999) defined teams as groups in work settings. Forsyth and Elliott (1999) differentiate teams from groups in terms of behavior and explained that teams are structured

groups who work on defined common goals that need coordinated interactions to accomplish certain tasks.

Teams usually engage in sport or work. They have specific goals and roles of members are defined based on those goals. Members need to have specialization and skills relevant to their roles in a team. Teams are typically in a larger organization. Teams are more exclusive than groups. Groups size range from two to thousands, but teams have a much narrower size, a typical example would be between three to twelve members. Scholars research groups in a laboratory setting and teams in a field to better understand the organizational settings Kerr and Tindale (2004). Also, team members have complementary skills, while group members can do the same task with similar skill sets Levi (2017). Hayes (1997) defines a team considering the power of members. She believes a team must actively cooperate to achieve its goals that require to have independence, responsibility, and power to work. So, a group becomes a team if they have some authority to perform tasks independently and members assist each other in accomplishing a task.

Considering teams in organizational settings doing various types of function, Sundstrom et al. (2000) defined six types of work teams as follows:

- Production teams that manufacture or assemble products on a repeating basis, such as factory teams.
- Service teams that provide repeated transactions with customers such as food services.
- Management teams that plan and develop policy or coordinate activities of an organization. They are managers.
- Project teams that bring experts together to perform specific tasks within a defined period, such as research and engineering teams.
- Action or performing teams perform tasks in a concise time repeated in different settings and requires specialization and comprehensive training and preparation,or instance, groups of individuals work together such as surgery teams and firefighters.

- Parallel teams work outside the routine work and are temporary, such as advisory committees.

Researchers (Mohrman and Mohrman 1993, Sundstrom et al. 1990) have suggested classifying teams whether they are permanent or temporary, how much internal expertise and interdependence they require, how much combination and coordination with other parts of the organization they need to have.

While specific characteristics differentiate teams from groups, groups and teams have often been used interchangeably in the literature (Kozlowski and Ilgen 2006, Hackman and Katz 2010).

Another body of literature focuses on how groups are different from crowds. Henry et al. (1999) identified three main resources that affect group identification, including cognitive(social categorization), affective (interpersonal attraction), and behavioral (interdependence) resources. Henry et al. (1999) explained each of these resources help us to define a group, but they are not equal to group identification. The main difference between social categorization and group identification is how they are perceived conceptually. Social identity is a self-categorization (Flippen et al., 1996 and Tajfel et al., 1971) and dichotomous phenomenon (Turner, 1982), while group identification is defined and measured as continuous value members and outsiders can perceive.

Further, Henry et al. (1999) explained that cohesion is another source for group identification. Cohesion includes group attractiveness and group ability to keep a member in a group, but it is different from group identification as it is only defined as a group-level phenomenon. Group identification is a result of affective bonds among members. However, affective bonds may or may not exists in workgroups. Henry et al. (1999) also elaborated on the difference between behavioral resources and group identification. Behavioral interdependence that includes group processes and shared outcomes are the sources for group identification, but the group is more than its processes and outcomes. These resources are also interacting with each other. For instance, group interdependence plays a crucial role in intragroup attraction.

A more recent study by Leach et al. (2008) suggested a two-dimensional model including for group identification. The model includes two main categories including self-definition and

self-investment. Self-definition is explained by self-stereotyping and in-group homogeneity. Self-investment is explained by satisfaction, solidarity, and centrality- how much a group is central to a member's life. Leach et al. (2008) further explained that self-investment is what members purposefully choose to align themselves with the group while self-definition is what they define based on perceived similarity. However, int their model they did not mention that group identification varies depending on the individuals' attachment to the group. The model by Henry et al. (1999), however, elaborated on how group identification is different from social categorization and its strength varies for members. Similarly, Arrow et al. (2000) explained criteria that would define a "groupy" behavior in a given system. They defined six criteria as follows:

- "If individuals consider themselves as members of the group."
- "If they recognize each other as members and distinguish members from nonmembers."
- "If members feel connected to other members and projects of the group."
- "If members coordinate their behaviors in pursuing collective projects."
- "If members coordinate their use of shared tools, knowledge, and other resources."
- "If members share collective outcomes based on their interdependent activities in the groups."

These criteria are defined based on three cognitive, affective and behavioal resources that explain group identification. first three conditions explore cognition and affection sources of group identification and if members within a group consider themselves and each other as group members. The last three conditions explore behavioral resources of group identification by exploring the interdependencies among individuals; coordinating behaviors and use of shared resources and sharing the positive or negative shared outcome. These criteria are continua rather than dichotomies, and any given set of relations between at least two people fall either weakly or strongly in some or all criteria. Given two main functionality of work groups, completing the group project and satisfying group members' needs, different sets of connections may have various responses to each criterion. For instance, groups of individuals can work together (the

last three criteria) without feeling connected to each other and developing affective bond. However, their shared activities distinguish them from a bigger community. These criteria can be measured from two perspectives; an insider or outsider view. Campbell introduced the concept of entitativity Campbell (1958) as perceiving a group as a single entity by group members or people outside the group. Perceiving a collection of individuals as a group from a member's or an outsider's perspective becomes fuzzier in online settings rather than collocated groups.

To understand how group forms and works, the next two sections presents studies of group beginning and processes.

2.3 Group Beginning

The process of teamwork starts with group beginnings. Tuckman and Jensen (1977) defined five stages of group development as follows:

1. Forming: getting to know each other
2. Storming: disagreement about roles and plans
3. Norming: establishing rules and social relations
4. Performing: focus on accomplishing a task
5. Adjourning: completion of a task and end of the group

Simmel (1950) argues that individuals organize a group around first primitives, egoistic ones, secondary to "finesse and intellectuality." These fundamental issues remain the same even a group undergoes different artifacts and changes. Newcomb (1961) suggests that a group coalesces around the shared attitude. Later in 1963, Scott summarized significant evidence that affection, similar likes or dislikes, has a critical role in shaping a group and sharing cognition. Simmel also argues that emotions are the primary reasons to shape a group, as emotional appeals persuade individuals more than intellectual appeals. While similarity among group members explains their associations, individuals tend to present their dissimilarity to keep their unique identity. Ziller (1964). Ziller (1964) summarized several social psychology studies and

He concluded that individuals within a group are vacillating between uniting and separating from their associations.

From a social psychology perspective, people form a group for a reason, and group members have a shared goal. Members either have a relationship or a connection to one another that helps them share what happens to their fellows. Communication is a central process of a group. They recognize their membership and have rules and norms that control their interactions. They influence each other, and the desire to stay in a group develops reciprocal influences. Eventually, a group satisfies members' physical and psychological needs. The interdependency among members is the most important feature of a group Johnson and Johnson (1987).

Group beginning or formation relies on the assembly of people and resources and their interactions within each other and with its embedding context. These interactions make a group a distinguishable component in its embedding context; both internal and external forces affect the group's assembly and formation. Whether there are planned programs for forming a group or emergent processes, group formation will be different. Arrow et al. (2000) defined four categories based on external and internal forces, planned assembly, and emergent processes (see Figure 2.1. They defined four types of group assemblies as follows:

- Concocted groups (external, planned forces predominate)
- Founded groups (internal, planned forces predominate)
- Self-organized groups (internal, emergent forces predominate)
- Circumstantial groups (external, emergent forces predominate)

The model is also applicable to online settings. The section virtual groups explains how Gómez-Zará et al. (2020) developed a new taxonomy for team assembly in online settings. They specifically focused on teams while used the term groups and teams interchangeably for their analysis. The rest of this chapter uses these two words interchangeably.

2.4 Group Processes

Scholars in different disciplines such as organizational studies and social psychology have examined group process and its impact on group sucess. In the following sections, three main

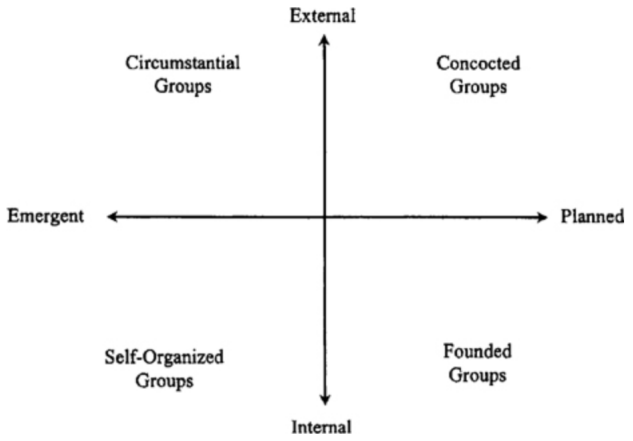

Fig. 2.1 Forces in the Group Formation Space by Arrow et al. (2000)

studies that inspired the study of groups in organizations and online platforms are explained.

2.4.1 Groups and Organizing by Weick

Weick employed revolutionary theory to explain how groups work through interlocked behaviors and how they maintain their collaboration. He started explaining how groups start working despite having a dilemma at the beginning of group formation. While it might be assumed that individuals should have a shared goal to work in a group, Weick (1979) believed that at the beginning of forming a group, they do not have tangible actions that rely on determining the goal. Despite the dilemma, they form a group and create interlocked behaviors. Scholars conceptualize collective actions or interlocked behaviors by studying collective structures. Individuals are exposed to instant evolution of events that make it difficult for them to predict and perceive the world. So, if they want to recognize what is happening and how they should react or initiate any behaviors, they need to know when an event starts and terminates or understand cycles. Defining a cycle in a group is more difficult, but it happens when a person acts in a certain way, and their action is valuable to another person. The person who has benefited from action does something in a reciprocal manner that satisfies the first person. The mutual benefit that their actions have made for another one is shaping a collective structure Allport (1962). Allport explains the give-and-take acknowledgment is another element of shaping collective structure. He mentioned that every group member trusts another person's interlocked behaviors but not the person and all their behaviors. Since every individual may have several group memberships

that lead them to invest their behaviors among different groups, only their interlocked behaviors are reliable. Allport (1962) mentioned individuals within a group provides structural assurance, which has two pieces of evidence; 1) the structure is reliable and keeps operating and 2) every individual maintains their place within this structure.

Scholars can examine the processes of organizing within a group by understanding the interlocked behaviors. Sociocultural evolution theories can help us understand the process of organizing with no focus on the result of organizing. Weick specified three processes: variations, selection, and retention according to evolutionary theory adopted to social behavior by Campbell (1975). As he mentioned in his book, the most apparent feature of sociocultural is the variation among individuals. Having several heterogeneous individuals in a group makes the group wiser than each individual, and consequently, the variations prepare the context for selection. The selection is an action of choosing someone as being the most suitable match for an organization. The Selection is one of the complex constructs that makes it difficult to answer a simple question, such as how we know if something is selected; is it selected because it is there, or is it there because it is selected? Campbell defines selective systems and selective criteria separately. There are six selective systems, according to him:

- "Selective survival of complete social organization": A big part of an organization or a group can be omitted because they did not have effective collective actions.
- "Selective Diffusion among groups": The groups who are less wealthy borrow characteristic of groups behavior who are wealthier.
- "Selective perpetuation of temporal variations": A group will repeat behaviors that are interconnected with pleasant memories.
- "Selective imitation of inter-individual variations": The individual within a group emulates other group members' behavior.
- "Selective promotion to leadership roles": The group chooses individuals who have various actions that are adaptive for all other members and promote them as leaders.

- "Rational Selection": If societies plan, predict, and anticipate, which makes a rational selection.

In addition to these six selection systems, countless selection criteria exists, such as accepting what is pleasant and bringing pleasure, and rejecting what is not. However, the social system should consider these endless criteria regarding its internal and external functionalities.

A group removes equivocality of the actions by creating interlocked behaviors. Hollander and Willis (1967) proposed three principal elements in a cycle between two people, including acting, interacting, and double interacting. A person acts; the second person either accepts or rejects the action, then the first person responds to what the second person did. The person has a choice point in the second and third stages of the cycle; the second person can either accept or reject the action, and the first person can either abandon, revise or maintain in response to the second person interaction. Each process of organizing is a set of assembly rules and interlocked behavior cycles. Rules are limiting selection of interlocked behavioral cycles. There are different rules or criteria such as effort, frequency, success, performance, duration, availability, personnel, relevance, reward, and disturbance. For instance, the group will select cycles that could remove ambiguities the most.

The final evolutionary process is retention that is repository of selected actions and affects future actions. Concurrently, if the actions evolve a lot, some of the selected actions can be contradictory to those stored in the repository. Barnard (2004) and Thompson (1967) suggested the social contract is why workers stay at work, such as satisfaction, productivity, and interpersonal ties. Other scholars like Marks et al. (2001) having a cyclical perspective suggested a recurring phase model of teamwork that teams perform a temporal cycle of activities that creates a rhythm for the team. The developmental perspective shows how teams go through different phases of activities (i.e., low activity vs. high activity).

2.4.2 Groups and Organizing by McGrath

Studying group process through evolutionary theory helps to understand the process within a group, however, it does not take into account how the system, a bigger context, plays a role in this process. McGrath developed a model considering the organizational setting.

The model by McGrath (1984) defines three ways of organizing people into work groups, including a work group, a team, or a self-managing team. Work groups are part of an organizational hierarchy. And managers or supervisors decide that members can independently accomplish a task based on their decision. Managers select team leaders who employ various techniques (i.e., constative mode) to include team members in decision-making processes. Team members work independently, and the leader coordinates their activities. Self-managed teams have significantly more power and authority than other types of teams. They are also more independent, and members select the leader, and the leader has a facilitator role than a managerial role. The leader employs democratic or consensus decision-making. Team members are highly interdependent, and they work together to coordinate activities. He proposed a conceptual framework for the group elements that impact the group dynamics as follows:

- Characteristics of individuals such as behavior and belief
- Relationship among individuals as a group structure, for instance, if they like each other or not, or how long they have known each other
- Features of the environment, whether it is at a workplace, sports competitions or a family vacation
- Task features that have a critical role in conceptualizing a group as it defines the interactions of individuals within the group and
- Behavior settings of a group is the pattern emerged from the group, a structured entity, and the task, a structured set of requirements and demands.

In his conceptual framework, the study of groups starts with the study of specific individuals' characteristics and the study of the physical and social aspects of the environment. As members become interrelated and the environment's features define a particular task, these input elements' changes are parallel to the primary input element before their change. McGrath describes the conceptual framework as a metatheory. Later in 1991, He proposed a theory of groups and how they do what they do. The theory focused on three dimensions, including time, interaction, and performance McGrath (1984). He emphasized that scholars need to consider

Table 2.1 Small Group Theory by McGrath (1991)

	Production	**Group well-being**	**Member support**
Inception	Production demand /opportunity	Interaction demand /opportunity	Inclusion demand /opportunity
Problem-solving	Technical problem solving	Role network definition	Policy conflict resolution
Conflict resolution	Policy conflict resolution	Power and payoff distribution	Contribution and payoff relationships
Execution	Performance	Interaction	Participation

the social, physical, and temporal context of the group besides the duration of membership affect their relationships. He added that group members perform a complicated series of interdependent tasks that help them reach the goal. McGrath theorized three proposition about the nature of groups as follows:

- Groups are supposed to be complex social units with many interdependent functionalities on multiple tasks while being moderately linked within and roughly linked to the surrounding environment. According to this proposition, functionalities are in three levels, including the members, the group, and the organization, respectively have three products: member support, well-being, and production. As they take specific actions to fulfill multiple tasks, they require coordinating their (sub)tasks with other group members. They have multiple memberships that imply partial membership for each group, and their group memberships are roughly linked to a larger social unit.

- There are four modes of group actions when members want to work on a project, including the inception and approval of a project, choosing the right solution, policies in resolving the conflicts, and execution. In other words, group actions encompass choosing the goal, means, policy, and goal accomplishment. He emphasized that not all group activities include means and policy choice, but they have the goal choice and the accomplishment. The relations of functions and modes presented in Table 2.1.

  As McGrath interprets, scholars need to go through each functionality and study the action modes. Production is an organizational function that goes through the goal choice, production demand/opportunity, the means choice, technical problem solving, policy

choice, policy conflict resolution, and execution, performance. The choices and execution modes are defined for the group's level, well-being, and individual, member support.

- The four modes of activities do not have a fixed order. They are a set of substitutes and alternative activates that a group and its members may employ. However, the default linear path is from the first mode inception and initiation of the opportunities for production, interaction, or participation to the last mode, execution of the production, interaction, or participation.

McGrath emphasized that groups going through different paths depending on the group needs does not mean they do not have efficiency. His perspective is opposed to scholars in the 1970's like Steiner (1972). McGrath explained that there are multiple time-activity paths that groups choose to do what they do. He proposed four propositions for the second element of group theory, time, as follows:

1. Behaviors in workgroups present many complicated temporal patterns. Workflows temporal aspects include scheduling, organization, time allocation. There are difficulties in matching the appropriate amount of time for a particular activity. There are entertainment processes for coordinating group members' behavior and group behaviors with external events.

2. Organizations and individuals struggle with three common and basic temporal problems, including temporal ambiguity, conflicting temporal interests and requirements, and insufficiency of temporal resources. Organizations respond respectively by scheduling works, coordinating works through different parts of the organization, and time allocation for each project. Individuals react respectively by making temporal obligations, discuss norms for behavior sequencing, and adopting and regulating the task flow and interpersonal interactions. He noted the mismatch between organizational and individuals' responses can cause residual temporal problems such as deadlines, coordination, and regulating the interaction flow. These problems weaken the group.

3. A time-based effective workflow in groups obliges a complicated matching of work collections with specific periods. Some periods are more useful for a particular activity than

other times. For instance, some activities can be done faster overnight and can slower over the weekend. The same concept applies to the activities. Some activities can be done efficiently at any time and are more flexible than others.

4. Social entertainment is one of the main temporal patterns. It is about the temporal coordination of two or more processes. Social coordination is the coordination of behavioral processes and not the psychological ones. As mentioned earlier, it can be between individuals, groups, and the surrounding environment that can be internal or external considering each of these social units.

The pattern of entrainment or coordination varies for each social unit, their task, and social context. McGrath noted that entrainment, functionalities, and modes are interdependent. As groups choose a different path to accomplish a task, they may or may not improve the task's quality depending on the available temporal resource.

The last element of McGrath group theory is the interaction. He proposed four proposition for interaction as follows:

1. Group interactions are about the workflow within a group at a micro-level, the act of an individual. The act has three dimensions; type of the act and its relation to the ongoing group work, source and destination of the act that specifies the act's duration, and the time of the act that determines when an individual should accomplish it.

2. Any point of interaction is either ”purpose” or ”objective” of a group that is its principal task. The focal or principle task of a group is what they are working on and accomplishing. These tasks are related to the production, well-being, or member support functions that can include ”socio-emotional” features excluding ”instrumental” substances.

3. Each act may or may not be relevant to the group's current focal task. Different tasks are relevant to the ongoing group's focal task, such as proposing solutions to a task problem, proposing a new plan, suggesting a different set of activities and strategies, and evaluating whether positive or negative. However, examples of irrelevant acts to the ongoing group focal task are irrelevant anecdotes about project content or dissatisfaction with environmental situations.

4. Acts are situated: they have meanings regarding the functions, modes, and activity path of the group. They can be either related or unrelated to the functions and modes, and activity path. However, in either case, they are situated and meaningful depending on the context. That implies act aggregations are vital to analyze the group interactions. Any form of act aggregations reflects different aspects of the workflow within a group. Group's workflow is understood by analyzing the flow of interactions for specific events, acts, the aggregation of acts, members, and periods of time. These independent variables, act types, members, and periods of times can be calculated differently, such as distributions of frequencies of activity types, comparing duration for different activity types, and the sequence analysis of activities. All the calculations over a specific activity provide different and useful information about group work's temporal pattern.

2.4.3 Groups and Organizing by Ilgen

McGrath elaborated on the details of team processes within a system and how their activities change over time. His model was adopted in many studies and was a basis for developing another outstanding model by Ilgen et al. (2005).

While McGrath (1984) proposed a model of input-process-output, Hackman (1987) suggested that a group is evaluating their performance while they are working, and that helps them change the process of their work to have a higher performance. Later in a study of teams in organizations, Ilgen et al. (2005) developed an input-mediators-output-input model. This model considers the social aspects of a team and includes the feedback loop that Hackman had suggested before. Besides, it separates emergent states from processes. The principles of this model are input, mediators, and outputs. Team members' characteristics, task characteristics are the input. Mediators are aspects that facilitate inputs' effect on outputs, and outputs are the tasks and non-task results of a team's functions. Mediators are into two classes, including processes and emergent states. Processes are dynamic interactions among team members while working on a task, and emergent states are dynamic properties of the team that is different in each given context. These properties are cognitive, motivational, and affective states based on their definition. They defined the loop between output and input to emphasize how feedback

from output to input stage affects the team's future regarding processes and emergent states. Consequently, according to a group's functionality, there are different active input, members' and tasks' characteristics, and process, dynamic interactions of members. Scholars in different disciplines adopted Ilgen's model to understand the nature of organizing in different settings.

Ilgen's model captures the dynamics of groups in many ways. It reflects while they are working and interacting with the environment, they practice maintaining as a group through the feedback loop. The model can be adapted for many studies to capture group processes and, if they manage to evolve to stay as a group in a given environment.Another approach derived from TIP theory and added complex system theory was by Arrow et al. (2000). They considered small groups as "a complex, adaptive, dynamic, coordinated, and bounded set of patterned relations among members, tasks, and tools." They studied local and global dynamics of a group within a system.

2.5 Factors Contributing to Group Success

Group formation and process are critical parts of group studies. However, it matters to understand what contribute to group success to better understand group processes and what affect their success.

The group composition, task characteristics, group process, and organizational context affect the success of a group. A team must have members with the required expertise and relevant knowledge who can work through interpersonal skills and if there is a need they must represent the relevant parts of an organization. The team performs different tasks that require various skills, and we may analyze a task by observing and checking a member's input relates to the product. For instance, if a team can accomplish a task regardless of all indviduals' contributions, the most mediocre performance of an individual will not affect the team performance. And if the team should make a quality decision, the team outcome will be better than individuals. A good team task is what motivates the team and requires a coordinated work to fulfill the task. The group process is what connects the team members to its tasks. A successful and capable team organizes itself to accomplish tasks, develops supportive social relations, and selects leaders who help and facilitate the teamwork and progress. The last element, the or-

ganization, gives the context to the team. First, the organization supports teams by creating a culture that fosters collaboration and enables the team to control its internal operations. Then it provides directions, resources, information, and assistance to support the teamwork. Finally, it gives feedback on their performance and rewards successful teams. Scholars from diverse disciplines defined successful teams mostly through interviewing successful teams. However, as there are various types of teams and interview questions depending on scholars' knowledge, we have different definitions of successful teams as well.

Among organizational psychologists, Hackman (1987) with the specialization of job design, defined five factors for team success based on a study of teams at work and in the laboratory as follows:

1. Clear direction and goal that helps them to focus their efforts and measure their performance

2. Good leadership that manages internal and external team relations and facilitates team orientation toward their goals

3. Complicated and challenging tasks that require coordinated actions by team members with a variety of expertise

4. Resources such as material, training, and personnel supports

5. Allocating power and authority to teams for decision making and implementing plans

While the model Hackman (1987) has been adopted in developing more comprehensive models and study of online teams, the groups success has been also studied from other disciplines. From management perspective, Katzenbach and Smith (1993) studied upper-level management teams in large organizations. They noticed clear goals, conventional approaches and methods for fulfilling a task, and mutual accountability are essential factors for the team success. They saw a team has the best performance if it has a small number of members who have complementary expertise and are committed to reaching a shared goal.

From the group communication perspective, Larson et al. (1989) studied teams from business, sports, and government. They mentioned similar factors such as clear goals and standards

of excellence, ethical leadership, external support and appreciation for the team success. Moreover, they found that a result-oriented structure, qualified team members, a unified commitment, and a collaborative environment are essential.

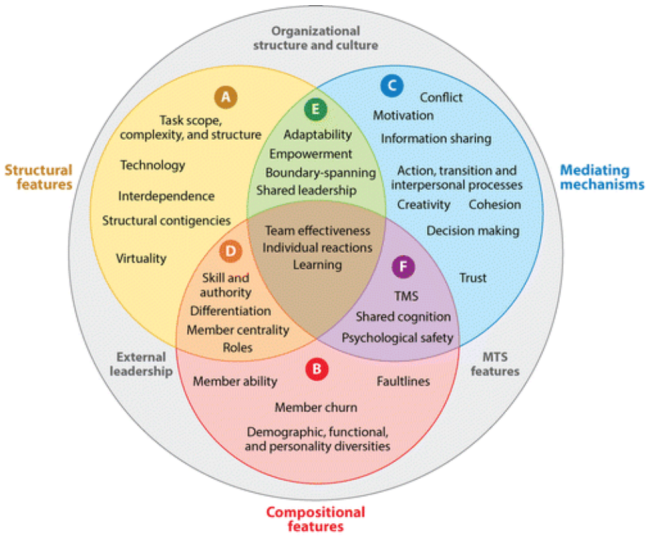

Fig. 2.2 The model describes how team inputs, mediating mechanisms, and structural features in addition to overlapping co-evolving facets affect team success by Mathieu et al. (2019)

A more recent study by Mathieu et al. (2019) also adopted the model by Hackman (1987) and Ilgen et al. (2005) to create a comprehensive model for team/group effectiveness based on their earlier studies Mathieu et al. (2017), (See Figure 2.2). They defined three main categories, including structural features, compositional features, and mediating mechanisms that consider factors beyond team performance and success, including the consequences of team performance on its members and the future work. They also focused on the intersection between each pair and considered how each intersection affects team success, including structural features and compositional features, structural features and mediating mechanisms, compositional features and mediating mechanisms. Structural features include structural contingencies, task scope and complexity, interdependence and team virtuality. Compositional features include average member attributes, diversity, and faultlines. Mathieu et al. (2017) discussed three types of diversity, including surface-level, deep-level and functional diversity. Surface-level diversity consists of demographic characteristics of team members. Deep-level diversity focuses on psychological characteristics such as personality factors and attitudes. Functional diversity focuses on functional areas and background varies among team members. And faultlines are hypothet-

ical splitting lines that divides groups into subgroups. Mediating mechanisms includes team processes, information sharing, emergent states, and conflict. Compositional and Structural features include skills and authority differentiation, member centrality, roles, and team size. Structural and Mediating features include adaptability, team empowerment, boundary spanning, and shared leadership. Compositional and Mechanisms features include psychological safety, shared cognition/mental model, and transactive memory systems.

As mentioned above, the model by Mathieu et al. (2019) is comprehensive and includes all principal and overlapping factors, it borrowed its principles from the Hackman model. The benefit of the Hackman model is its simplicity that would give a clear understanding of what makes a group successful.

2.6 Virtual Groups

Group studies in traditional settings provide a foundation for studying teams in online or virtual settings. Studies of teams, team formation and processes, in online setting are inspired mostly by McGrath and Ilgen model (e.g., Gilson et al. 2015, Wiggins and Crowston 2010).

A recent study by Gómez-Zará et al. (2020) developed a model insipred by Arrow et al. (2000) to explain how online teams or groups form in online platforms. The model by Arrow et al. (2000) explained how internal/external forces and planned/emergent forces affect group formation and the model by Gómez-Zará et al. (2020) explained how systems and user agency play role in team formation.

Gómez-Zará et al. (2020) identified two dimensions for group formation; the authority of individuals in forming a group and the number of people who participate in forming a group. Gómez-Zará et al. (2020) explained that depending on how a socio-technical system is designed, individuals have different levels of the agency to form a team. For instance, in online communities where individuals can choose to work with any individuals on a project, they have full agency versus the system that assigns individuals to a team based on the components through its algorithm. In some online systems, the algorithm creates groups, and individuals do not have the authority to form a group. The second dimension of group formation is the number of participants who contribute to forming a group. For instance, in some socio-technical sys-

tems, a few individuals can form a team, regardless of their membership in the group, versus in some systems, any number of individuals can form a team. Gómez-Zará et al. (2020) defined four categories for group assembly (see Figure 2.3) as follows:

- Self-assembled teams: a system enable users to self-organize in their own teams.
- Staffed teams: a user customizes the team-assembly criteria used by the system to simulate and form teams.
- Optimized teams: a system assembles teams given particular team-formation criteria.
- Augmented teams: a system augments users' actions by suggesting potential teammates.

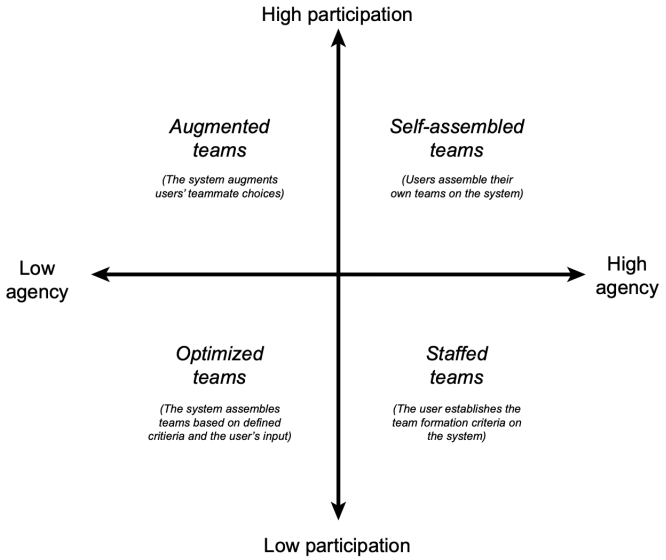

Fig. 2.3 The Taxonomy of Team-Assembly Systems by Gómez-Zará et al. (2020)

Scholars like Warkentin and his colleagues Warkentin et al. (1997) studied performance and relational links in virtual teams according to the TIP theory developed by McGrath (1991). Piccoli and his colleagues Piccoli et al. (2004) explained that the research for virtual teams focused on the input, socio-emotional and task processes, and outputs based on the IPO (input, processing, output) model by McGrath (1984). They identified early studies of virtual teams in each of these areas as follows:

- Input: structure, culture, technology, and training.

- Socio-emotional processes: cohesiveness, trust, relationship building.
- Task processes: communication, coordination task-technology-structure fit.
- Output: performance, satisfaction.

Research studies (e.g., Ilgen, Hollenbeck, Johnson, and Jundt, 2005; Marks, Mathieu, Zaccaro, 2001) also highlighted the importance of team processes on increasing the quality of output and collaboration. Team processes, including communication (i.e., information exchange) and coordination, are critical issues in virtual teams. Scholars (McGrath, 1991; Warkentin, Sayeed, Hightower, 1997) identified that the significant challenge for online teams is coordinating the temporal patterns of group behavior. Scholars in organizational behavior studies (Olson 2000, Cramton 2001, Huckman et al. 2009, Hinds et al. 2011) explained geographic distribution, technology-mediated communication, and fluid membership hinder coordination and team success. Lack of shared understanding of the whole task, roles, and responsibilities limits team coordination. Virtual teams deal with many obstacles to reach a shared understanding as they do not have a rich communication as teams in traditional settings Cramton (2001).

Gilson et al. (2015) conducted an inductive analysis of studies around virtual teams based on the IPO (input, processing, output) model by McGrath (1984). They identified ten themes including, research design, team inputs, team viturality, technology, globalization, leadership, mediators and moderators, trust, outcomes, and ways to enhance virtual teams' success.

Virtual teams have been studies in different settings such as online organizations and crowdsourcing platforms. The next section focuses on formation of virtual teams in crowdsourcing platforms as the focus of thesis is crowdsourcing platforms.

2.6.1 Virtual Teams in Crowdsourcing Platforms

Crowdsourcing platforms support collective action on a variety of tasks, such as problem-solving and innovations (i.e., Open Source Systems), data processing (i.e., data collection, data analysis), knowledge aggregation and knowledge creation (i.e., writing an article) that require either simple or expert skill sets Geiger et al. (2012).

Individuals, institutions, non-profit organizations or companies can initiate a crowdsourcing

project and invite the public or a specific group of people to perform paid or unpaid tasks Estellés-Arolas and González-Ladrón-De-Guevara (2012). Initiators may require homogeneous contributions that consider all valid responses equally or heterogeneous contributions that consider responses depending on the individual's expertise. Initiators may require an emergent response that considers all responses and relationships among them or non-emergent response that considers isolated responses Geiger et al. (2012)(Figure 2.4).

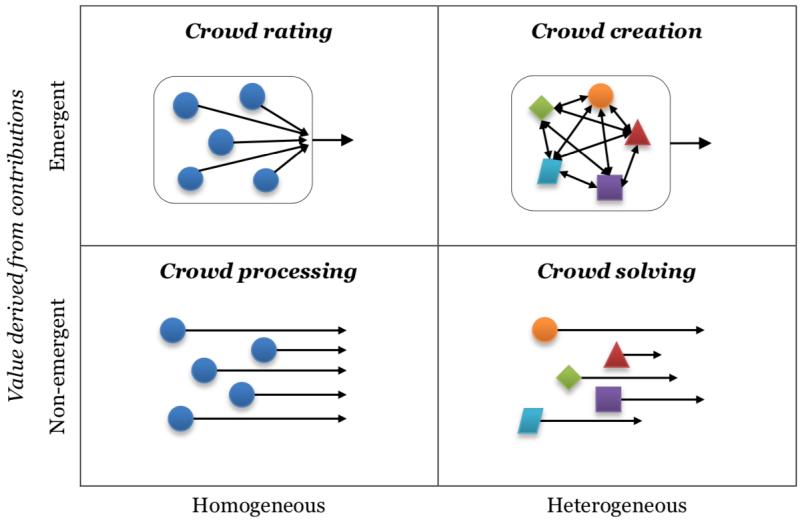

Fig. 2.4 Crowdsourcing types by Geiger et al. (2012)

Citizen science project tasks in particular include crowd processing, collaborative problem-solving contributions Huang et al. (2018) or creating knowledge Rughiniş (2016). Processing projects such as the majority of citizen science projects are examples that seek homogeneous and non-emergent responses. Online opinion systems and collective rating are examples of crowdsourcing projects that rely on heterogeneous and non-emergent responses. Problem-solving projects such as Netflix Prize Bennett et al. (2007) and Kaggle benefit from heterogeneous and non-emergent responses. And projects that require innovation and creation such as user-generated content (YouTube) and knowledge aggregation systems (i.e., Wikipedia) require heterogeneous and emergent responses. There have been several studies designing workflows to help workers to increase the quality of their work and designing efficient voting algorithms to allocate the accurate score to each response (e.g., Lin et al. 2012, Dai et al. 2013, Salehi et al. 2015).

The design of crowdsourcing platforms and voting algorithms vary depending on what kinds of responses are required, emergent, or non-emergent values. Crowd creating or rating that requires emergent responses bring more interdependencies and collaborations among individuals compared to crowd processing and problem-solving. Haythornthwaite (2009) described two extremes of crowd engagements, Lightweight Peer Production (LWPP) vs. Heavyweight Peer Production (HWPP), that each requires non-emergent and emergent responses respectively.

Individuals in Lightweight Peer Production (LWPP) work independently as their contributions are not dependent on others. Crowd initiators define a singular unit for the task, and the system considers each response independent on other individuals' responses. It usually does not require registration in the system and allows anonymous contributions. Consequently, interactions among individuals are minimum, having a weak ties model of interaction. Consequently, interactions among individuals are minimum, having a weak ties model of interaction.

In contrast, individuals in Heavyweight Peer Production collaborate to accomplish an interdependent task, and interdependency is reciprocal. They revise responses, negotiate, and address equivocality to reach a consensus and deliver the outcome. Collaborators have registered accounts and usually have specific roles in the project. This kind of system usually has fewer contributors who do the majority of work in the system and require group support to accomplish the task.

While crowdsourcing platforms mostly design work for individuals, there are examples of platforms that enable team work, ad hoc teams, according to the project goals. However, these teams are mostly temporary and have a short cycle. Virtual ad hoc teams in crowdsourcing platforms encounter more obstacles in coordinating their activities and perform a task as a team. They emerge depending on specific circumstances, and consequently, the level of shared understanding, clarity of roles, and responsibilities are more uncertain than virtual teams. Despite all uncertainties, they should be able to reach a shared understanding while exchanging information. Scholars identified the value of the ability to coordinate temporarily (Ancona Chong, 1999; Janicik Bartel, 2003; Okhuysen Waller, 2002; Zellmer-Bruhn, Waller, Ancona, 2004) and diversity of exchanged information (Bell, 2007; Mello Rentsch, 2015) on

team success in online emergent collaborations. Virtual ad-hoc teams emerge and coordinate temporarily through discussing and sharing information. Discussions are mostly asynchronous on different topics. Individuals join teams according to their regularity, bringing more temporal coordination patterns (Kittur et al., 2013). Consequently, there might be a burst in a load of discussion and long pauses in between Riedl and Woolley (2017). In a study by Riedl and Woolley (2017), they found out the temporal "burstiness" of team activity and the diversity of information exchanged among team members are substantial predictors of performance despite cash incentives and membres expertise. In another study by Kou and Gui (2014), they found out there are productive social interactions within temporary teams, and they create disciplines to facilitate collaborations.

While the notion of light-weight and heavy-weight peer production presents collaboration types in crowdsourcing platforms and some studies specifically focused on emergence of teams in crowdsourcing platforms, to the best knowledge of author, there is no study exploring if a new concept emerge in this continuum.

2.7 Categorization

Since the work in many crowdsourcing platforms is data analysis and categorization requiring group work, this section explains how individuals and teams categorize items. Categorization is a primary step in knowledge creation. Human beings create categories to manage loads of information required to live in a complex information-loaded world. Categorization is a process in which individuals put similar things in one category, spatial, temporal, or spatio-temporal segmentation, that would help them to understand their world (Bowker and Star 2000). Individuals form categories as a social construct based on the existing knowledge; they form categories having a specific goal (Berger and Luckmann, 1967). Bowker and Star (2000) explained how classification and standardization are two sides of the same coin; classification may become standardized, and every successful standard forces a classification system. Three principles define different classification systems, including a consistent rule that defines a class, mutual exclusivity of categories, and completeness of a classification system that would describe its embedding system. However, not all working classification systems would cover these three

principles. There are visible and invisible classes depending on constraints imposed by social, economic, and political reasons.

Similar to the group theory, classification theory has been studied in different disciplines such as psychology, social science, organizational theories, and management. Findings in psychology studies have been borrowed in other fields. Psychology studies focus on human cognitive, neuropsychological, and neuroimaging processes to understand how they categorize objects Ashby and Maddox (2005). Psychology scholars aim to understand the micro-foundations of categorization, considering properties of categories and the cognitive processes humans take to perceive categories. Examples of psychology studies are as follows:

- Rosch (1978) explained that basic-level categories exist in abstraction.
- Brooks et al. (1978) investigated the internal properties of categories' constructs.
- Cohen and Murphy (1984) and Hampton (1988) explored how individuals perceive items that combine components from various categories.
- Rosch (1999) proposed organizing categories into hierarchies makes each category group either less or more inclusive.
- Johnson and Mervis (1997) explained how individuals use different schemas to categories the same objects.
- Spalding and Murphy (1996) studied the different bases for forming categories.
- Ashby and Maddox (2005) explained individuals learn perceptual categories differently depending on how categories are constructed; various qualitatively distinct systems mediate human category learning.
- Newell et al. (2011) explained psychological phenomena are explained by varieties of explanation and multiple-systems often provides an illusory sense of scientific progress.

Social science studies focus on how people use categories and share the meaning of categories. Social science and social psychology studies (e.g., DiMaggio 1991, Douglas 1986, Duquenne et al. 1998) centered around meaning embedded in category systems to understand

role conformity, social sanctioning, collective identity dynamics, and evaluation of categories by an external audience. For instance, Researchers in economic sociology Pachucki and Breiger (2010) studied how categories separate markets and consequently provide the required economic exchange infrastructure.

Management studies have drawn from both psychology and sociology literature to understand organizational phenomena. Porac and his colleagues (Porac and Thomas, 1990; Porac et al., 1989, 1995) relied on cognitive psychology studies to understand how strategists perceive the external environment and their firm's position within the environment. In the late 90's, Zuckerman (1999), through the lens of social science studies on categorization, theorized categories as features of the external environment and studied how categories create disciplines. Zuckerman, in different studies, explored how categories form the allocation of attention among organizations (Zuckerman, 1999; Zuckerman and Kim, 2003). Scholars (Hsu and Hannan, 2005; Phillips and Zuckerman, 2001) studied how categorization theory facilitates comparing firms and products. Management studies kept applying this approach in further studies; later in 2014, Vergne and Wry (2014) defined a lexicon of categories through reviewing the literature and emphasized that the management and organizational studies would benefit integrating more with studies in cognitive psychology. They defined the lexicon as follows:

Classification Hierarchy: a cognitive description of structural relationships between categories that gained consensus among category members. The structure has various levels, and the hierarchy defines how these levels are nested and related to each other.

Category Boundaries: absolute boundaries define exactly what falls inside and what does not, but the relative boundaries help to differentiate categories.

Category Membership: if audiences and other members of a category believe an organization's offerings are within the boundaries of a category and the organization is focused enough, they consider the organization a category member.

Partial Membership: if offerings hold some category attributes and the organization is not focused enough to be recognized as a full member, it is partial membership.

External Identity: the material and symbolic support and resources that help audiences to evaluate category membership is external identity.

Category Straddling: if an organization has multiple concurrent memberships in more than one category in the classification hierarchy, it is spanning categories.

Category Legitimacy vs. Category Legitimation: scholars inspired by institutional theory (Suchman, 1995) defined legitimacy as compatibility with broad social norms, whereas scholars having an ecological perspective (Hannan et al., 2007) defined legitimation as compliance of feature values with schemata. As a result, legitimation increases as the audience reach a consensus about the meaning of a label.

Category Prototype: a prototype is the most typical/representative or central member of a category based on an audience's perspective. However, there are discussions about whether the category prototype should be the average member of a category or the most salient one (Jones et al., 2012).

Category Properties: category attributes, fuzziness, contrasts, leniency, saliency, similarity, and stigmatization from the perspectives of category members and the audiences.

People categorize known and unknown data as a primary step to create knowledge and make sense of their world simultaneously. Mainly, if a group of people categorizes new data, they collectively make sense of what they create to share within a group and with outsiders. However, collective sensemaking requires to have a shared identity; as Weick et al. (2005) mentioned, identity is central in the process of collective sensemaking. After establishing an identity, they would enact their surroundings in a narrative (Currie and Brown, 2003) and plausible stories to save and share with other people (Maitlis, 2005). Since sensemaking is an ongoing process, favorite stories and narratives change over time as individuals project their stories to their world and receive feedback from their audience. Individuals extract cues from their world to understand what is relevant and acceptable to develop their stories. Collective sensemaking would help interpret categories developed within a group, but it does not provide a framework to identify analytical steps to create a categorization system.

Grodal et al. (2020) applied categorization theory to understand the analytical actions of qualitative researchers improving existing theories or developing a new theory. Grodal et al. (2020) identified eight analytical moves that qualitative researchers apply to analyze data. They emphasized that qualitative researchers may apply more than one move in analyzing their data

in non linear order. For instance, while a scholar is considering the relation between categories in the stabilizing phase they may still go back to the refining stage and merge categories. However, specific moves are dominant in each stage of the process. Figure 2.5 shows the theoretical framework developed by Grodal et al. (2020); darker shades shows which analytical moves are dominant in each analytical stage. Since scholars may have repeated previous analytical steps while moving forward, the figure includes the earlier steps in the later stages but with a lighter color; earlier steps are not dominant actions in the later stages. Analytical moves are asking questions, focusing on puzzles, dropping categories, merging categories, splitting categories, relating and/or contrasting, sequencing categories, and developing or dropping working hypotheses. Grodal et al. (2020) explained researchers create initial categories either by asking a question or focus on parts of data that do not conform their existing knowledge.

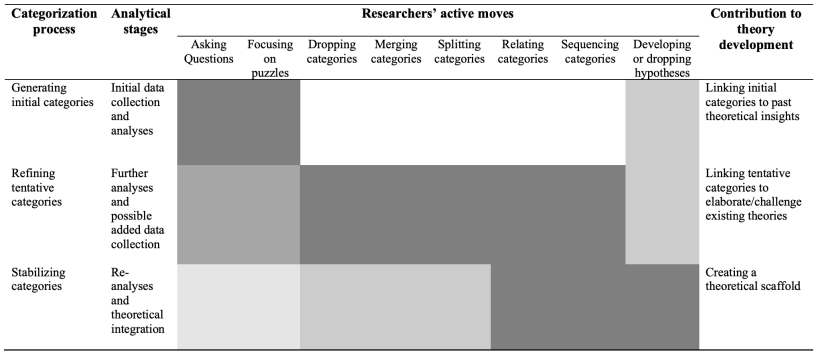

Fig. 2.5 An Active Categorization Framework for Theory Development by Grodal et al. (2020) - Darker shades shows the importance of each step in the analytical stage.

Asking questions: researchers may approach data having specific questions. As mentioned earlier, human classify data based on existing knowledge and may categorize the same object differently depending on the goal of classification (Bowker and Star, 2000). Researchers do the same; they rely on existing theory and have specific questions while collecting data and analyzing. However, they may modify their questions along the process of analysis.

Focusing on puzzles: researchers may initialize analysis by focusing on parts of data that are surprising. As (Bowker and Star, 2000) explained, not all pieces of information are equally important for categorization. Humans categorize using accumulated knowledge about a particular topic (Durkheim and Mauss, 2009). Consequently, individuals may not recognize a piece of information perfectly aligned with their knowledge (Clark and Wilkes-Gibbs, 1986, Colyvas

and Powell, 2006, Vygotsky, 1987). Instead, when observation does not conform with the accumulated knowledge, individuals focus on that piece to expand their knowledge. (Ahn, 1999) explained when observations are different from the existing knowledge, salience arises; the salience has a key role in category formation as it draws attention to the novelty of a specific situation. Similarly, researchers by focusing on odd findings expand their analyses to examine possibility of theory development. Becker (2008) defined this move as "over-focusing on strange elements". He explained focusing on findings that do not fit the existing theory shows the necessity of theory development.

After initializing primary categories, researchers improve the initial categories through three moves, including dropping categories, merging categories, and splitting categories.

Dropping categories: as Murphy (2004) and Rosch (1978) mentioned, humans mostly face an overwhelming amount of information for the purpose of categorization that would make them initially form categories that are wrong, biased, and/or irrelevant. Also, as humans focus on salient categories, the stable patterns become less critical (Weick et al., 2005). Consequently, dropping initial categories to focus on more meaningful categories is part of the process that would better explain the phenomenon of study (Murphy, 2004).

Merging categories: according to categorization theory, people initially form detailed categories to better elaborate a new phenomenon (Bloom, 2000). Consequently, they create categories that are detailed and could be merged with similar ones. So, they merge similar categories to create a super-ordinate one that covers all instances with the same properties. Murphy(2004) explained the process of merging categories results in complex hierarchies of categories.

Splitting categories: according to categorization theory, people initially form detailed categories to elaborate on a new phenomenon (Bloom, 2000). Consequently, they create categories that are detailed and could be merged with similar ones. So, they merge similar categories to create a super-ordinate one that covers all instances with the same properties. Murphy(2004) explained the process of merging categories results in complex hierarchies of categories. Researchers gain a more nuanced understanding of different ways a category is formed in their data.

Researchers may improve the primary categories by either sequencing them temporarily or by relating and/or contrasting.

Relating and/or contrasting categories: scholars studying categorization suggested that categories are typically interrelated semantically (Collins and Loftus, 1975, Quillian, 1969). Categories become related if they are in the same semantic network and are distantly related in they never occur together. Moreover, categories may be perceived as contradictory when each opposes the other (Douglas, 1966, Lévi-Strauss, 1969). As a result, the way individuals relate or oppose categories affects the meaning of categories. For instance, people from the meaning of "natural" in comparing it with its opposite meaning, "artificial" (Weber et al., 2008). Researchers relate or contract categories to specify the relation or lack of relations among them.

Sequencing categories: an important part of the categorization is creating a dynamic understanding of categories related to the world (Durand and Paolella, 2013). A fundamental part of the categorization process is to create sequential relationships among categories (Ahn, 1999, Murphy and Medin, 1985). The sequence determines if all actions within a category conform with the definition of a category. Qualitative researchers identify a sequence of mechanisms, objects, persons, and concepts that would create a theory.

In the last stage of categorization, researchers aim to create a theoretical framework to explain the studied phenomenon through revising and re-analyzing categories and combining specified mechanisms and thoughts. Consequently, researchers conclude if data support the existing theory or develop a new one.

Developing or dropping working hypotheses: researchers (Durand and Paolella, 2013, Murphy and Medin, 1985) explained the way humans form categories is not dissociated from the more comprehensive theories that create an integral part of a category. So, if an instance does not conform to mechanisms and definitions of a category, it implies that it does not belong to the category. Developing or dropping categories is an iterative process that humans experience to either expand or drop current assumptions they employ to perceive the world. Researchers, similarly, form an overarching theory that would explain a phenomenon, and as they iterate the process of data categorization, they may find instances that would either develop or drop a theoretical perspective.

It is expected to have a similar analytical moves if a team of scholars conducts a qualitative analysis to improve a current theory or develop a new theory. A relevant study to the team categorization process is studies of the team classification. Specific characteristics of classes set them apart from categories; classes within a classification system are mutually exclusive, jointly exhaustive, and have relationships and notation (Clarke 2021). However, a classification system still includes a categorizing process while new classes are developed. Similarly, the categorization process aims to develop relations between categories. Due to partial similarity in the process, the author examined the literature to understand how teams develop classes or categories.

Scholars like Young and Mandelstam (2013) showed public institutions support team classifications for the sake of transparency and accountability, rather than individual classification systems. Similarly, in online settings such as Wikipedia, many collaborations emerge for creating categories to tag Wikipedia articles (Thornton and McDonald 2012). An important aspect of team classification is whether the classification designer supervises the team of workers developing classes or is a member of the team to work with other experts to build a classification system. In a study by Soergel (1976), the classification designer assigned two main tasks to his staff and subject expert. Staff employed a series of detailed flow charts for each step of the process to enter terminologies into the system, and the subject expert was responsible for identifying semantic relationships between terms. The classification designer extracted information from the subject expert to provide a satiable source for terminologies. The classification designer was responsible for defining the broad subject fields, specifying the system structure, and coordinating the teamwork.

While many systems support hierarchical classification as the classification designer has the authority to dictate the process of classification, some systems support a flat structure of teams where classification designers participate in the process of classification, and multiple views are considered developing the system. A study by Albrechtsen and Jacob (1998) showed a case study of a classification work at Ballerup Public Library where a team of librarians, reference and cataloging experts, invited a group of people to collaborate on each step of the classification process. The classification designer translated ideas of groups to an organizing system, and the

community recognized their role as a communicator for the library environment. The classification designer had the authority at the final stage of decision making rather than defining the working flowchart for team members from the beginning.

An example of a flat organization is seen in online social tagging systems that share some similarities with classification systems. Both classification and social tagging systems help to organize and locate information. However, social tagging systems apply a loosely defined method, folksonomy (Sturtz 2004). Folksonomy, however, has some structure compared to open tagging. While an authoritative organization creates a classification system using predefined rules and structures, in social tagging systems, a community of individuals creates a tagging system, usually using an uncontrolled vocabulary (Trant 2009). People tag objects to either collect similar items within an online community or increase the findability of an item for themselves and others. Individuals' efforts may end up a collective action if people use the same tag for the same item; they create a taxonomy resulting from the community work called folksonomy. Depending on the number of people using the same tag for the same item, folksonomy might be broad or narrow (Trant 2009). In collaborative projects such as social bookmarking, a group of people tag items known as a collaborative effort. However, as Trant (2009) mentioned in most cases in online communities, it is social tagging rather than collaborative tagging. Social tagging and collaborative tagging have been used interchangeably in some literature.

Social tagging systems allow the whole community gets involved in the process of developing folksonomy. However, as there are not controlled vocabularies as a guideline for tagging, the community would need to apply different techniques such as tag gardening to organize tags and identify what is best to choose as a tag within a community (e.g., Peters and Weller 2008, Jackson et al. 2018). Additionally, folksonomy may not be an accurate representation of an item or an entity. Scholars like Kroski (2005) consider folksonomy as an inclusive way to capture the wisdom of crowds, but other scholars like Kroski (2005) argues relativism of folksonomy makes its accuracy flawed compared to the accuracy of a classification result.

Despite contradictory evaluations of folksonomy, tagging has several functionalities that would serve a community, such as identifying what an item is, identifying characteristics of an

item, refining categories, and task organizing (Golder and Huberman 2006). However, tagging is not equal to categorizing despite its usage for task organizing. Tags are similar to a book index, whereas categories are like a book chapter. A category covers a broader concept that describes a set of items, but tags describe details of an item within a set or an item that still does not belong to a category. While categories are mutually exclusive in some levels, tags might be used on different items that belong to different categories. However, in a categorization system, individuals tag items to develop categories; descriptive tags help distinguish which items share specific characteristics to create a category. While tags and categories are different, in some studies such as Thornton and McDonald (2012) the difference between tagging and categorizing in Wikipedia is not well defined. Thornton and McDonald (2012) showed what are the main concerns of Wikipedia editors while creating a category system, including the hierarchy of categories, the proper scope for each category, and how they could be used to navigate different articles.

Communities like citizen science apply social tagging, while the ultimate goal is a categorization requiring the science team's approval. The science team holds power compared to the rest of the community due to their comprehensive knowledge about data. As a result, the authority of the science team affects a team's outcome in developing a new category. Consequently, citizen science projects bring an exciting opportunity for a study where social tagging and categorization happen in one setting with different purposes and authoritative levels.

2.8 Discussion

Scholars have studied teams in different disciplines and settings to understand the nature of groups and processes of their work (e.g., Weick 1979, McGrath 1984, Ilgen et al. 2005, Arrow et al. 2000). The model developed by McGrath (1991) and Ilgen et al. (2005) inspired several studies to understand the team formation and process of their work in online settings. Scholars examined how team formation and work processes change in online settings Gómez-Zará et al. (2020). The processes vary as the user agency in forming a team depends on what a system allows. For instance, virtual ad hoc teams emerging depending on task requirements usually have a shorter collaboration period and they should be able to coordinate their actions despite

challenges such as lack of history of teamwork.

While virtual ad hoc teams form depending on task requirements as a flash organization Valentine et al. (2017) in crowdsouring platforms, there is a assemblage of individuals emerge depending on task interdependencies but do not officially form a team. A group of individuals come together to accomplish a project with no particular role description and official organization. While individuals often work on pieces of the product independently in the system, they may also want to occasionally work and collaborate, either implicitly or explicitly, forming what the author calls an occasional group.

To understand the characteristics of an occasional group in a crowdsourcing platform, the author employs the definition of groups by Arrow et al. (2000) to investigate to what extent they resemble group characteristics. While most literature definitions have definite boundaries to define a group, the criteria defined by Arrow et al. (2000) are continua and cover fuzzy boundaries. The model includes cognitive, affective and behavioral resources that help to identify groups. For instance, it covers groups with loose bonds between members coming together for a short period to accomplish a task such as a group laboratory who work for an hour; also, it covers groups with strong social bonds such as multigenerational family households. This model separates social identity from group identity. As Henry et al. (1999) explained despite social identity that is a dichotomous phenomena, group identity varies and has different strengths. Another aspect of criteria is the number of people within a group; any assembly of individuals including two or more people are considered a group. While in some literature teams should have a minimum of three, every two individuals can form a group. Arrow et al. (2000) also explains another side of the spectrum that focuses on the distinction between groups and the bigger collective; to what extend group members perceive themselves belonging to a group that is a part of a bigger collective. The model by Arrow et al. (2000) includes criteria that help to understand if a collection of people can be considered as a group and if the group is distinguishable from the community.

The study of characteristics of occasional groups is the first step to recognize their existence. Characteristics of occasional groups inform and facilitate recognizing their work processes in a given system. Like past studies of teams and ad-hoc teams, applying the input-mediators-output

model would bring a high-level understanding of how these groups work. But, it may not help to understand how they succeed despite not being recognized in a given system and having resources for a successful performance. While teams have clear boundaries indicating who are members and their roles, occasional groups do not necessarily conform to the same set of rules. Teams in crowdsourcing platforms have a specific identity as a team and may assign different roles to coordinate their actions; occasional groups do not have a formal identity as teams. As a result, coordination of actions within an occasional group becomes more critical as they should know with whom for how long they should coordinate their actions to accomplish a particular task analyzing data. As Hackman (1978) explained five factors help team success within an organization, including clear direction and goal, good leadership for managing internal and external relations, enough complicated task that require coordination, training and personnel resources, organizational support for team authority. However, groups that emerge occasionally in a crowdsourcing platform accomplish shared products despite lack of these resources and support.

To comprehend the work process by occasional groups, the author focuses on categorization processes through the lens of a model developed by Grodal et al. (2020) based on the categorization theory. Individuals analyze data to develop new categorizing data through three main stages: generating initial categories, improving tentative categories, and stabilizing categories (Grodal et al. 2020). However, the model yet is not explaining how teams would categorize data. While the categorization model is developed for individual categorization, the author believes the model still helps her to identify analytical moves within a group as the principles of categorization is the same for individuals and groups. Comparing folksonomy, collaborative tagging and categorization, the categorization model should better inform group processes as group categorization goes beyond tagging; categories are what can be identified by tags and categorization also considers the relations between categories.

A study by Bullard (2017) specified that including a domain expert in the classification process helps members better coordinate their actions and avoid having irrelevant or repetitive tags in a classification system. Teams categorizing data have specific roles and guidelines to develop a new category. Different informational resources, specific roles, and expertise

would support team coordination and decision-making processes. As occasional groups are not recognized within a system, coordination and decision-making should raise more problems compared to teams that have a specific identity within a system. Visibility of works among team members affects the process of their work, and the effectiveness of their collaboration (Carroll et al. 2006, Christensen 2013). Teams employ different techniques to increase the visibility of members' work to improve awareness, such as calling another member's attention to share what they are doing. As a result, they can coordinate their next move better. However, occasional groups should face an extra challenge as the beginning of their group work is not clear in a given system. These groups emerge depending on task requirements in a system and still do not officially form a team having members listed early in the project. Another aspect of work that may hinder the work process among occasional groups is allocating required resources to facilitate the work processes. Crowdsourcing organizations provide different resources to support team processes and success, but when the entity of an emerging group is unknown to the organization, their requirements for success remain unclear to the organization.

Recognizing occasional groups would help to better provide resources to facilitate their work. Knowing how they already accomplish a creative task would bring insights into the support processes within a framework. This thesis explores the nature of occasional group work in crowdsourcing platforms and investigates in high-level and detailed analysis how they go through different steps to analyze data and develop a shared creative product.

2.9 Conclusion

In this chapter, the group definition, formation, and processing are explored to provide a theoretical framework for examining occasional groups. As the author explores occasional groups in crowdsourcing platforms, she also explained the nature and different types of crowdsourcing platforms and studies of virtual teams in such platforms.

Categorization theory is also studied from different disciplines to understand which one is appropriate for the study of data analysis by occasional groups. In the end, the author discussed why she chose each framework to identify characteristics of occasional groups and their work practices in crowdsourcing platforms.

CHAPTER 3

Methodology

3.1 Introduction

This dissertation employs a multilevel study to understand the nature of occasional groups, how they develop new categories analyzing scientific data at individual and group levels. The author first explains the research setting and a high-level definition of work that volunteers perform in the project. The second section explains the data elicitation for each study, including virtual ethnography, interviews, and trace ethnography. The last section focuses on data analysis for each study and how the author analyzed data.

3.2 Research Setting

3.2.1 Citizen Science Project: Gravity Spy

Gravity Spy (www.gravityspy.org), hosted on the Zooniverse platform aims to improve the instruments used to search for gravitational waves in the Laser Interferometer Gravitational-Wave Observatory (LIGO) scientific collaboration (Zevin et al. 2017). Due to the high sensitivity of the detectors needed to detect gravitational waves, they also record the noise (”glitch”) that hinders detecting the gravitational waves. So, understanding glitches types and removing their sources is a crucial activity. There are three sites that record gravitational waves and glitches: LIGO Hanford (in Washington State), LIGO Livingston (in Louisiana), and Virgo (near Pisa, Italy).

While some crowdsourcing platforms require one type of participation, some projects require a mix of involvement, such as crowd processing and creation. Projects with a mixture

of requirements have more varieties of commitments that fall between Lightweight Peer Production and Heavyweight Peer Production. Consequently, task interdependencies, information needs, individuals' interactions require different infrastructural support compared to projects that support one specific type of crowd's participation. For instance, the process of creating a new class of data in a citizen science project brings more task interdependencies compared to the standard process of data classification. That requires more interactions among participants to negotiate and reach a consensus (Jackson et al. 2018).

Volunteers join Gravity Spy to classify glitches and identifying possible new glitch classes to help scientists to remove the glitch data from the dataset. Volunteers communicate with each other and the science team through discussion boards. They specifically can discuss features of each image through the classification interface linked to the talk page. Volunteers will face challenges not only to find new classes but also classify images that their boundaries are vague and require more expertise to specify the exact class(es).

3.2.2 Two Main Activities

There are two main activities that volunteers do to organize the massive dataset of glitches including classifying the glitches and identifying possible new glitch classes.

Classifying glitches into official classes: volunteers classify images through the classification interface. A glitch image, converted from omega scan to a readable format for humans and machines, is presented on the left side of the interface. Each image shows the intensity of glitches through its frequency (y-axis) over time (x-axis). There are four slides of each image with different times slots. Depending on each level of classification, volunteers see various numbers of classes. There are different tools to guide volunteers to choose the right class. They can click on each category and read what specific features belong to each class. They can also use three filters to see different classes based on the duration of the glitch (long or short), frequency of the glitch (high, mid, low), and where it is evolving or not. Figure 3.1 shows level 4 having fifteen classes of glitches. The glitch belongs to the "chirp" as it is a spiraling compact objects sweeping upwards in frequency over time. There are twenty nine glitch classes in the system.

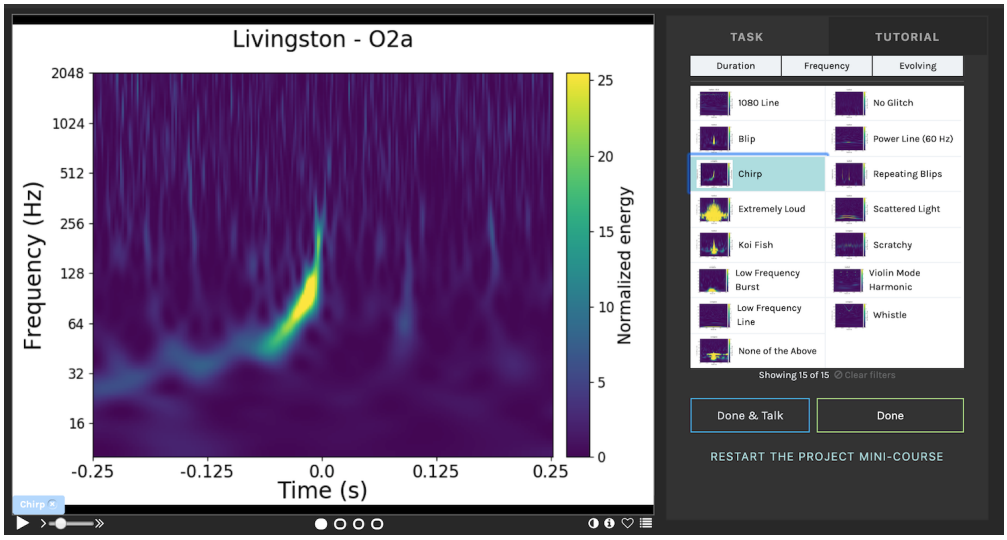

Fig. 3.1 The Gravity Spy classification interface.

Classifying possible new glitch classes: In addition to images, volunteers classify to the known classes, they also need to identify possible new glitch classes. This task in Gravity Spy compared to other Zooniverse projects, is almost unique. As in other projects, volunteers use classes that have been already identified by the science team, and they do not need to discover a novel class of data. Identifying possible new glitch classes on the talk page brings more task interdependencies among volunteers as there are not any structured tutorials or field guides to teach volunteers how to classify possible new glitch classes. Gravity Spy, like other Zooniverse projects, supports individual work and some collaborative work. For instance, a volunteer can invite others to collaborate on their collection, or they can discuss different topics on the talk page. However, for the process of finding possible new glitch classes, there is no specific support. Over this process, different groups of volunteers emerge due to the task interdependencies to work together occasionally to find out what label reflects the best what a glitch type is. Figure 3.2 shows one thread of discussion among three volunteers discussing a possible new glitch class, ”MicroHF”.

There are multiple boards that volunteers and the science team can discuss the project and relevant issues. The Science board serve as a space that volunteers can discuss the science of gravitational wave research and other interesting relevant research papers. The Help board is a space that volunteers ask questions about the interface or the project. The Collections board is

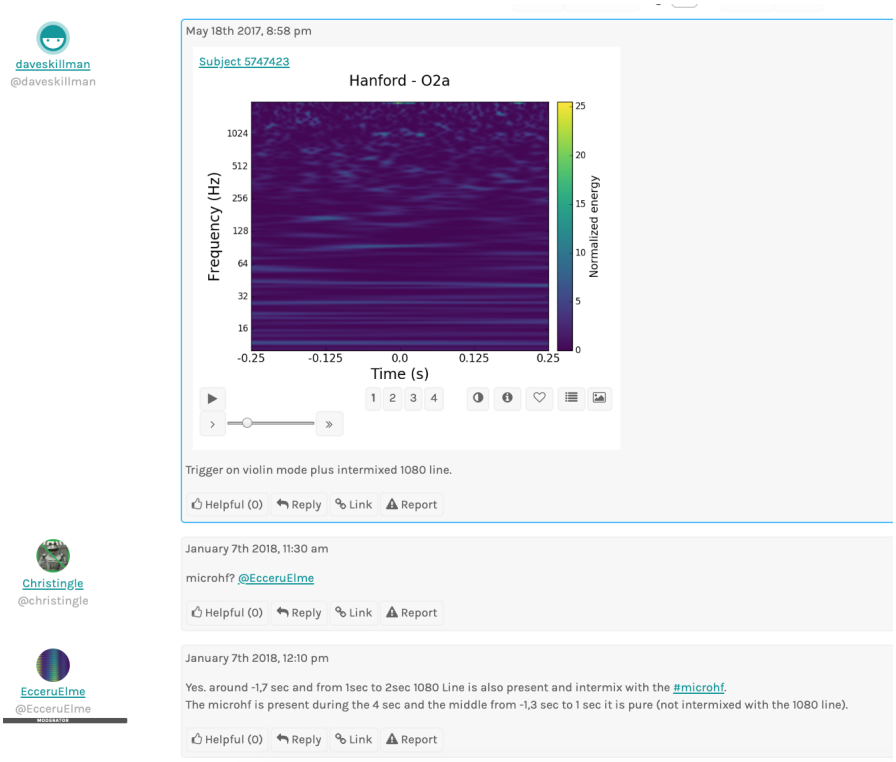

Fig. 3.2 A discussion thread around a possible new gltich class, MicroHF

where volunteers discuss the collections they develop and search for collection collaborators. The chat board is where some volunteers introduce themselves when they join Gravity Spy and can talk about any topics related to the Gravity Spy or anything else they would like to talk about. And the Notes board is a spaces that volunteers comment about each image that initiates a discussion with other volunteers interested in the images.

In addition to the boards mentioned above, there are two boards for submitting new class proposals, including LIGO New Glitch Class and Virgo New Glitch Class. Virgo is another interferometer to detect gravitational waves in Italy. As LIGO and Virgo decided to share and jointly analyze data, Gravity Spy also added glitch images of Virgo a couple of years ago to the system. Moderators can submit a new glitch proposal to the system, and the science team will review the proposal to decide if they should add a new type to the system. The moderator should submit a proposal including a prototype image, glitch name, the date of the proposal, a short description suitable for the field guide, characteristics of hashtags, a list of all hashtags used to describe an image, a collection of hundred images, whether they used the similarity search tool, and collaborate with other volunteers. Figure 3.3 shows a proposal named campfire. While

moderators are the only one who can open a new discussion thread on the New Glitch Class board, all volunteers can leave comments on each proposal and discuss the new type.

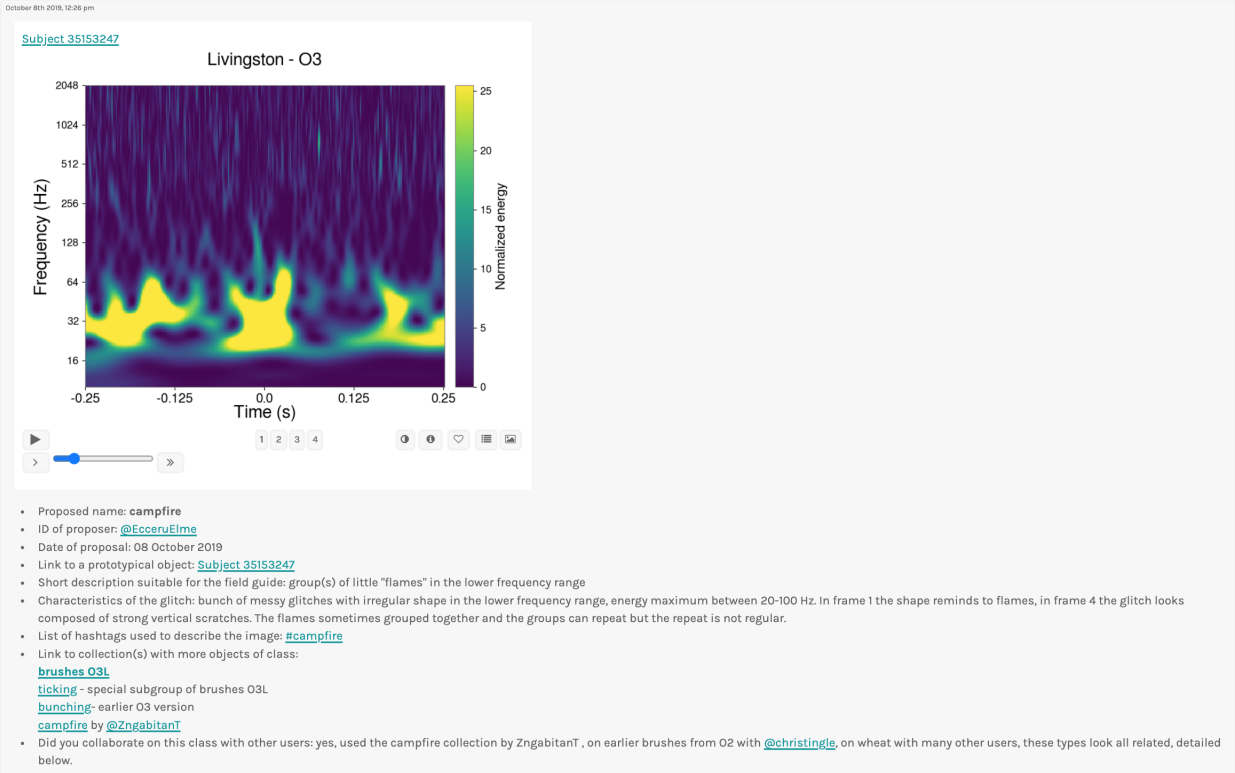

Fig. 3.3 A typical new glitch proposal

3.3 Data Elicitation

The author employed virtual ethnography (Hine 2000), semi-structured interviews, trace ethnography (e.g., Geiger and Ribes 2011, Østerlund et al. 2020), to understand group behaviors in Gravity Spy as an example of a crowdsourcing project. She conducted a multi-level study to specify characteristics of groups who work occasionally, how they find new classes of data, and how they analyze data to develop a new category.

Virtual ethnography allows scholars to be part of an online community and, depending on the research setting, observe, document, and interview with individuals within the community. The author created an account before the launch date and observed volunteers' discussions around various topics. As a virtual ethnographer, she observed and took notes rather than engaging in the talk page. The author employed the model by Arrow et al. (2000) as a theoreti-

cal framework to lead her observations and collect threads of conversation that would manifest ”groupy” behavior. She went through the talk page to observe volunteers' behavior by studying their comments on different images and topics. She also did a few hundred classifications and reached level four when conducted the study. As a result of classifications and observing different discussion threads, the author could take notes of comments showing group behavior and recognize challenges groups of volunteers would face in the process of labeling possible new glitch classes. She specifically observed and analyzed what group behavior look like among volunteers over six months.

After conducting the first study, the author noticed volunteers have more in-depth collaboration and group behaviors around the advanced work; she conducted virtual and trace ethnography and semi-formal interviews to better understand the nature of advanced work and collaborations among volunteers.

The author conducted virtual and trace ethnography for four proposals submitted in 2017, 2018, and 2019 to understand how occasional groups develop a shared outcome. The author went through discussion threads of proposals submitted from 2017 to 2019 and randomly selected five proposals with several, few, and no comments left by different volunteers. Proposals submitted under the names of Spirograph and Cord were discussed by several volunteers on the proposal page and consequently provided rich data to analyze ”groupy” behavior. Selecting the proposals with fewer comments, including Crown and Fireball or no comments, including tealight, would decrease the bias for identifying ”groupy” behavior and provide a more realistic view of work patterns in the platform. Through trace ethnography techniques, scholars can study individuals' historical records and make a history of events based on their traces. As a trace ethnographer, the author can gain more specific data to learn fine-grain details of interactions' history among volunteers in different circumstances. The author conducted virtual and trace ethnography to analyze volunteers' comments around finding possible new glitch classes to understand the processes of analysis developing a new proposal by occasional groups.

The author also conducted interviews to expand her understanding of how volunteers perform advanced work of categorization and if there are ”groupy” behavior and analytical moves beyond the findings through ethnography. She invited thirty volunteers who had either left sig-

nificant numbers of comments, over 300 comments and were in the process of developing a new glitch proposal. She sent invitation messages to volunteers on Gravity Spy site. Six volunteers attended the interviews, three volunteers preferred to write their responses to interview questions, and the rest were either not interested or did not respond to the invitation message. She specifically focused on Crown and Tealight proposals.

3.4 Data Analysis

3.4.1 Study 1- Group characteristics

In the first study, the author did a virtual ethnography to understand how volunteers manifest group behavior. She employed the model of ”groupy” behavior defined by Arrow et al. (2000) while observing volunteers' communication on the talk page. Since the current analysis focuses on understanding the behavioral examples of ”groupy” behaviors, the analyses of comments left by individuals who did not interact with any other volunteers are not included in this study. The author investigated conversations between at least two volunteers who exchanged at least two comments to identify if the conversation manifest group behavior. Exchanging at least two comments usually guarantees that two volunteers communicate and understand each other about a particular topic. The author conducted analysis between April and October 2019 and discussed her notes with three other researchers to get their feedback while writing the results.

According to Clark and Brennan (1991) understanding is a criterion for grounding in a conversation. The speaker should believe that contributors in a conversation have understood the utterances, although understanding is not flawless. As soon as the speaker utters, the listener should provide evidence of their understanding. If the speaker receives contrary evidence, they restate what they have already said, and if there is no contrary evidence, they assumed that the listeners have understood. For instance, in Gravity Spy, volunteer A asks a question, and volunteer B answers the question. If volunteer B leaves no further comments, volunteer A assumes that volunteer B has understood their answer. They may continue the conversation in the private messaging system that is not what researchers have access to for analysis. Having two rounds of commenting implies a conversation between two people.

After focusing on volunteers who had a conversation, the author investigated if any of six

”groupy” behavior manifest in one or more than one round of conversation between them. The overarching theories that provide context for different aspects of group identification are group formation, group processes and group success. For instance, group formation is a precondition for cognitive and affective resources of group identification come (Henry et al. 1999). Consequently, the author considered three stages of a group, how a conversation started (group formation), how they worked together on what (group processes), and if they ended up with a result (group success) to identify if they manifest six criteria of ”groupy” behavior.

As mentioned above, cognitive and affective aspects of group identification comes after group formation and the author took into account whether a group of two people started conversing together over the six months of observation or earlier. She queried the log data of comments to see when two people started conversing together. The comments dataset shows the board id, board title, board description, discussion id, discussion title, comment id, comment body, commentor user id and the timestamp of the comment. Then she investigated how their conversation exhibits six criteria in a discussion thread in processing and result stages.

3.4.2 Study 2- Categorization processes by occasional groups

After identifying ”groupy” behavior of volunteers through analysis of six months observations, the author investigated their analytical moves through analysis of volunteers’ comments around possible new glitch classes. She employed the group model by Arrow et al. (2000) and the categorization model by Grodal et al. (2020) to understand how volunteers work together to develop a new category of data and describe different episodes in a process of developing a proposal. She went through the following steps to identify the processes:

1. She collected hashtags mentioned in a few proposal that has been used to describe a pattern.

2. She collected all discussion threads, including any variations of each hashtag (i.e., the term or the hashtag, singular or plural).

3. She started an investigation to understand who has introduced the hashtag by the first mentioning and if they have used a term before using the hashtag (i.e., using caterpillar

before using #caterpillar).

4. She read discussions around subjects in which at least one person used the term or the hashtag to get involved in a discussion. To expand understanding of a discussion, she also read discussion threads linked to the thread to better understand the context.

5. Since discussions around a possible new glitch class continues after submitting a proposal, she continued the analysis until there were no group discussions around the hashtags listed in the proposal.

6. She took notes of discussions that show group behaviors using the model by Arrow et al. (2000) while considering their analytical moves using the model by Grodal et al. (2020).

7. She identified different techniques of categorization based on the model by Grodal et al. (2020) and how individual perform these activities as a group using the model by Arrow et al. (2000). She also considered themes that were not defined by the models.

Meanwhile, the author conducted thematic analysis of interviews using analytical moves framework to understand how individual perform categorization and work as a group and to make sure she has expanded her analysis to include all relevant themes to group categorization.

The author conducted thematic analysis on transcripts of interviews by analyzing responses to questions and new questions that emerged over the interviews; she first labeled words that would belong to one theme and labeled each. Since the goal of interviews was to understand how volunteers find new possible glitches, the author identified three themes, including how volunteers perform advanced work, naming a new glitch class, and collaboration patterns among volunteers.

3.5 Conclusion

This chapter explained the research setting and data elicitation and analysis for two studies. The first study was informed by the group theory developed by Arrow et al. (2000) to identify characteristics of ”groupy” behaviors and to explore the nature of work and collaboration among

volunteers. At the same time, the author considered overarching theory, the input-mediators-output model, to provide a context for identifying a group. Since occasional groups are formed based on their interdependent activities, identifying groups is feasible if they are formed and performed a shared activity. The second study applied the group identification theory by Arrow et al. (2000), and the model developed based on categorization theory by Grodal et al. (2020) to understand how occasional groups analyze data to develop new categories of data.

By conducting these analyses, the author answers the following questions:

- What are the characteristics of occasional groups in a crowdsourcing platform?
- How occasional groups perform analytical moves categorizing a possible new class of data?

Answering the first question advances our understanding of groups who do not conform to conventional groups and teams in organizations but still benefit from certain aspects of group behavior. The author hopes to elaborate on how groups make different analytical moves towards finding a new data class by answering the second question. The second study is complemented with interviews to have an inclusive and comprehensive understanding. The second study shows how an occasional group can accomplish innovative tasks despite issues raised as part of their nature as an occasional group.

CHAPTER 4

Results

4.1 Introduction

This chapter reports the findings of two studies and explains how the results of each study informed the next one. The first section shows the findings of virtual ethnography on behavioral examples of groups in the Gravity Spy project. The second section reports the findings of virtual and trace ethnography that investigates occasional groups categorization in addition to findings of interviews.

4.2 Study1- Group characteristics

The author investigated comments exchanged between volunteers from April to October 2019 to understand how individuals manifest group behavior in Gravity Spy. She employed group criteria developed by Arrow et al. (2000) to identify themes showing group behavior. She specifically answered the following question:

RQ1: What are the characteristics of occasional groups?

First, she found out that volunteers communicate differently, and their communication is not limited to using the reply-to button. Volunteers communicate with each other around various topics mostly on a relevant board using different techniques as follows:

- Leaving a comment mentioning someone else using their user_id, or name
- Leaving a comment in reply to someone else's comment using the reply-to function
- Leaving a comment in reply to someone else's comment mentioning their name or user_id

- Leaving a comment in reply to someone else's comment without using the reply-to function, mentioning their names or user_id

Learning how volunteers communicate with each other helped to better follow their conversation and identify if "groupy" behavior is emerging and define its characteristics. For example, since the author started observing volunteers, she noticed two advanced volunteers, whom I refer to as Sarah and Oliver, discussed several glitch types. They are one of the groups working together since the project launched officially in October 2016. They also had several discussions over the six months of observation, and in July 2019, they had a long conversation about falcon. The author focused on this discussion to see if there is evidence of group behavior. Oliver started a conversation explicitly mentioning Sarah and asked her opinion on an image and if it is a falcon- an example of a type of glitch called a falcon by other volunteers. The act of mentioning one specific person out of many volunteers who left comments around the same time, in addition to their prior conversations, shows that Oliver recognizes Sarah as a group member from the rest of the community- one of "groupy" criteria by Arrow. Also, as they dedicated lots of time to discuss falcon and several other topics to develop knowledge, the author concluded they might both feel connected to the work due to the time investment. The author then focused on what they have been discussing about the falcon and whether they coordinate their actions and use a shared tool or resource. They conducted two sets of analysis, including comparing characteristics of two glitch types and curating their collections. They both used the same measures to compare two glitch types, falcon and chirp. Also, they shared what images could be added to a collection of chirps. Sarah and Oliver shared their understanding of the image and collection that eventually helped them reach an agreement. So, the author concluded that Sarah and Oliver coordinated their analyses by sharing their knowledge that ended in a shared understanding and agreement as an outcome.

Below, the characteristics of groups are described based on each criterion.

Individuals consider themselves as members of the group: volunteers leave no comments explicitly saying they consider themselves as members of a group. The closet behavioral example that suggest they may consider themselves as group members is when a group of volunteers discuss one or several images only together rather than other volunteers who also left comments

on the discussion thread.

Individuals recognize each other as members and distinguish members from non-members: volunteers who work together on different images, recognize each other by mentioning either names, user_logins, or abbreviation of names to discuss images. Their interactions with specific volunteers over different topics on each board, either task-related (i.e., know what other person would like to know about particular topic) or social-related (i.e., knowing about each other daily life) imply behavioral examples of how volunteers recognize specific individuals as group members. Not having similar conversations with other active volunteers on the talk page implies how they distinguish members from non-members.

Members feel connected to other members and projects of the group: extended activities around different topics with specific volunteers imply certain volunteers have a sense of belonging to other members and projects. Regarding the feeling of connectedness to projects, members of a group interact with each other several times to identify glitch types and relevant tasks (i.e., providing information resources). They also show the sense of being connected to other members by disclosing their personal life and sharing interesting topics of the project with them. However, since most group interactions focus on tasks, the feeling connected to other members is not as clear as the feeling of connection to the project.

Members coordinate their behaviors in pursuing collective projects: volunteers who work closely on different tasks such as identifying a glitch type, labeling a possible new glitch, or projects like developing shared collections and proposing a new glitch class to the science team coordinate their actions over these matters. They may collaborate on developing a collection of images by collecting and adding subjects that are similar to each other. Volunteers who work on classification together mention each other on different discussion threads to seek consultant, assessment, or informing about a new pattern or naming a new pattern in glitch images.

Members coordinate their use of shared tools, knowledge, and other resources: volunteers share and coordinate the use of different resources. Considering two main activities of volunteers, classifying glitches to known classes and finding possible new glitch classes, volunteers working together share and coordinate the use of collections, hashtags, external resources such as publications, and helpful posts found on the internet. Once one of them share such

Table 4.1 Characteristics of groups in Gravity Spy according to group criteria by Arrow et al. (2000)

Group Characteristics in Gravity Spy
Individuals consider themselves as members of the group
A behavioral example of this feature is when a participant discusses tasks with specific people out of all who left comments. However, the data does not speak to what an individual perceive about group membership.
Individuals recognize each other as members and distinguish members from non-members
A behavioral example of this feature is when participants converse with specific people using their user_id to discuss a task. There are also participants with a more extended history of interactions. For instance, they know what another member is doing outside Gravity Spy and converse about their personal lives while working together. They do not have such a conversation with other individuals.
Members feel connected to other members and projects of the group
A behavioral example of this feature is when a participant keeps working with a group for months and shares personal matters beyond the project. However, the data does not speak to if an individual feels belonging to a group.
Members coordinate their behaviors in pursuing collective projects
They leave hashtags and mention a person when they want to inform them what they have labeled an image. These two people usually have a history of conversation about a particular glitch and coordinate further actions using hashtags and @mention. They mostly use this kind of practice while collecting images to create a new class of glitches.
Members coordinate their use of shared tools, knowledge, and other resources
Volunteers use hashtags and collections as two resources to either collect glitches of current classes or find new classes of data, they start sharing these tools and coordinate their actions over the use by adding more similar images to the collection and using the same hashtag for similar images.
Members share collective outcomes based on their interdependent activities in the groups
Volunteers who propose a new glitch class through submitting a proposal share it with their collaborators and the rest of the community

tools, other members would ask to remind them where they could find the resource to better analyze an image.

Members share collective outcomes based on their interdependent activities in the groups: members of a group share two main collective outcomes; they share collections and proposals for possible new glitch classes as collective outcomes. However, some collections are also the means to develop a proposal. Table 4.1 summarizes findings.

Considering the first three criteria as a social-related category and the last three criteria as task-related category, groups manifest different variations of these two main categories. On one hand, there are groups who work together over months on different tasks and topics; they know each other beside what they accomplish together. The heavy activity of these groups distinguishes them from the larger community. This type of groups are long-lived and highly interconnected. On the other hand, there are groups work together on a smaller set of tasks

over weeks; they usually focus on tasks only rather than sharing any personal stories or events. The boundary between this type of group and the bigger context is a bit fuzzy as the amount of their activities and shared outcomes are not as outstanding as compared to the first group. This type of groups are ephemeral and moderately interconnected. While either type of groups accomplish one or several tasks together, neither of them have a dedicated space and identity in the community. Also, interactions among members is time to time, what the author called occasionally. Below examples of these actions are reported. The conversations happens asynchronously among members.

Sarah, Oliver, and Sherry, discussed many glitch images on the Notes board over six months. They discussed these images occasionally, depending on what system was showing to them. These volunteers chose each other to discuss different topics, know each other's names, and know who knows what about a particular topic. They also know what might be interesting for another member regarding what studies they are pursuing together. One of the conversations they had was about whether a particular image is a whistle variation, one of the official classes in Gravity Spy. They had a conversation around analyzing the image, while three other volunteers left their individual opinion about the image, but were not involved in their conversation. Sherry and Oliver had talked about how to analyse a certain type of a whistle, a double whistle. Below is a thread of conversation among Sarah, Sherry, and Oliver. Figure 4.1 shows the image of discussion.

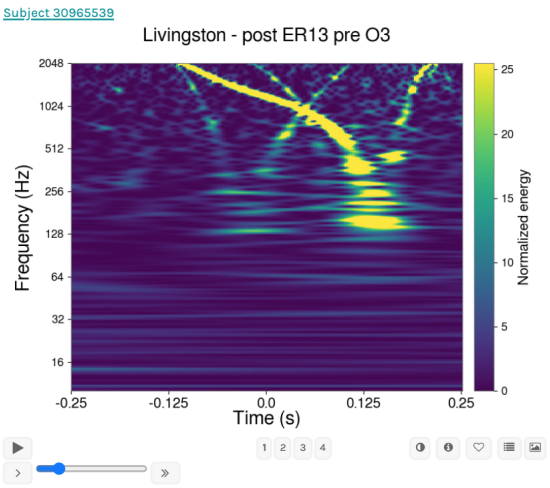

Fig. 4.1 A double Whistle discussed by Sarah, Oliver, and Sherry

Sherry: "*Where on earth did the fifth arm come from? @Sarah* "

Sarah: "*@Sherry, I had exactly the same thought :) an orphan arm of the whistle which stops at the point where the other two whistles cross. (or maybe it continues and overlaps with the second strong whistle. in this way, the shape of the first and second whistles would be pretty mirror symmetric)*"

Sherry: "*Ha Ha, hi Sarah (smiley emoji), but which one is the 5th arm? After looking at this again and especially as seen in the inverted image, could there be a 6th originating at 750 Hz and -0.05 s? or just an illusion? @Oliver hi Oliver, can you calculate if the whistles cross each other in the middle of the glitch? I think I asked for your help with this type of subject before but cannot locate the conversation can you (smiley emoji, thinking emoji)? Sherry*"

Oliver: "*Hi, Sherry. I remember the conversation, but I can't remember the context. The best estimate I can make of the crossing point at .04 sec is 950-1000 Hz (a bit indeterminate). My impression is that the strong arm could divide at the cross-over point. Like you, I can see several signals that could be further, shallow, whistles, including the one you can see. I can see a further 2 or 3, faint, dipping down from the LH side; at -0.2sec, 700, 1100, and 1500 Hz. Oliver*"

Sarah: "*I can see these weak components too now you both have described; the negative image helps a lot; thanks for the suggestion!*"

On a different discussion, Oliver started a conversation by mentioning Sarah, discussing the image type and if it belongs to the falcon category. Sarah explained how it also resembles chirps in addition to falcons. Later, Sarah mentioned she needs some images for a collection related to the image, and Oliver offers his help and shares his collection with her. Sarah examined his collection and identified correct and relevant images to the current image. Below is the thread of conversation between Sarah and Oliver and figure 4.2 shows the image.

Oliver: "*Could this be a poorly formed falcon? @Sarah*"

Sarah: "*Yes, I think it is exactly a poorly formed #falcon. I would also mention a possible weak #chirp candidate in the background at 0.3 sec 100-300Hz.*"

Oliver: "*Well spotted! I can trace it from -1 sec, 15 Hz to 0.35 sec, 250 Hz. It looks interesting @Sarah*"

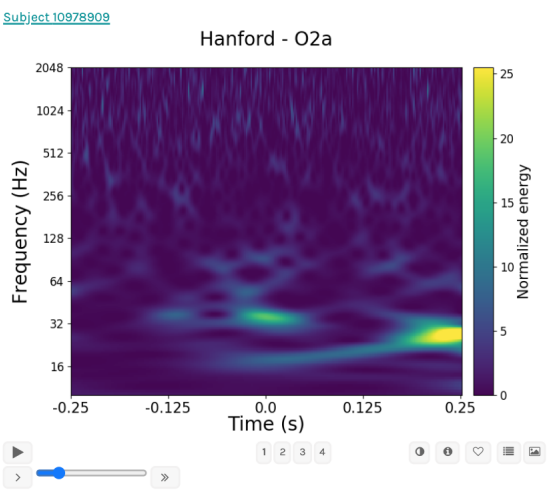

Fig. 4.2 A falcon discussed by Sarah and Oliver

Sarah: "*Thank you Oliver, I can trace it back too. If the falcon was not there maybe the SNR could be more in favor of the chirp. During O2 Hanford was more sensitive than Livingston as Jack mentioned it earlier, plus because of the coincident falcon this chirp might have been unnoticed. I think it would worth to check the other detector for possible coincidence.*"

Sarah: "*... I would like to collect several good candidates for this purpose.*"

Oliver: "*I will go through my collection of chirps, and let you know which ones look potentials for your new collection*"

Oliver: "*I have been through my collection, and the following subjects are worth considering for the Weak chirp candidate collection: ...* "

Sarah: "*@Oliver thank you for the list, i went through on them. these two I think are injections ,...* "

There are also volunteers who work as a group on a smaller set of images and accomplish a few tasks together. However, this type of groups do not have many shared projects and tasks as the first type.

Patricia and Sarah discussed some images over two months. The following figure 4.3 shows an image where Sarah and Patricia discussed together if the new pattern could be a variation of scattered light, one of the official classes the system. Patricia responded to Sarah's comment after about two months once she had started commenting on the talk page. In response to Patricia's comment, Sarah agreed with her that there is a new pattern and added another volunteer,

John, defined a new hashtag medusa to describe the new pattern. Patricia and Sarah discussed more images over these two months.

Sarah: "*very unusual, looks like extremely strong scattered light and strong low frequency bursts*"

Patricia: "*dunno, light scatter usually doesn't have teeth. I'd call it midfreqlens and see if I find something alike to call it possiblenewglitch I definitely seen some extremelyloud things with this kind of jagged border, but never it was in a shape of a lens.*"

Sarah: "*@Patricia yes, the teeth are new features. Even if it is related to scattered light it is a new type of it, if it is unrelated then it is a new type on its own right. @John started to call them medusa*"

Patricia: "*@Sarah, medusa certainly has a nice ring to it. I'll call them that too. it does look like a jellyfish)*"

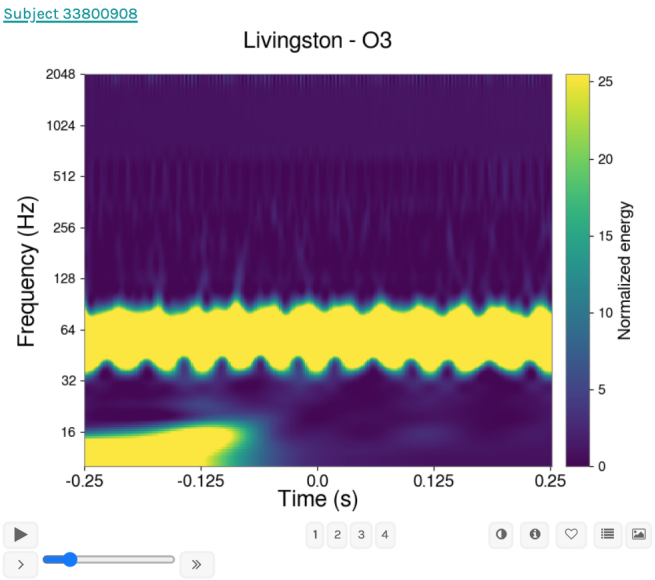

Fig. 4.3 A new pattern discussed by Sarah and Patricia

Patricia discussed another new pattern with Sarah on a few weeks later; She suggested that the pattern is a new one as she sees it very often. Sarah agreed with her and told her that they should find a name for it. Patricia explained how the boundary of the pattern has overlapped with other classes. Sarah replied that she creates a collection of similar images using the similarity search tool to see if the algorithm could find more images of the pattern, but there were only 19. Consequently, Sarah and Patricia added some more images to the

collection. Figure 4.4 shows the image of their discussion.

Patricia: "*repeating #tomte #tomtekoi I see this pattern a lot. maybe I should make a collection.*"

Sarah: "*I see them too. We should find a good name for them.*"

Patricia: "*to me it looks like a spaceship (but honestly, what doesn't look like a spaceship?) or a pine forest (but it's almost always symmetrical, and the name says nothing about that) marching tomtes? I mean, they do look like they're always going in some kind of pig-head formation with the middle tomte ...*"

Sarah: "*Thank you for the suggestions! I agree, the linked subjects are member of the same highly variable group. I think if we can separate one or several different sub-groups with fixed morphology it still can be useful. The subject seems to be a good candidate to start with because of the symmetry and regular period and the simsearch returned many good matches, I haven't saved the results but I will repeat it and will see then how to continue after.* "

Sarah: "*Subject 35582971 is present in 19 copies and .... and the link collection is here..* "

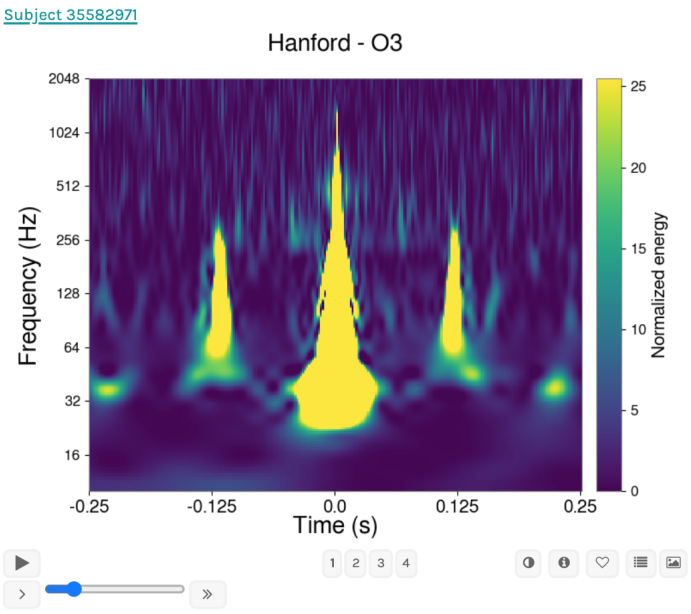

Fig. 4.4 A tomte like pattern discussed by Sarah and Patricia

4.2.1 Summary

Through virtual ethnography, the author identified characteristics of occasional groups. According to the findings, occasional groups are task-focused groups that come together to accomplish different goals occasionally. Individuals come together to work on a task (i.e., analyzing a glitch type) rather than forming a group before working on a task or a project or the existence of groups before starting a new project. The beginning, process of work, and shared outcomes help identify them as a group and understand their characteristics.

These groups share their understanding of a glitch type to reach a common ground for further analysis, create and curate a shared collection of glitches, and sometimes develop a new glitch proposal. They coordinate their actions using @mention to call another member's attention to the work. They also mention another member to recognize their work and/or share an outcome. While most occasional groups are focused on accomplishing tasks and do not develop social bonds, some groups develop social bonds. This type of occasional group meets six criteria of a group definition by Arrow et al. (2000). However, their activities are still occasional compared to their individual contributions to the community.

While occasional groups succeed in accomplishing a task or a project, they do not have all resources that would support teams within an organization. Occasional groups do not have a clear direction and specific plan to accomplish a task or a project. Depending on members' involvement, groups work together to accomplish a task (i.e., identifying a pattern in an image) or a project (finding a new category). As a result, group interactions are temporally protracted, the group may last for months or weeks, and group accomplishments vary accordingly. For instance, some conversations have been activated after passing two months from the first comment on a glitch. Most groups last for weeks and there are a few groups that stay active over months and years. Consequently, groups stayed longer accomplished developing more proposals compared to the rest.

Occasional groups emerge organically due to the task complexity that would increase interdependency. As a result, group members voluntarily coordinate their actions and shared resources to accomplish a task or a project. Consequently, occasional groups do not have specific roles (i.e., leader, tag gardener) and task assignments as there is no explicit group membership

and commitment. The only defined role in the community, not within a group, is a moderator. While moderators are the only ones who can submit the proposal to the science team, they do not assign a task to group members or manage interactions of group members. Occasional groups also lack training resources to develop a new category and work as a group. While occasional groups can work and submit a proposal through a moderator, they still need the science team's approval on their proposal submission. The following study investigates their interactions and analytical moves to understand how occasional groups develop new categories despite the lack of a specific structure, plan, and resources.

4.3 Study 2- Categorization processes by groups

The second study investigates how groups make analytical moves to find a new possible glitch class. The authour employed the framework suggested by (Grodal et al. 2020) to understand and explain how occasional groups find new categories of data. The study answers the following question:

RQ2: How do occasional groups do perform analytical moves required for categorizing a possible new class of data?

The author explored all proposals and investigated traces of groups around four possible new glitch classes to identify the main analytical moves that groups do while categorizing images and identifying a possible new class. Table 4.2 shows the result of analysis. For example, the author selected one of the early proposals, Spirograph, submitted to the LIGO New Glitch Proposals board, the Spirograph proposal, to analyze how occasional groups develop the category. The discussion thread of the proposal shows that several people were involved in the process of categorization that implies the individuals' interactions and group categorization would increase. The author retrieved all comments from the comments dataset posted by more than one volunteer, including all variations of hashtags, including loophole, pixlated, helix, anthill, and scatteredlight. The author also retrieved comments, including all variations of labels added later to the proposal, including tapestryconv15 and organ pipes.

After retrieving comments, the author retrieved discussions on the Talk page of the Gravity Spy to reach through the whole discussion and understand how a group started categorization.

For instance, the term loophole was the first hashtag to describe the pattern of Spirograph and introduced by Mina. Mina asked Sarah and Sherry if they agree that there is a new pattern and if they agree to use the term loophole to describe the pattern. Sarah and Sherry agreed with her and started using the loophole on more images. Sarah and Sherry also compared some images to decide if they belong to the official classes, light modulation and scratchy or the new category loophole. Since this group of volunteers started categorizing a new pattern by asking a question and speculating if there is a new pattern, using a new name on more images and labeled more images and related the new category with prior categories, the author concluded they performed analytical moves of categorization. As they asked each other's opinion about the existence of a new pattern, use a new name, label more images, and compare a new category with other categories to reach common ground before coordinating their actions, the author concluded they performed categorization as a group. Since these group analytical moves are occasional compared to their individual analysis and contributions, the author considered them occasional group categorization.

The section is organized around three main analytical moves, asking a question, merging categories, relating and/or contrasting categories. Themes emerged from occasional groups working on developing a new proposal is reported for each analytical move. Then further analytical moves by occasional groups are explained.

4.3.1 Asking a question

(Grodal et al. 2020) explained that qualitative researchers might investigate data having a specific question for the purpose of categorization. In Gravity Spy, while individuals may investigate each subject by asking questions, such as what the pattern looks like in an image or what possible causes are for a particular glitch, they share their questions while working as a group. For instance, a volunteer would ask about the pattern in an image or speculate what physical causes are for a particular glitch. Volunteers who work as a group coordinate their behaviors in pursuing a collective project through asking questions and seeking agreement. They also coordinate their shared tools (i.e., a hashtag, an image) and knowledge (i.e., knowledge about a cause) to better identify causes.

Speculating on a new pattern: One of the primary steps to starting developing a proposal is when volunteers notice a new pattern. As an occasional group, they consult with active members around the same time- active members leave comments on different discussion threads frequently). on June 23rd, 2017, three advanced volunteers labeled a possible new glitch type by using a new hashtag. Mina was the first person who tagged the subject 10539005 as #loophole and added that it is similar to #curtain and #scratchy. Hashtags #curtain and #scratchy were used by these volunteers earlier. Also, she asked Sherry and Sarah (by mentioning their @user_ID) to see if they agree the subject is a new type and loophole is an appropriate hashtag to use. Sherry and Sarah started using this hashtag on similar images.

While these volunteers started finding and labelling similar images using the new hashtag, using a new hashtag on more images is not always an easy task, as the pattern may not appear the same in different images. The complexity of a pattern brings more interdependence among volunteers. For this particular hashtag, two volunteers, Sarah and Sherry, were working closely from the beginning; they would often ask each other questions about the pattern. For instance, once Sherry was not sure what name would describe a subject best and asked Sarah's opinion. Sarah responded to her by recommending using the hashtag, loophole that Mina had introduced earlier. A few minutes later, Sherry consulted about another subject with Sarah and mentioned that the background looks like scattered light and scratchy but she chose light modulation. Sarah agreed with her about choosing light modulation and mentioned that the background reminds her of the loophole.

These advanced volunteers started using the hashtag simultaneously and checked several images together to decide if the pattern is a loophole. Since three members learn about the new pattern simultaneously, they usually do not have disagreeing opinions about the existence of the pattern. A bit later, another advanced volunteer learned about the new possible type and joined the rest of the journey of finding more similar images a bit later.

Speculating causes of a glitch type: Volunteers speculate about the causes of a glitch to better understand if a new pattern should form a new class of data. They share their perspectives and seek what would causes a glitch. This understanding would help them to better understand the glitch type and if it is a new possible glitch class.

For instance, Sarah, Oliver, and Sherry worked more intensively together and tried to understand some glitches' causes. A week later, in late June 2017, Oliver started a discussion thread about subject 10524237 on the Note Board, commenting on its unusual resolution compared to other images. Sherry agreed with him and explained that the high-energy signal in frame 1 has stepwise edges or is pixelated. She asked Sarah if she had further explanations. Sarah responded that the pixelated appearance might be a consequence of how spectrograms are created. However, she added that some physical causes might result in stepwise edges or pixelated images in rare cases. She further elaborated how different components of the wavelet detection filter algorithm or the time window of data sampling changes the resolution and, in some cases, patterns like loophole or tapestry emerge. She emphasized that it would be difficult to separate the effect of data processing from the precise details of a physical waveform, but with knowledge of the physical cause, they could better understand a glitch type's morphology. She also shared one of her older comments about fixed points to support her explanation of this pattern and the cause. Oliver, Sarah, and Sherry continued the discussion on the Chat Board and asked a science team member if there has been a new way of creating spectrograms that causes pixelated images.

After the science team member informed them that there was not a new style for creating spectrograms, these volunteers considered pixelated as a prefix for official classes with the same pattern. A couple of weeks later, by the end of July 2017, Sherry described a new subject as a variation of scattered light that is pixelated. Sarah added that the glitch type variation is very informative regarding the physical causes and the math and methods of creating the spectrogram. Then Sherry mentioned that this pattern reminds her of a favorite toy named spirograph. Sarah expressed her interest in this analogy and found it relevant for describing a complex oscillatory motion. Since then, these two started tagging more images using the spirograph. This new hashtag is what Sarah used later to name the glitch proposal that she submitted for review by LIGO scientists.

Sarah and Sherry agreed to use a new name easily. The decision-making process was quick. As they both developed a shared understanding of the pattern, they could tag several images and agree with the name. Also, Sarah could relate to Sherry's choice of spirograph as she believed

Table 4.2 Analytical moves of volunteers in identifying possible new glitch classes based on the model by (Grodal et al. 2020)

Analytical moves	**Behavioral patterns in Gravity Spy**
Asking a question	Speculating on a new pattern
	Speculating causes of a glitch type
Merging categories	Elaborating characteristics of a newer pattern
	Creating and updating a list of hashtags
Relating and/or contrasting categories	Specifying boundaries of similar glitch types
	Marking what does not belong to a category
	Sharing resources to identify relationships among glitch types
	Use of similarity search tool to create/improve collections
Developing working hypotheses	Submitting a new proposal
	Updating the proposal by including similar hashtags

the name presents the pattern perfectly.

4.3.2 Merging categories

(Grodal et al. 2020) explained qualitative researchers categorizing their data might combine two or more current categories to create a superordinate category. In Gravity Spy, merging categories is common as volunteers usually create several categories for one overarching pattern that appears slightly different in each category. Volunteers who work in a group may create a new category but they may figure out if it can be merged to the bigger category. They coordinate their behaviors and shared knowledge to elaborate the characteristics of a new glitch type and eventually combine it with an existing category.

Elaborating characteristics of a newer pattern after proposal submission: Volunteers share characteristics of a new pattern within a group to reach a common ground and better coordinate their actions for labeling further similar images. This process also helps them to understand if they should merge two types and consider them as one possible new glitch class. They rely on different resources to communicate what a pattern means and if it belongs to a bigger category. They either explain characteristics of a pattern, share their collection to show more examples, use the similarity search tool to retrieve more similar instances, or raise questions that would shed light on the importance of merging categories.

As an example, Sarah, Sherry, and Oliver discussed a new type named tapestry. While Sherry and Sarah were more familiar with the pattern, Oliver kept exploring and asking ques-

tions that eventually helped Sarah identify a new glitch type related to the spirograph and combine two classes. Late February 2018, Sherry first commented on an image and described the shape. Then Oliver expressed that type is new to him because of the lower band's constant frequency and repeating features. Sherry and Sarah helped him to learn more about this by sharing more similar images. Also, Sarah shared her collections of tapestry to better explain the new type. Oliver responded how he analyzed the image and asked if it is reasonable to expect a horizontal line to be generated at frequency X if the glitch repeats at a frequency X/sec. Sarah agreed that this might be the case in this type of glitch and mentioned that she would explore this in her collection to find similarities.

A bit later, in March 2018, Sarah, Sherry, and Oliver discussed an image similar to what they have already discussed in February and agreed that the image is tapestryconv15. Oliver mentioned the same pattern is repeating for about 15 seconds. Sarah agreed with him and told him that she created a dedicated collection for tapestryconv15. She shared the collection with Oliver for further exploration and observing similar images for this category.

In May 2018, Sherry and Sarah discussed a couple of images and decided both images are tapestryconv15. Sherry shared the similarity search tool's result, and Sarah told her that one of those images is tapestryconv15. Sherry pointed out that another image in the search result is tapestryconv15 and would be helpful for Sarah's collection. Sarah agreed with her and added the image to the collection on which both were working.

In June 2018, Oliver raised a question about a subject tagged as spirograph and asked Sarah if it is related to tapestryconv15. Sarah had to do a comprehensive comparison of two types in two collections and found out these two are similar enough to be merged. Subsequently, Sarah added this type to the spirograph proposal. Since Sarah had also shared her collections with Oliver, Oliver could examine more images that made him wonder about the pattern of tapestryconv15. As he shared his question with Sarah, Sarah could reexamine the collections and realize these two types are similar enough to be merged. A close collaboration of these two members within a specific period facilitated the process of merging categories. However, as the collaboration is sporadic, it took a while to come to a new understanding that the new hashtag should be merged with what they already proposed.

While several proposals have a list of hashtags or collections added later, showing merged categories, merging categories also happened at the beginning of proposal submission. As 5.1 and 5.2 show 12 out of 28 proposals submitted over three years, have a list of hashtags. The primary list of hashtags is what the moderator collected based on similar names on the same discussion thread describing a pattern. Sarah, Sherry, Oliver, Coryn were working on a different proposal; cord and its segments. After a while, they doubted if the cord and caterpillar are the same categories. They discussed several images, whether they belong to the cord or caterpillar. Sherry and Coryn left comments on several images asking if they belong to the cord or caterpillar. Oliver and Sarah talked together that they think there should be a similarity between cord and caterpillar. Both volunteers learned caterpillar has a lower Q value than the cord in the last frame of an image, but since the first frame was very similar, Sarah combined two categories under one proposal and submitted it for revision by the science team.

Creating and updating a list of hashtags:As a part of organizing hashtags suggested by volunteers, Sarah and Mina were working on adding new hashtags to a hashtag list to keep track of all tags on the Collection Board. Sarah created the list in January 2017, and Mina and other volunteers, including Sherry, Oliver, Harry, Sandy, and David, helped update the list. Mina were going through many subjects and identified two groups of frequent and infrequent hashtags and asked Sarah to update the list. David and Sherry provided some descriptive statistics on the frequency of hashtags and explained since half of the hashtags are mentioned once and if it worth deleting those tags. However, other group members did not agree and as a result, no one deleted the tags.

Sherry, Oliver, Harry, and Sandy helped identify a few new hashtags and mostly specified similar hashtags that describe the same morphology or shape. They asked Sarah to edit the list and merge those hashtags. Sarah edited the list and later in September 2017, Sarah created a document on Google and invited all of them (Sherry, Oliver, Harry, and Sandy) to edit the list and make the editing process as a collective action. However, as the process is tedious, the group did not update the list.

4.3.3 Relating and/or contrasting categories

According to (Grodal et al. 2020) qualitative researchers compare different categories to identify relationships between them. They might use this method to better understand what categories should be merged or spilt. In Gravity Spy, volunteers who work together try to identify boundaries of a pattern that make it distinguishable from other similar categories. They specify relations among different categories and conclude if the new pattern should form a new category. In this regard, they coordinate shared knowledge (i.e., shared external resources) and tools (i.e., collections) to conduct the analysis.

Specifying boundaries of similar glitch types: volunteers were comparing different images to better specify the boundaries of a pattern. Sarah and Sherry started a long analysis of some subjects, starting analyzing the subject 10539443. They shared more images that looked like each other and tried to understand the boundaries of the pattern for similar images and what makes one subject different than the rest. Sarah listed a subject as a possible similar image to what Sherry had shared. They also discussed other pattern, falcon and its variations concurrently to decide the boundaries for each type. However, they did not have a clear vision how these categories are related.

After a couple of months, early November 2017, Sarah wrote a post describing how different clusters of official classes and hashtags are related to each other. She focused on violin mode harmonics and new types, 2secZipFalcon and 1secZip, which fall under this official class and linked them to a cluster of two other new types, ETMY scattering and spirograph, which she identified as a subcategory of extreme scattered light. She provided collections of images for each type and explained the spirograph pattern is very sensitive to different Q values and for lower values of Q it resembles helical patterns and for higher Q values resembles extreme scattered light. She added because of this sensitivity she will propose spirograph as a possible new glitch class. She also mentioned that the name spirograph was inspired by a conversation with Sherry. Figure 4.5 shows her hypergraph comparison. Sarah could help Sherry and other members to better understand how different categories are related, but they still form a separate cluster.

After sharing this comparison, she chose the subject 10538965 as a typical example of

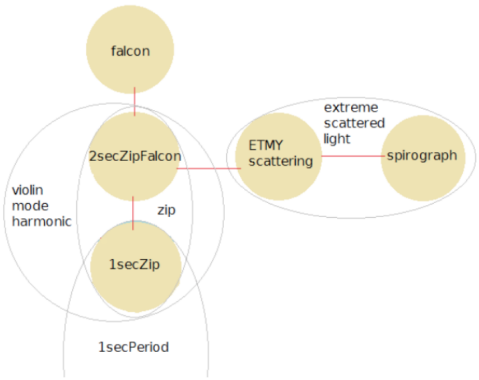

Fig. 4.5 A hypergraph comparison for related clusters of images

spirograph and submitted the proposal using this image- Sherry had tagged this image as loophole and scattered light earlier in July.

Marking what does not belong to a category: while groups develop new categories identifying what belongs to a possible new class, they also share what instances do not belong to a possible new class. Labeling what does not belong to a category is one of the activities that groups do to make sure not to collect subjects that do not resemble a new pattern.

One example of labeling instances that do not belong to a category was between Sarah and Oliver. They started noticing a new pattern in October 2018 and named the pattern fireball. They started investigating the new pattern by exploring several subjects together. Like other proposals, they discussed the glitch causes, relation of a fireball to another similar class, helix. Meanwhile, they labeled several images that do not belong to the fireball category. They did not leave any comments discussing if an image does not belong to the fireball category once they identified the fireball's boundaries. Marking what does not belong to a category helped both of them to better specify the boundaries of a new glitch type.

Sharing resources to identify relationships among glitch types: volunteers share internal and external resourced to facilitate the process of finding relationship among gltich types. Sarah referred to metadata that would help them better categorize images; she explained while a small difference in Q value has a huge impact on the background of these two images, the complex fractal pattern of two subjects is very similar. While Sarah explained the details of how to differentiate patterns, Sherry reminded her of a research paper that Sarah had shared before.

Sherry added that the paper compared the same simulated glitches with different Q factors as Sarah was explaining about the subjects. A couple of days later, Sherry and Sarah tagged more subjects using the loophole. In one of the images that Sarah tagged as loophole, Sherry left a comment adding the same hashtag and mentioned it is similar to subjects recorded by Virgo Detectors. She shared the a-log file, a useful external resource reported by LIGO scientists, to help Sarah better understand the rational behind her comparison.

Volunteers also shared their collections to better explain the difference between similar types of glitches. Early December 2017, while Sarah, Oliver, and Andy discussed whether a subject was spirograph or gearwheel, Sarah shared her two collections types to better explain how morphology of these types were similar, but the causes could be different. That would help Oliver and Andy to go through instances of each type and see how the patters are different in each type. Similarly, in different discussion threads in February and March 2018, Sarah shared her collections of tapestry and tapestryconv15 with Oliver and Sherry to better explain these two types of glitches and how they are connected to the spirograph. Sharing collections is one of the steps that helped group members to reach common ground about particular glitch types. Consequently, it allowed them to find similar images and expand the collection faster. However, as these work instances are periodic depending on what the system shows for classification, it takes time to find enough cases for a particular type.

Use of similarity search tool to create/improve collections: To analyze potential new glitch classes, participants employ various practices involving the technical infrastructure within the Gravity Spy website and auxiliary tools developed by project organizers that exist outside the platform. The process of analyzing a glitch after having classified images centers around practices to curate glitches scattered around the website.

Core members of occasional groups know how to use an auxiliary tool developed by project organizers, the similarity search tool. They use the similarity search tool to identify and curate new glitches. This group of volunteers, moderators, and advanced participants, have a more comprehensive knowledge of all current classes. They accumulated knowledge benefiting from classifications over years, various onsite resources, and external ones such as a-logs, public reports published by LIGO scientists, to understand various glitches' characterizations. They

first choose an image either through the classification system or tagged by other volunteers as a possible new glitch class and then use the similarity search tool to retrieve similar images. Then they select the search results that happened in the same period and create a collection using similar images. Once they have a suitable collection, they decide if the collection could be a new glitch class and try to increase the number of images within a collection to meet the proposal's requirement. These volunteers usually share their thoughts with each other to decide if the pattern is a new one. One of the moderators said:

> *The result of similarity search also shows a bar chart. There is usually an outstanding bar that contains most of the candidates, and it would be possible to create a new collection exclusively from those subjects that are correctly covered by the bar. Sherry and Oliver helped to expand the collection.[However], I find to cooperate on collections difficult: it is difficult follow if a contributor adds new subject to the collection unless they bring it into my attention or I bring to the attention of the owner of the collection when I am the contributor.*

While the similarity search tool helped core members of occasional groups better specify possible new classes, depending on the complexity of a glitch type, they still conducted an in-depth analysis of the pattern and causes of the glitch. Sarah said:

> *"Sometimes a pattern is very obvious and you can find a very homogeneous cluster through the similarity search tool. I know proposing this type of new glitch class to the science team is not very helpful as machine learning algorithms already clustered them very well, but it could be useful for other volunteers to see the collection and learn from it. However, in most cases patterns are complex and require more investigation."*

4.3.4 Occasional groups and Further analytical moves

While previous moves are aligned with analytical moves that qualitative researchers would have analyzing their data, occasional groups have other activities as the nature of their interactions and responsibilities are different. The main characteristic of occasional groups is also reflected

in the process of categorization of new glitch types. As they are working periodically based on what the system asks them to classify and their active time in the system, their activities are spread in the system on different discussion threads and the level of involvement in a group activity varies. Consequently, the process of collecting and synthesizing becomes complex and challenging. Also, occasional group categorization work at the beginning and end of the processes is different from team categorization as they need to apply techniques for coordination within the group and the community.

4.3.4.1 Different levels of involvements

Like other online communities, volunteers allocate time in different levels. While there are volunteers engage in different discussions to gain or share their knowledge, some volunteers consider classification as a way of meditating. This type of volunteers mostly focus on classification and only leave a hashtag for a subject to share their thoughts with other members classifying similar images. However, they do not engage in discussions describing glitch types as they do not want to slow down the pace of classification as a medium for meditation. One of the volunteers, Tom, said:

> *"I was not involved in discussions around characteristics of crown with other people. I want to see that most of the lot of images and commenting takes a lot of time. It's easier to do classification and see if there is a new pattern then leave a hashtag [for a possible new glitch class]. That helps you find the people that are looking at the very similar pictures and helps you find the most used name for it."*

Other members, however, engage in several shared activities to figure out if a new pattern could form a new glitch type. They would work on a collection together and define characteristics of a new pattern. They mainly invite each other to their collection to expand their collection and gather all similar images. They would also voluntarily inform the collection owner if they see similar images without being a collaborator on a collection. One of the volunteers, Ron, said:

"*When I leave a comment on an image, I check if someone has created a collection of that image, and I will mention the person if I see more images similar to their collection.*"

Moderators have more number of collaborations compared to other volunteers as they can only submit the proposal. They would work closely with some volunteers on several proposals and different tasks around finding new glitch classes. However, recording what every member contributed to developing a new proposal is not easy. Volunteers leave comments on many discussion threads, and reading all threads over time is not feasible for all moderators. One of the moderators said that she might have overlooked some of the contributions, but she believed another moderator follows and records all contributions.

As volunteers have different collaboration patterns, the output of their work varies. Volunteers who have been tagging images, collaborating on collections, and share their analysis of a pattern with a moderator have a more outstanding role in developing a proposal. However, for those who only left tags, their contributions were either not recognized or recognized later.

Sarah explained that some volunteers like Oliver, Sherry, David, and John helped create a collection necessary to submit recent proposals. She added that some other volunteers helped with finding and tagging some images. Another moderator said that he recorded all volunteers who tagged images he collected to create a collection for the proposal. He also included volunteers who helped to identify the boundaries of a pattern.

4.3.4.2 Informing group members at the early stage

One of the actions that helped groups of individuals to coordinate their work towards developing a new category was informing each other at the early stage of categorization. While three members of the group working towards developing a new category informed each other about using loophole or spirograph, the members of another group who worked on developing a different proposal for the same glitch instances did not inform each other to use a particular hashtag.

Cord and its segments is an example of proposals that group did not coordinate their actions at the early stage. In this work, Oliver and Sherry analyzed some subjects, identified a new pattern, and decided to name the pattern caterpillar, but meanwhile Sarah had chosen a different hashtag rope. However, after a few days, when Sherry and Oliver were using caterpillar more extensively, and Sarah noticed that they were using different hashtags for the same pattern,

after seeing Sherry's comment, caterpillar, on an image that she had analyzed. She told her that they are using different hashtags. Later in the afternoon, she created a discussion thread on the Collections board and mentioned Sherry and Oliver to explicitly communicate they are using different hashtags for the same pattern. She mentioned that she would use caterpillar and later she edited the subjects that she had labeled rope by adding "edit:caterpillar."

These individuals who worked together on the rest of the subjects identifying the same pattern did not coordinate their actions early in the project. While one of the members had to invest more time to edit what they have already tagged, she had the opportunity to use a name that she had thought would best describe the image.

4.3.4.3 Reaching a consensus on naming a new possible glitch class

Volunteers identify instances of a possible new glitch class using hashtags or describing characteristics of an image. However, most volunteers leave comments using hashtags to name similar images rather than describe an image's characteristics. This particular action makes categorization distinguishable from the act of tagging.

Volunteers tag various labels for the same image as each individual observes a different pattern in the series of images. They may also focus on different boundaries of a pattern within an image; some volunteers identify a particular pattern in an image, and others prefer to specify a broader pattern. They also use metadata about the image which contains the period the data object was produced (e.g., 20180825) and a q_value which is a numeric representation of a glitch's length across time (i.e., stretched or squeezing) and its height (e.g., narrowing or expanding) in frequency. Depending on their expertise on recognizing official glitch classes, they may add a tag representing a variation of an official glitch class.

An example of different names for the same pattern is a recent official class named crown. Sarah noticed a new pattern and named it sparrow. She tagged a few more images on the same day that drew the attention of Sherry. Sherry checked a few more images with Sarah to see if the pattern is a sparrow. Sarah explained to her that it looks like sparrows are sitting on a wire. Sherry implicitly agreed with the sparrow name, and both Sarah and Sherry tagged several similar images as sparrows. A couple of days later, Sarah submitted a proposal under the name

of the sparrow. A bit later, Sarah thought the pattern in in these images resembled other glitches that had been tagged with the label "crown", so she tagged them as sparrows and crown_subset. After investigating the characteristics of the pattern on several images, she decided to change the proposal's name to the crown and explained the change to Sherry, as they were working closely on this type.

In addition to the sparrows and crown, there were also other tags for several images. For instance, a couple of volunteers saw the pattern as paw prints or dumplings. These names were not recognized at the time of proposal submission, but the moderator noticed these names soon and added them to the proposal. These two members were in peripheral network whose work were recognized despite not having many interactions with the core members, Sarah and Sherry. Since a group of volunteers discuss and name a new glitch earlier, other volunteers with fewer number of interactions follow what the core members have chosen as a name. They believe using the most popular tag would help to have a less chaotic process to identify which images belong to a possible new glitch class. One of the volunteers said:

> *"I prefer to use the most popular hashtag as long as it looks like the same glitch. And so if I make a post and let's say I tag the subject with two different hashtags. And then later, I see that one is more popular than the other. Sometimes I've gone back and modified my comment. I've added the most popular hashtag and removed the less popular one."*

4.3.4.4 Managing the community

One of the possible ways to inform the bigger community of volunteers about a new hashtag is labeling more subjects using the hashtags. Consequently, the system shows the hashtag in the list of most popular tags.

Once the group agreed on using the term spirograph, Sherry told Sarah that there are not enough subjects tagged with spirograph, and they should tag more subjects. They both started naming more images using the spirograph. Several volunteers adopted the new term and tagged several subjects. Oliver and Andy tagged more images with spirograph than others. They both developed more confidence tagging more images using this hashtag after consulting a few

images with each other and Sarah. After a while, Sarah, Sherry, and Oliver checked back some older subjects tagged as loophole, scratch, helix, and scattered light and added the new hashtag spirograph to the discussion thread for those subjects. Informing the community about the internal decision after submitting the proposal under the name of Spirograph helped the group and the bigger community collect more instances of this new pattern.

4.3.5 Summary

The second study identified various analytical moves among occasional groups developing a new category. By applying the model by (Grodal et al. 2020), the author showed how volunteers categorize images as a group and what kind of their activities go beyond what qualitative researchers apply in analyzing their data.

Groups usually start exploring a new possible class by speculating a new pattern and its causes. Then they try to specify relations between similar images by specifying boundaries of similar glitch types, marking what does not belong to a category, and sharing resources to identify relationships among glitch types. They also update the library of hashtags for each new glitch type. They usually carry on discussions around possible new glitch classes after the proposal submission and merge newer categories with what they have already identified. They also update the list of hashtags as they use newer hashtags to describe the same pattern.

The study also showed that occasional group members might adopt a different term initially and later adopt what other members are using. Decision making is mostly a smooth process for these groups. Occasional groups also manage the community by using the new label on more images and promoting it among other volunteers.

CHAPTER 5

Conclusion

This chapter includes the discussion of results, thesis conclusion, limitations of the research, and further work directions.

5.1 Discussion

The investigation of behavior in Gravity Spy revealed a type of group that comes together in different intervals to discuss a task or a project in a crowdsourcing platform. This type has been named here as an ”occasional group”. The beginning, process of work, and shared outcomes help identify them as a group and understand their characteristics. The existence of such groups depends on the task that brings individuals to work together; they are emergent rather than being assigned by the organization. These groups stay active for the short or long term as members may either stop contributing to the platform or decide to do less work for a while. Most group members create bonds only around a task, however, a few developed social relationships beyond their work; this group has more substantial contributions in the system over different tasks and projects. The boundary between this type of group and the community is more visible than the first type as their existence regarding their internal interactions, fulfilled projects and tasks is more apparent from an outsider perspective. Groups stayed active longer in the project, have multiple cores and the structure of the group has core, peripheral and extra peripheral, similar to the structure of social support network in WebMD studied by Introne et al. (2016).

Employing the group criteria by Arrow et al. (2000) helped to recognize different variations of group work that go beyond the individual assignment and goals. Since the model considers

cognitive, affective, and behavioral aspects of groups, the author could identify what a collection of individuals do to manifest a group. As the model guided the virtual ethnography, the data analysis revealed details of occasional group characteristics. For instance, the work is temporally delayed, groups have multiple cores, group membership has no criteria, and groups apply different techniques to work as a group despite not having an explicit identity in the organization.

Occasional groups have either moderate or high interconnectedness compared to the rest of the community. However, considering the light-weight and heavy-weight continuum by Haythornthwaite (2009) they fall in the middle for two reasons. First, they are not intensely connected as teams in heavy-weight communities such as a team of editors in Wikipedia. Second, there are still individual works and achievement by group members while they are working as a group occasionally. Occasional groups are more than light-weight communities as they create new knowledge through their interactions with each other. Creating new knowledge is more than what light-weight communities would produce. Figure 5.1 visualizes the notion of occasional group in the light-weight and heavy-weight continuum.

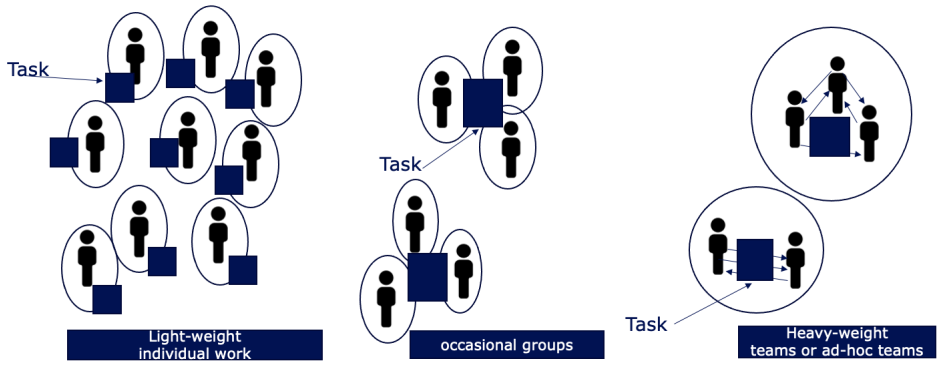

Fig. 5.1 The notion of occasional groups compared to light-weight and heavy-weight peer-production by Haythornthwaite (2009)

A similar concept to the notion of occasional groups is the intensional network introduced by Nardi et al. (2002). Nardi et al. (2002) explained the intensional network through investigation of cross-organizational collaboration. They found out that workers replaced the organizational settings and roles with the personal social network to accomplish a task. Intensional networks coexist with conventional teams while workers create their assemblage of people to

collaborate for a short or long period. Once workers perform joint work, they employ different techniques to maintain their network for future collaboration. The dynamic of an intensional network depends on what people are doing on a particular task at a specific time; subnets are active at different times.

While intensional networks emerge in teams working in an organization, occasional groups appear in settings where individuals rather than teams complete the work. The complexity of a task or a project brings individuals together, and they keep working together to accomplish a shared work. However, the emergence of "groupy" behavior is occasional compared to their individual contributions. Another difference between intensional network and occasional group is the group dynamic. The dynamics of groups are different from intensional networks. Workers in an intensional network are aware of a project early in their collaboration and have live subnets at a particular time to work on the project. However, occasional groups are not aware of a project, its goals, and phases before forming a group. Also, the number of group members increases depending on individuals' decision to join a shared task or project. As a result, the dynamic of occasional groups is temporally delayed, but for a different reason than intensional networks. Workers in an intensional network employ various communication techniques to maintain their network, and different subnets are salient at a particular time. However, the dynamic of an occasional group changes depending on who joins the group and works on what task.

Occasional groups share some similarities with ad hoc team. Ad hoc teams emerge to perform different projects or tasks depending on what is needed to be done in an organization(i.e., a crowdsourcing platform). The emergence nature of ad hoc teams is similar to occasional groups, but the activity period is more consistent compared to temporally protracted nature of occasional groups. Also, occasional groups still do not officially form a group or a team and their existence as a group are not recognized by the system.

In some crowdsourcing platforms, the system enables individuals to self-assemble their teams (Gómez-Zará et al. 2020). Self-assembled teams form a team by choosing any members they would like to work with (i.e., Wikipedia article authors) or perform an activity (i.e., virtual games). Some systems allow individuals to choose and invite their teammates as they have a

history of interactions on the platform. Some teams have more strict criteria than others, and the team composition and goals change as members may leave a team and join a new one. Whether team formation is well planned or spontaneously shaped, the entitativity of teams is visible in the system. Consequently, teams are supported while working on different projects.

Occasional groups are not forming a team and they are not recognized as a team. As a result, they do not have resources that would support team success within an organization. For example, occasional groups do not have clear directions and goals when they start working together. However, they managed to collaboratively accomplish various tasks and projects in the system. As the second study showed, groups of volunteers managed to find many possible new glitch classes. Although group contributions were sporadic over months compared to their individual contributions, they coordinated their actions to go through different analytical stages and created a new category proposal.

While categorization and tagging have been used interchangeably in some work settings, such as Wikipedia, other settings such as citizen science projects focus on categorization. Consequently, the act of tagging becomes a tool for data categorization and developing a new category. As a result, applying the categorization theory would better help to understand how either individuals or groups perform the work. Other concepts such as social tagging would not be still helpful as it focuses on tagging a subject rather than tagging and collecting multiple subjects and developing a new category that requires more intellectual work than naming and labeling. Applying the model by Grodal et al. (2020) helped to identify major analytical moves and explain what activities occasional groups had for each move to reach a common goal. The author could identify how occasional groups merge and relate categories to develop a new category by submitting a proposal to the science team. However, as the model was developed for individual works, further themes emerged from the data.

Occasional groups applied different techniques to bring awareness to the group and coordinate their actions at the early stage of their work. Awareness has a critical role to increase the visibility within a group (e.g., Dourish and Bellotti 1992, Gutwin and Greenberg 2002, Carroll et al. 2003, 2006). For instance, some group members informed each other about a new pattern by mentioning each other @user_login or names and reached an early agreement if the pattern

was a new one and if they all agreed to start using a new hashtag to mark more similar images. The member who noticed a new pattern decided to bring awareness among the group by calling their attention and explicitly coordinate their actions.

Not all groups, however, applied the same technique at the beginning of their work. For instance, one of the groups did not coordinate their actions at the early stage and as a result, one of the members used a different name to describe the same pattern that the other two members used. Lack of active awareness in this group allowed one of them to use a label that she thought would best describe the pattern, but once she noticed other members were using a different term that better described the pattern, she decided to use their label. Consequently, she edited the first label she had used for categorizing some images by adding a new line to what she had already posted. She also created a post and called other members' attention to acknowledge that their choice better explains the pattern, and she applies the same name for future similar images.

Another technique that all groups applied to better coordinate their actions was the use of shared outcome, a proposal. While teams working to develop a new tag in classification systems merge tags before announcing a new tag to the system, occasional groups used a different strategy to organize categories. Occasional groups used proposals as a shared outcome and a medium to coordinate future actions. The act of submitting a new proposal would inform all group members that the pattern they have discovered has specific boundaries and characteristics that make it different from current official classes. As a result, group members actively label more images using the name submitted in the proposal.

The glitch proposal is also a document for future work, a tool to coordinate future actions around categorization. As group members continue to identify more instances that resemble the new pattern, they also identify additional patterns emerging in data using new hashtags. They have post-proposal discussions to determine if these newer patterns are similar to what is already proposed in the system. As they realize the newer pattern resembled the proposed class, the moderator updates the proposal and adds the newer hashtags to the document. Consequently, it increases group awareness to know what categories have been merged. As shown in the second study, Sarah mentioned Oliver and Sherry to inform what new categories have

Table 5.1 Proposals submitted between in 2017 and 2018

Proposal Name	**Primary list of hashtags**	**Similar collections/hashtags added later**	**Date**
Spirograph	#loopholes #helix #anthill #scatteredlight	tapestry conv15 ETMY scattering Mont: Brushy Monet, 1sec+highscat organ pipes	11-03-2017
Cord and its segments	Cord: #cord #snake cord Caterpillar: #caterpillar , #rope Cord segment: #cord cord #lowfrequencysplatter	tangle wavy LF scratchy LF	11-23-2017
Sheaf	#sheaf, sheaf	brushes low Q, brushes, broom, wheat	11-23-2017
MicroHF	#microHF, #2000, #2000Hz 1000, 2000	MicroHF with 1080 Line, 1080 Line with underparallele, Shaker	01-05-2018
Forest	forest: #forest tuning forks: #tuningforks trees: #foil, #trees Logs: teddy bear face, #logs, #pinchers, Ghouls		04-08-2018
Zero signal	#zerosignal		04-22-2108
Falcon	#falcon		05-22-2018
Zip violin	#zip2secfalcon		05-22-2018
Shower towers	#showertowers, #showertower #shower-tower, #shower-towers #nota-shower_tower		07-29-2018
Sparkling	#sparkling, #kazam, #curtain		10-24-2018

been added to the spirograph proposal. As shown in tables 5.1 and 5.2, half of the proposals have a new list of hashtags or collections that have been added later.

The subsequent use of proposals was using them as a document for managing the community. After a proposal is submitted to the system, group members pick the preferred name and start using it more extensively on similar images to make the new term popular. As a result, individuals who do not interact with a group in the categorization process will be informed of what term has been chosen to label similar images. As one of the volunteers mentioned in their interview they pick the most popular term as they want to reduce the number of hashtags and

Table 5.2 Proposals submitted between in 2019

Proposal Name	**Primary list of hashtags**	**Similar collections/hashtags added later**	**Date**
Fireball	#ler13_fireball, #ligo-fireball		03-07-2019
Fly	#fly_er13		03-07-2019
Cordon bleu	#cordonbleu		03-07-2019
Centipede	#centipede		03-08-2019
Crown	#sparrow, #crown sparrows -crown subset: #crown-subset #dumpling		03-14-2019
Serpent	#serpent-er14	sparrows - serpent subset	03-20-2019
Hfb500	#hfb500er14, #ufo, #pizzicato	#jewel, #diamond	03-20-2019
High Frequency Burst	#hfb1500er14, #hfb_mini		03-20-2019
Slow parallels	#slow-parallels	Freddy	03-21-2019
Gnarly Whistles	#drizzle		04-03-2019
Pile	#pile	#bigpile	04-08-2019
70 Hz Line	#70Hz		04-21-2019
Vibration	#vibration, MF_clattering jolt		05-24-2019
Vibration lace	#vibration_lace	#vibration, #lace, #lacy	07-27-2019
Sparkling cloud	#sparklingcloud	#Gnarly #Sparkling	08-08-2019
Tealight	#tealight	gandalf-hat	09-18-2019
Hf-segments	#hf-spray	hf-segments	10-05-2019
Campfire	#campfire	brushes: ticking, bunching, brushes lowQ, brushes, wheat, greekfire	10-08-2019

make the system more organized.

Another group function, decision making, is mostly a smooth process for occasional groups. Members share instances of an image to elaborate on the boundaries of an image and how a particular name would describe it well. They may also share some scientific resources to speculate the causes that would help them better decisions about the glitch type. Members who share their knowledge by showing more examples or scientific resources help other members understand the pattern characteristics. As a result, they come to an agreement with no severe arguments. While members of occasional groups mostly come to agreement through sharing resources to reach a common ground for further analysis, there are instances that an advanced volunteer had a conflict a couple of core members and after a while left the community.

5.1.1 Implications of Occasional Groups

Recognizing occasional groups and their work helped us learn how knowledge creation results from group work even when the work setting is designed for individual assignments. The concept of occasional groups also helped to realize how groups apply different strategies to coordinate their actions and achieve a shared goal despite the temporally delayed communication and lack of supportive resources for group work. According to the findings, the identification of occasional groups is after they accomplish a shared activity. As a result, not all group members have a sense of group membership.

The concept of occasional groups applies to similar communities where the work is complicated enough and requires collective actions while group members still accomplish individual work. Communities like Wikipedia as a knowledge creation platform enable individual and team contributions. However, they encourage collaboration on articles to increase credibility through peer evaluation. While there are groups of editors that keep collaborating on several articles as a team, they still do not have an explicit identity like teams in FoldIt. However, their identity is more expressive than occasional groups. Occasional groups in such a platform are groups of editors whose cooperation is periodic compared to single edits but still do not have interconnectedness as teams of editors. Consequently, analyzing their behavior before consistent collaboration as a team would inform researchers in two ways. First, it helps to understand how

they can manage to create an article through periodic group work in what level of quality and what resources would help them to improve the quality of articles; depending on the quality of their shared outcome, each occasional group may require different resources to improve their work. Second, after analyzing the quality of mutual outcome, the system could recommend further collaboration to the same group members.

Another platform, Facebook, supports individual and group contributions for different purposes, such as political discourse. While groups on Facebook have a clear identity and shared space for group activities, such as exchanging information and holding events, occasional groups may merge in a different space, such as discussions on similar posts around a particular topic. Recognizing occasional groups would shed light on how each group would suggest a set of actions to solve a societal problem and how it is different from those who have an explicitly shared space as a group. Having a different space and level of consistency of contributions might affect their outcome and its quality.

The notion of occasional groups and their work could go beyond organizational or civil works. Occasional groups could emerge for a different reason, such as social support in online communities. Communities like Yahoo answers and Reddit provide social support in various areas such as health issues. Once individuals come together to support a patient about either mental or physical health issues, they form a group that would stay active periodically depending on the patient's needs. Recognizing a group of individuals who periodically provide support would benefit the community in different ways. For instance, once an occasional group helps patients with their social support, if the members become aware of their group work, they can employ different techniques (i.e., task assignment) to support more patients with similar issues.

Different stakeholders would benefit accordingly by understanding when occasional groups emerge and how they accomplish a shared task or a project. The group members would have an opportunity to deliberately continue working as a team and achieve more shared outcomes or keep their structure as is if they can not commit to the teamwork. The platform organizers would provide resources to better support them. Other members within the same community would benefit from their shared results and have a chance to join them for future work and accomplishments.

5.1.2 System Design Implications

Considering the existence of such groups in Gravity Spy, the system could provide different supports to facilitate their collaboration. As occasional groups do not have an identity like teams, the system do do not provide a shared space for their work. As a result, they have to employ different techniques to increase visibility of their work and coordinate their actions on multiple images spread in the system. While they already managed to work on images that matter to develop a proposal, they could still benefit from a shared space to work on a shared product (i.e., developing a proposal).

The platform could bring more visibility within a group and the community to maintain the current state of occasional groups. While a group of individuals is working on a new pattern, allocating a shared space by collecting all discussion threads about one topic would improve coordination and access to shared informational resources. The shared space would also serve the rest of the community in two ways. First, the group work could be used as learning resources for other volunteers, as mentioned earlier, the current structured training materials are focused only on regular classification tasks. Second, it may increase participation in existing groups or encourage volunteers to join future groups while working on new proposals.

Another strategy to increase visibility among group members is improving the use of shared tools. An example of shared tools is collections that is not well suited for group work. The current system does not notify volunteers when someone else has added them to their collections of images. Also, it does not order collections of individuals in a meaningful way. As a result, members working together may miss seeing a new collection on their page unless the other member notifies them explicitly. By increasing the visibility and order collections in individual pages chronologically, collaboration would be facilitated.

The platform organizers would also think about providing a different space for volunteers to work on their collections. The current organization does not inform members of an occasional group if a new image is added or a member leaves a comment on an image. A different organization could notify all members when a member shares their thoughts about an image. Consequently, the collaboration in such a shared space would improve and make the work more pleasant for such groups.

Another approach to support occasional groups is make team formation in the platform official and recognize a collective of individuals as a team. Consequently, there might be benefits if the system supported occasional groups. Like other citizen science projects supporting team work, Zooniverse could also support volunteers to form a team working on creative tasks and projects. Team work is not required when the analysis process is designed to be done by several independent volunteers and the system collects the majority vote as a final category for a particular item. However, team work might help individuals and the community when the analysis process does requires collaborations and different point of views and skills.

While the support for team formation would help occasional groups to establish their identity and better plan for a shared goal, it might hinder easy group membership. The current state of occasional group allows individuals join the group at any stage and have an input to the work. The trade off after team formation would be lack of diversity as teams keep working on several projects with the same members. One strategy would be promoting open teams that welcome new members at anytime in the project.

5.2 Limitations and Future Work

To expand the analysis of virtual and trace ethnography, the thesis could benefit from further analysis. First, the current analysis focused on successful groups. However, identifying occasional groups that did not succeed would provide further insights about the processes of works and obstacles they faced. Reasons such as conflict among members, not having enough information to develop a particular proposal, not having enough data to propose a new proposal would each provide a different direction to better understand characteristics of occasional groups and support their work.

Second, the log analysis would cover all the shared outcomes and their processes. The current analysis is based on some of the shared works that eventually do not provide insights about all the work practices and challenges over four years. The log analysis, however, would help to analyze all data if the system would better stored all the interactions. Since the current system is designed for individual work, the current log system does not record traces of collaboration among volunteers. While volunteers can collaborate on different collections, the log files do not

show who is collaborating with whom. Also, the log system does not capture who is replying to whom. As a result, the log data did not help design an algorithm to show all group interactions using different resources to either develop or not develop a proposal. Another limitation is how volunteers interact with each other. There are many threads of conversations when volunteers are talking together, but as they do not call each other or use the reply-to function, the reader does not realize if they are talking together unless they know the history of interactions and the topic conversation between volunteers.

Future work would apply computational methods to capture all interactions among volunteers around each possible new glitch class and identify threads of discussions as traces of a group working on a proposal. By conducting a similarity measure to accumulate all relevant discussion threads around a possible new glitch class and estimating a time threshold to collect individuals' comments on a discussion thread, the algorithm could create a new collection showing traces of group activities relevant to a possible new glitch class.

The result of such an algorithm could serve future research and a better design of the system. The outcome of the algorithm could be a new source of data for future research to show if structural, and compositional characteristics of such group change over time and if that affect their shared outcomes (i.e., shared collections, proposals).

The result also could inform a design of a new form of collection in Gravity Spy. Such a new collection would facilitate collaboration among a current group who already developed a proposal as they have access and discuss all images in one place and be aware of a new person joins the work and adds something interesting to the work. The new collection of group traces would also serve other community members to access all discussions around a proposal in one place for learning purposes.

5.3 Conclusion

The author examined and identified characteristics of occasional groups in a crowdsourcing platform by conducting virtual ethnography and employed "groupy" criteria defined by (Arrow et al. 2000). She also showed how two types of groups emerge around finding new classes of data depending on the intenconnectedness and longevity of groups. Recognizing the existence

of such groups would increase the visibility of their work in the community that is the first step to better support and facilitate their work in a system.

Then the author interviewed a handful of number of volunteers to show how they work around forming a new class of data and conducted another round of virtual and trace ethnography to understand the categorization process done by occasional groups. The author employed the framework by (Grodal et al. 2020) to identify analytical moves among volunteers who work together to develop a new class of data. Since occasional groups are not recognized as a group or team, members within a group applied different techniques to coordinate their actions and develop a new category together. Understanding how occasional groups accomplish a categorization project first implies their competency in fulfilling such a complex analysis. Second, it provides guidelines for the organization on how to support occasional groups in their current and future work.

While occasional groups managed their work and developed different techniques to find a new class of data, they could benefit from some platform changes. Most citizen science projects focused on data processing design tasks and assign a singular task to individuals; platform organization design and develop algorithms to collect and choose one vote out of many votes submitted to the system. However, once the task evolves and gets complicated, emergence of groups is inevitable in a system. Recognizing the groups and providing a shared space for easier collaboration facilitates their work. The current system allows individuals to collaborate on a collection, however, it does not provide required tools to facilitate the process of collaboration between two or more individuals working on a collection. By providing a shared space to discuss each instances of a collection occasional group members would manage the load of information and burden of finding relevant images that they already discussed would be lifted.

By recognizing occasional groups and their work, the system would bring more visibility to the process of categorization and developing a new class of data. Providing access to past and ongoing categorization processes would be an invaluable learning resource for volunteers.

The author expects to see a similar phenomenon in other crowdsourcing platforms such as Wikipedia, where individuals tend to create new content while forming a team is not required by the system. Recognizing a group of individuals who occasionally work and develop new

content (i.e., part or whole of an article) would encourage their work and provide options for team formation to work on further articles. Team formation is not required but it can be suggested to individuals who have a history of group work. As a result, individuals tackle complex problems by having organizational support for teamwork.

Appendices

Appendix A

Members of occasional groups

The following tables show members of each occasional group for developing a proposal.

Table A.1 Group members for each proposal submitted between in 2017 and 2018

Proposal name	**Group members**
Sparrow	P2, P8, P14, P19, P30, P37, P70, P86, P95, P105, P106, P111, P113
Cord and its segments	P2, P8, P14, P16, P17, P19, P21, P25, P30, P32, P37, P52 P70, P78, P86, P95, P105, P106, P107, P109, P110, P111, P113
Sheaf	P8, P14, P19, P30, P37, P88
MicroHF	P2, P10, P14, P19, P21, P22, P23, P24, P25, P29, P30, P32 P33, P37, P42, P47, P52, P54, P64, P68, P79, P80, P81, P83, P84, P89, P99, P104, P106, P109, P111, P112, P113
Forest	P2, P8, P14, P18, P19, P20, P22, P25, P27, P30, P32, P37, P39, P48, P52, P54, P56, P58, P61, P63, P64, P65, P72, P79, P80, P94, P102, P103, P109, P111, P113
Zero signal	P2, P14, P19, P30, P37, P50
Falcon	P2, P5, P7, P14, P19, P30, P32, P37, P51, P62, P106, P113
Zip violin	P5, P7, P14, P19, P30, P37, P62, P106
Shower towers	P19, P30, P48, P68, P79, P106, P113
Sparkling	P2, P14, P17, P19, P25, P30, P34, P37, P65, P66, P68, P69 P80, P102, P106, P109, P111, P113

Table A.2 Group members for each proposal submitted between in 2019

Proposal name	**Group members**
Fireball	P19, P30, P106, P113
Cordon bleu	P14, P19, P30, P106
Centipede	P19, P30, P113
Crown	P2, P4, P6, P11, P14, P15, P17, P18, P19, P20, P26, P27, P28, P30, P35, P37, P40, P41, P42, P45, P46, P53, P55, P62, P63, P65, P67, P68, P72, P75, P77, P79, P80, P86, P87, P91, P93, P94, P96, P101, P102, P106, P108, P109, P111, P113
Hfb500	P1, P2, P9, P12, P14, P19, P25, P30, P35, P37, P43, P44, P48, P49, P52, P56, P60, P63, P71, P72, P73, P76, P79, P82, P85, P90, P98, P104, P106, P107, P109, P110, P113
High Frequency Burst	P19, P30, P79, P113
Slow parallels	P30, P113
Gnarly Whistles	P19, P30, P31, P57, P80, P113
Pile	P14, P17, P19, P30, P37, P55, P58, P68, P113
70 Hz Line	P2, P14, P19, P30, P36, P37, P58, P68, P75, P106, P109, P111, P113
Vibration	P2, P8, P13, P14, P19, P24, P30, P38, P52, P58, P59, P90, P92, P100, P106, P111, P113
Vibration lace	P19, P30, P113
Sparkling cloud	P19, P30, P80
Tealight	P14, P19, P30, P113
Hf-segments	P19, P30, P113
Campfire	P3, P6, P12, P19, P30, P44, P60, P74, P97, P106, P113

Appendix B

Interview Questions

General Questions:

- If we hope to understand how new glitch class proposals emerge, what should we pay attention to in the process?
- How did the definition of a new class evolve?
- How did you decide on terminology? What other resources/methods/data did you draw on? E.g., a-logs
- How other volunteers were involved?

For volunteers who did not propose a new glitch class:

- What is the process of finding possible new glitch classes for you?
- How did you decide on terminology?
- What resources/methods/data did you draw on? E.g., a-logs
- Do you collaborate or work with moderators or other volunteers when you are finding a new possible glitch class?
- If yes, how do you collaborate/work with them?

Work Process:

- Do people share specific work processes that you need to know or share with each other or have adapted from science team members?

- To what extent is this process a collaborative work?
- How do they collaborate with others?
- Why do they collaborate with other volunteers?
- Do other volunteers in the process have an implicit role?
- What is the process of creating a specific collection/labeling when there are different opinions on similarities to different classes?
- How would they decide a possible new glitch class is similar to which official class?
- Is it more an individual-based decision or more reaching a consensus with others?
- If the latter, what does a typical decision-making process look like?
- Changes over time: Does the importance of these features change over time?

REFERENCES

Albrechtsen, H. and Jacob, E. K. (1998), 'The dynamics of classification systems as boundary objects for cooperation in the electronic library'.

Allport, F. H. (1962), 'A structuronomic conception of behavior: Individual and collective: I. structural theory and the master problem of social psychology.', *The Journal of Abnormal and Social Psychology* **64**(1), 3.

Arrow, H., McGrath, J. E. and Berdahl, J. L. (2000), *Small groups as complex systems: Formation, coordination, development, and adaptation*, Sage Publications.

Ashby, F. G. and Maddox, W. T. (2005), 'Human category learning', *Annu. Rev. Psychol.* **56**, 149–178.

Barnard, C. I. (2004), *Organization and management: Selected papers*, Routledge.

Bechky, B. A. (2006), 'Talking about machines, thick description, and knowledge work', *Organization Studies* **27**(12), 1757–1768.

Bennett, J., Lanning, S. et al. (2007), The netflix prize, *in* 'Proceedings of KDD cup and workshop', Vol. 2007, New York, NY, USA., p. 35.

Berkowitz, M. W. and Gibbs, J. C. (1983), 'Measuring the developmental features of moral discussion', *Merrill-Palmer Quarterly (1982-)* pp. 399–410.

Bowker, G. C. and Star, S. L. (2000), *Sorting things out: Classification and its consequences*, MIT press.

Brooks, L., Rosch, E. and Lloyd, B. (1978), 'Cognition and categorization'.

Bullard, J. A. (2017), Classification design: understanding the decisions between theory and consequence, PhD thesis.

Campbell, D. T. (1958), 'Common fate, similarity, and other indices of the status of aggregates of persons as social entities', *Behavioral science* **3**(1), 14–25.

Campbell, D. T. (1975), 'On the conflicts between biological and social evolution and between psychology and moral tradition.', *American psychologist* **30**(12), 1103.

Carroll, J. M., Neale, D. C., Isenhour, P. L., Rosson, M. B. and McCrickard, D. S. (2003), 'Notification and awareness: synchronizing task-oriented collaborative activity', *International Journal of Human-Computer Studies* **58**(5), 605–632.

Carroll, J. M., Rosson, M. B., Convertino, G. and Ganoe, C. H. (2006), 'Awareness and teamwork in computer-supported collaborations', *Interacting with computers* **18**(1), 21–46.

Christensen, L. R. (2013), 'Stigmergy in human practice: Coordination in construction work', *Cognitive Systems Research* **21**, 40–51.

Clark, H. H. and Brennan, S. E. (1991), 'Grounding in communication.'.

Clarke, R. I. (2021), 'Library classification systems in the us: Basic ideas and examples', *Cataloging and Classification Quarterly* pp. 1–14.

Cohen, B. and Murphy, G. L. (1984), 'Models of concepts', *Cognitive science* **8**(1), 27–58.

Cramton, C. D. (2001), 'The mutual knowledge problem and its consequences for dispersed collaboration', *Organization science* **12**(3), 346–371.

Dai, P., Lin, C. H., Weld, D. S. et al. (2013), 'Pomdp-based control of workflows for crowdsourcing', *Artificial Intelligence* **202**, 52–85.

Dascalu, M.-I., Bodea, C.-N., Lytras, M., De Pablos, P. O. and Burlacu, A. (2014), 'Improving e-learning communities through optimal composition of multidisciplinary learning groups', *Computers in Human Behavior* **30**, 362–371.

DiMaggio, P. (1991), Social structure, institutions, and cultural goods: The case of the us, *in* 'Social theory for a changing society', Westview Press.

Dissanayake, I., Zhang, J. and Gu, B. (2015), 'Task division for team success in crowdsourcing contests: Resource allocation and alignment effects', *Journal of Management Information Systems* **32**(2), 8–39.

Douglas, M. (1986), *How institutions think*, Syracuse University Press.

Dourish, P. and Bellotti, V. (1992), Awareness and coordination in shared workspaces, *in* 'Proceedings of the 1992 ACM conference on Computer-supported cooperative work', pp. 107–114.

Duquenne, V., Mohr, J. and Le Pape, A. (1998), 'Comparison of dual orderings in time', *Social science information* **37**(2), 227–253.

Dynes, R. R. and Aguirre, B. E. (1979), 'Organizational adaptation to crises: Mechanisms of coordination and structural change', *Disasters* **3**(1), 71–74.

Estellés-Arolas, E. and González-Ladrón-De-Guevara, F. (2012), 'Towards an integrated crowdsourcing definition', *Journal of Information science* **38**(2), 189–200.

Forsyth, D. R. and Elliott, T. R. (1999), 'Group dynamics and psychological well-being: The impact of groups on adjustment and dysfunction.'.

Geiger, D., Rosemann, M., Fielt, E. and Schader, M. (2012), 'Crowdsourcing information systems-definition typology, and design'.

Geiger, R. S. and Ribes, D. (2011), Trace ethnography: Following coordination through documentary practices, *in* '2011 44th Hawaii International Conference on System Sciences', IEEE, pp. 1–10.

Gilson, L. L., Maynard, M. T., Jones Young, N. C., Vartiainen, M. and Hakonen, M. (2015), 'Virtual teams research: 10 years, 10 themes, and 10 opportunities', *Journal of management* **41**(5), 1313–1337.

Golder, S. A. and Huberman, B. A. (2006), 'Usage patterns of collaborative tagging systems', *Journal of information science* **32**(2), 198–208.

Gómez-Zará, D., DeChurch, L. A. and Contractor, N. S. (2020), 'A taxonomy of team-assembly systems: Understanding how people use technologies to form teams', *Proceedings of the ACM on Human-Computer Interaction* **4**(CSCW2), 1–36.

Grodal, S., Anteby, M. and Holm, A. L. (2020), 'Achieving rigor in qualitative analysis: The role of active categorization in theory building', *Academy of Management Review* (ja).

Gutwin, C. and Greenberg, S. (2002), 'A descriptive framework of workspace awareness for real-time groupware', *Computer Supported Cooperative Work (CSCW)* **11**(3), 411–446.

Hackman, J. (1987), 'The design of work teams. in jw lorsch (ed.), handbook of organizational behavior, englewood cliffs, nj: Prentice-hall'.

Hackman, J. R. and Katz, N. (2010), 'Group behavior and performance.'.

Hampton, J. (1988), *Hobbes and the social contract tradition*, Cambridge University Press.

Hayes, N. (1997), *Successful team management*, Thomson Learning Emea.

Haythornthwaite, C. (2009), Crowds and communities: Light and heavyweight models of peer production, *in* '2009 42nd Hawaii international conference on system sciences', IEEE, pp. 1–10.

Henry, K. B., Arrow, H. and Carini, B. (1999), 'A tripartite model of group identification: Theory and measurement', *Small group research* **30**(5), 558–581.

Hertel, G. and Hüffmeier, J. (2011), 'Many cheers make light the work: How social support triggers process gains in teams', *Journal of Managerial Psychology* .

Hinds, P., Liu, L. and Lyon, J. (2011), 'Putting the global in global work: An intercultural lens on the practice of cross-national collaboration', *Academy of Management annals* **5**(1), 135–188.

Hine, C. (2000), *Virtual ethnography*, Sage.

Hollander, E. P. and Willis, R. H. (1967), 'Some current issues in the psychology of conformity and nonconformity.', *Psychological Bulletin* **68**(1), 62.

Huang, J., Hmelo-Silver, C. E., Jordan, R., Gray, S., Frensley, T., Newman, G. and Stern, M. J. (2018), 'Scientific discourse of citizen scientists: Models as a boundary object for collaborative problem solving', *Computers in Human Behavior* **87**, 480–492.

Huckman, R. S., Staats, B. R. and Upton, D. M. (2009), 'Team familiarity, role experience, and performance: Evidence from indian software services', *Management science* **55**(1), 85–100.

Ilgen, D. R., Hollenbeck, J. R., Johnson, M. and Jundt, D. (2005), 'Teams in organizations: From input-process-output models to imoi models', *Annu. Rev. Psychol.* **56**, 517–543.

Introne, J., Semaan, B. and Goggins, S. (2016), A sociotechnical mechanism for online support provision, *in* 'Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems', ACM, pp. 3559–3571.

Jackson, C., Crowston, K., Østerlund, C. and Harandi, M. (2018), 'Folksonomies to support coordination and coordination of folksonomies', *Computer Supported Cooperative Work (CSCW)* **27**(3-6), 647–678.

Janicik, G. A. and Bartel, C. A. (2003), 'Talking about time: Effects of temporal planning and time awareness norms on group coordination and performance.', *Group Dynamics: Theory, Research, and Practice* **7**(2), 122.

Johnson, D. W. and Johnson, R. T. (1987), *Learning together and alone: Cooperative, competitive, and individualistic learning*, Prentice-Hall, Inc.

Johnson, K. E. and Mervis, C. B. (1997), 'Effects of varying levels of expertise on the basic level of categorization.', *Journal of Experimental Psychology: General* **126**(3), 248.

Joyce, E., Ozturk, P. and Pike, J. (2021), Social capital in online temporary organizations: Addressing critical, complex tasks through deliberation, *in* 'Proceedings of the 54th Hawaii International Conference on System Sciences', p. 306.

Katzenbach, J. R. and Smith, D. (1993), 'K.(1993)', *The wisdom of teams* .

Kerr, N. L. and Tindale, R. S. (2004), 'Group performance and decision making', *Annu. Rev. Psychol.* **55**, 623–655.

Kou, Y. and Gui, X. (2014), Playing with strangers: understanding temporary teams in league of legends, *in* 'Proceedings of the first ACM SIGCHI annual symposium on Computer-human interaction in play', ACM, pp. 161–169.

Kozlowski, S. W. and Ilgen, D. R. (2006), 'Enhancing the effectiveness of work groups and teams', *Psychological science in the public interest* **7**(3), 77–124.

Kroski, E. (2005), 'The hive mind: Folksonomies and user-based tagging', *InfoTangle Blog* .

Larson, C. E., Larson, C. and LaFasto, F. M. (1989), *Teamwork: What must go right/what can go wrong*, Vol. 10, Sage.

Leach, C. W., Van Zomeren, M., Zebel, S., Vliek, M. L., Pennekamp, S. F., Doosje, B., Ouwerkerk, J. W. and Spears, R. (2008), 'Group-level self-definition and self-investment: a hierarchical (multicomponent) model of in-group identification.', *Journal of personality and social psychology* **95**(1), 144.

Levi, D. (2017), *Group dynamics for teams*, Sage Publications.

Lewin, K. (1948), 'Resolving social conflicts; selected papers on group dynamics.'.

Lin, C. H., Mausam, M. and Weld, D. S. (2012), Dynamically switching between synergistic workflows for crowdsourcing, *in* 'Twenty-Sixth AAAI Conference on Artificial Intelligence'.

Marks, M. A., Mathieu, J. E. and Zaccaro, S. J. (2001), 'A temporally based framework and taxonomy of team processes', *Academy of management review* **26**(3), 356–376.

Mathieu, J. E., Gallagher, P. T., Domingo, M. A. and Klock, E. A. (2019), 'Embracing complexity: Reviewing the past decade of team effectiveness research', *Annual Review of Organizational Psychology and Organizational Behavior* **6**, 17–46.

Mathieu, J. E., Hollenbeck, J. R., van Knippenberg, D. and Ilgen, D. R. (2017), 'A century of work teams in the journal of applied psychology.', *Journal of applied psychology* **102**(3), 452.

McGrath, J. E. (1984), *Groups: Interaction and performance*, Vol. 14, Prentice-Hall Englewood Cliffs, NJ.

McGrath, J. E. (1991), 'Time, interaction, and performance (tip) a theory of groups', *Small group research* **22**(2), 147–174.

Mohrman, S. A. and Mohrman, A. M. (1993), *Organizational change and learning*, Center for Effective Organizations, School of Business Administration ....

Nardi, B. A., Whittaker, S. and Schwarz, H. (2002), 'Networkers and their activity in intensional networks', *Computer Supported Cooperative Work (CSCW)* **11**(1), 205–242.

Newcomb, T. M. (1961), 'The acquaintance process as a prototype of human interaction.'.

Newell, B. R., Dunn, J. C. and Kalish, M. (2011), Systems of category learning: Fact or fantasy?, *in* 'Psychology of learning and motivation', Vol. 54, Elsevier, pp. 167–215.

Okhuysen, G. A. and Waller, M. J. (2002), 'Focusing on midpoint transitions: An analysis of boundary conditions', *Academy of Management Journal* **45**(5), 1056–1065.

Olson, J. (2000), 'Distance matters', *Human-Computer Interaction* **15**.

Østerlund, C., Crowston, K. and Jackson, C. (2020), 'Building an apparatus: Refractive, reflective, and diffractive readings of trace data', *Journal of the Association for Information Systems* **21**(1), 10.

O'Rourke, J. (2011), *How to fold it: the mathematics of linkages, origami, and polyhedra*, Cambridge University Press.

Pachucki, M. A. and Breiger, R. L. (2010), 'Cultural holes: Beyond relationality in social networks and culture', *Annual review of sociology* **36**, 205–224.

Peters, I. and Weller, K. (2008), 'Tag gardening for folksonomy enrichment and maintenance', *Webology* **5**(3), 1–18.

Piccoli, G., Powell, A. and Ives, B. (2004), 'Virtual teams: team control structure, work processes, and team effectiveness', *Information Technology and People* **17**(4), 359–379.

Powell, A., Piccoli, G. and Ives, B. (2004), 'Virtual teams: a review of current literature and directions for future research', *ACM SIGMIS Database: the DATABASE for Advances in Information Systems* **35**(1), 6–36.

Riedl, C. and Woolley, A. W. (2017), 'Teams vs. crowds: A field test of the relative contribution of incentives, member ability, and emergent collaboration to crowd-based problem solving performance', *Academy of Management Discoveries* **3**(4), 382–403.

Rosch, E. (1978), 'Principles of categorization [ ]/eleonora rosch'.

Rosch, E. (1999), 'Reclaiming concepts', *Journal of consciousness studies* **6**(11-12), 61–77.

Rughiniş, C. (2016), Citizen science, gallaxies and tropes: Knowledge creation in impromptu crowd science movements, *in* '2016 15th RoEduNet Conference: Networking in Education and Research', IEEE, pp. 1–6.

Salehi, N., Irani, L. C., Bernstein, M. S., Alkhatib, A., Ogbe, E., Milland, K. et al. (2015), We are dynamo: Overcoming stalling and friction in collective action for crowd workers, *in* 'Proceedings of the 33rd annual ACM conference on human factors in computing systems', ACM, pp. 1621–1630.

Simmel, G. (1950), *The sociology of georg simmel*, Vol. 92892, Simon and Schuster.

Soergel, D. (1976), 'Is user satisfaction a hobgoblin?', *Journal of the American Society for Information Science* **27**(4), 256–259.

Spalding, T. L. and Murphy, G. L. (1996), 'Effects of background knowledge on category construction.', *Journal of Experimental Psychology: Learning, Memory, and Cognition* **22**(2), 525.

Steiner, I. D. (1972), *Group processes and productivity*, Academic press.

Sturtz, D. N. (2004), 'Communal categorization: the folksonomy', *INFO622: Content Representation* **16**.

Sundstrom, E., De Meuse, K. P. and Futrell, D. (1990), 'Work teams: Applications and effectiveness.', *American psychologist* **45**(2), 120.

Sundstrom, E., McIntyre, M., Halfhill, T. and Richards, H. (2000), 'Work groups: From the hawthorne studies to work teams of the 1990s and beyond.', *Group Dynamics: Theory, Research, and Practice* **4**(1), 44.

Thompson, E. P. (1967), 'Time, work-discipline, and industrial capitalism', *Past and present* (38), 56–97.

Thornton, K. and McDonald, D. W. (2012), Tagging wikipedia: collaboratively creating a category system, *in* 'Proceedings of the 17th ACM international conference on Supporting group work', pp. 219–228.

Trant, J. (2009), 'Studying social tagging and folksonomy: A review and framework'.

Valentine, M. A., Retelny, D., To, A., Rahmati, N., Doshi, T. and Bernstein, M. S. (2017), Flash organizations: Crowdsourcing complex work by structuring crowds as organizations, *in* 'Proceedings of the 2017 CHI conference on human factors in computing systems', pp. 3523–3537.

Warkentin, M. E., Sayeed, L. and Hightower, R. (1997), 'Virtual teams versus face-to-face teams: an exploratory study of a web-based conference system', *Decision sciences* **28**(4), 975–996.

Weick, K. E. (1979), 'The social psychology of organizing. addison', *Reading, MA* .

Wen, M., Maki, K., Dow, S., Herbsleb, J. D. and Rose, C. (2017), 'Supporting virtual team formation through community-wide deliberation', *Proceedings of the ACM on Human-Computer Interaction* **1**(CSCW), 109.

Wiggins, A. and Crowston, K. (2010), 'Developing a conceptual model of virtual organizations for citizen science', *International Journal of Organisational Design and Engineering* **1**(1/2), 148.

Wilder, D. and Simon, A. F. (1998), 'Groups: Implications for social perception and intergroup behavior', *Intergroup cognition and intergroup behavior* **27**.

Woolley, A. W., Chabris, C. F., Pentland, A., Hashmi, N. and Malone, T. W. (2010), 'Evidence for a collective intelligence factor in the performance of human groups', *science* **330**(6004), 686–688.

Young, J. L. and Mandelstam, Y. (2013), 'It takes a village: Developing library of congress genre/form terms', *Cataloging and Classification Quarterly* **51**(1-3), 6–24.

Zellmer-Bruhn, M., Waller, M. J. and Ancona, D. (2004), 'The effect of temporal entrainment on the ability of teams to change their routines', *Research on managing groups and teams* **6**, 135–158.

Zevin, M., Coughlin, S., Bahaadini, S., Besler, E., Rohani, N., Allen, S., Cabero, M., Crowston, K., Katsaggelos, A. K., Larson, S. L. et al. (2017), 'Gravity spy: integrating advanced ligo detector characterization, machine learning, and citizen science', *Classical and quantum gravity* **34**(6), 064003.

Ziller, R. C. (1964), 'Individuation and socialization: A theory of assimilation in large organizations', *Human relations* **17**(4), 341–360.

Mahboobeh Harandi

www.mharandi.com mharandi@syr.edu Google Scholar Github LinkedIn

Research Interests

Focus: Studying collective behaviors and expectations in online communities AI-augmented

Methods: A mixed methods approach to collect data, including online trace data, interviews, focus groups, and virtual ethnography. Data Analysis through NLP techniques, conducting parametric and non-parametric statistical tests, and thematic analysis.

Keywords: Computer Supported Cooperative Work, Social Computing

EDUCATION

Syracuse University, Syracuse, NY, USA

Ph.D., Information Science and Technology, 2021

Thesis Title: Occasional Groups in Crowdsourcing Platforms

Norwegian University of Science and Technology, Trondheim, Norway

MScEng, Information Systems Engineering, 2015

Thesis Title: Sentiment Analysis in Political News in News Recommender Systems

University of Science and Culture, Tehran, Iran

B.S., Software Engineering, 2005

Research and Industry Experience

Syracuse University, Syracuse, NY, USA

Research Assistant, Jan. 2016 - July 2021

Facebook Inc., Menlo Park, CA, USA

UX Research Intern, Jun. 2020 - Aug. 2020

Syracuse University, Syracuse, NY, USA

Visiting Scholar, Feb. 2015 - Aug. 2015

Norwegian University of Science and Technology, Department of Psychology, Trondheim, Norway

Student Researcher, Software Engineer, Oct. 2012 - May 2014

Compuco Inter S.A. Geneva Switzerland, Tehran, Iran

Software Engineer, Dec 2011 – July 2012

Sayeh Saman Food Distribution Co., Tehran, Iran

Software Engineer, June 2008 – December 2008

Electronic Computing Services Corporation, Tehran, Iran

Software Engineer, March 2007 – May 2008

Publications

Journal Papers

[J.5] Jackson, C., Østerlund, C., Crowston, K., **Harandi, M.**, Allen, S., Bahaadini, S., Coughlin, S., Kalogera, V., Katsaggelos, A., Larson, S., Rohani, N., Smith, J., Trouille, L., and Zevin, M. (2019). Teaching citizen scientists to categorize glitches using machine learning guided training. Computers in Human Behavior.

[J.4] Crowston, K., Østerlund, C., Lee, T.K., Jackson, C., **Harandi, M.**, Allen, S., Bahaadini, S., Coughlin, S., Katsaggelos, A., Larson, S. and Rohani, N. (2019). Knowledge Tracing to Model Learning in Online Citizen Science Projects. IEEE Transactions on Learning Technologies.

[J.3] Coughlin, S., Bahaadini, S., Rohani, N., Zevin, M., Patane, O., **Harandi, M.**, Jackson, C., Noroozi, V., Allen, S., Areeda, J. and Coughlin, M., (2019). Classifying the Unknown: Discovering Novel Gravitational-wave Detector Glitches Using Similarity Learning. Physical Review D, 99(8), 082002.

[J.2] Jackson, C., Crowston, K., Østerlund, C., and **Harandi, M.** (2018). Folksonomies to Support Coordination and Coordination of Folksonomies. Computer Supported Cooperative Work (CSCW), 1-32. Received an **honorable mention** in the 2019 ”Best Paper Award” competition of the Association for Information Science and Technology (ASISandT).

[J.1] Lee, T. K., Crowston, K., **Harandi, M.**, Østerlund, C., and Miller, G. (2018). Appealing to Different Motivations in a Message to Recruit Citizen Scientists: Results of a Field Experiment. JCOM: Journal of Science Communication, 17(1), A1-A1.

Peer-Reviewed Conference Proceedings

[C.5] Attari, S., Scull, CH., and **Harandi, M.** (2021). Leveraging Speculative Design to Re-Imagine Product Roadmaps. In *Ethnographic Praxis in Industry Conference Proceedings*.

[C.4] Jackson, C. B., Østerlund, C., Crowston, K., **Harandi, M.**, and Trouille, L. Shifting forms of Engagement: Volunteer Learning in Online Citizen Science. In *Proceeding of Computer Supported Cooperative Work and Social Computing (CSCW). Received an* ***honorable mention*** *in the 2019 "Best Paper Award" competition.*

[C.3] **Harandi, M.**, Crowston, K., Jackson, C., and Østerlund, C. (2020). The Genie in the Bottle: Different Stakeholders, Different Interpretations of Machine Learning. In *Proceeding of 52nd Hawaii International Conference on System Science.*

[C.2] Jackson, C. B., Østerlund, C., **Harandi, M.**, Kharwar, D., and Crowston, K. (2019), Linguistic Changes in Online Citizen Science: A Structurational Perspective. In International Conference on Information Systems (ICIS).

[C.1] **Harandi, M.** and Gulla, J. A. (2015, September). Survey of User Profiling in News Recommender Systems. Third International Workshop on News Recommendation and Analytics In *Conjunction with the 9th ACM Recommender Systems Conference (2015),* (pp. 20-26).

Peer-Reviewed Extended Abstracts

[P.4] **Harandi, M.**, Supporting Occasional Groups in Crowdsourcing Platforms, Doctoral Consortium In Companion of the 2019 ACM Conference on Computer Supported Cooperative Work and Social Computing.

[P.3] **Harandi, M.**, Jackson, C. B., Østerlund, C., and Crowston, K. (2018, October). Talking the Talk in Citizen Science. In Companion of the 2018 ACM Conference on Computer Supported Cooperative Work and Social Computing (pp. 309-312). ACM.

[P.2] **Harandi. M.**, Yu, B., and Gulla, J. (2016), Extracting and Presenting Different Viewpoints from Political News Articles, iConference 2016 Proceedings

[P.1] Khosravifard, M., **Harandi, M.**, van der Meer, A., and van der Weel, R. (2015). Perceptual Learning through Brain Computer Interface in Young Infants. In Studies in Perception and Action XIII: Eighteenth International Conference on Perception and Action

(pp. 141), *Psychology Press*.

Teaching Experience

Syracuse University, School of Information Studies, Syracuse, NY

Research Methods in Information Studies, Jan. 2021 - July 2021

Syracuse University, School of Information Studies, Syracuse, NY

Introduction to Applied Data Science, Jan. 2020 - May 2021

Syracuse University, School of Information Studies, Syracuse, NY

Sentiment Analysis in Social Media, Guest Lecturer, Nov. 2018

Syracuse University, School of Information Studies, Syracuse, NY

Pros and Cons of Social Media, Guest Lecturer, Oct. 2018

Syracuse University, School of Information Studies, Syracuse, NY

Chatbots and Natural Language Processing, Guest Lecturer, Oct. 2017

Patris House of Culture Institute, Tehran, Iran

front-end Web Development and Microsoft Office, Apr. 2010 - Nov. 2011

Grants

Syracuse University, School of Information Studies, Syracuse, NY, Mar. 2016 - Nov. 2019

- Doctoral Consortium In Companion of the 2019 ACM Conference CSCW ($2000)
- Syracuse University iSchool Doctoral Summer Research Grant ($6000)
- Syracuse University iSchool Doctoral Travel Grant ($1000)
- Syracuse University iSchool Doctoral Summer Research Support ($3000)
- Syracuse University iSchool Doctoral Travel Grant ($850)
- Syracuse University iSchool Doctoral Summer Research Grant ($6000)
- Syracuse University Katzer Doctoral Summer Research Support ($6000)
- Syracuse University iSchool Doctoral Travel Grant ($850)

Certificates

New England Complex Systems Institute **Cambridge, Massachusetts**

The Integration of Human and Artificial Intelligence *February 2018*

Johannes Kepler University **Linz, Austria**

Subject-driven Role-guided Externalization of Organizational Models *July 2014*

Kish Institute of Science and Technology **Tehran, Iran**

Advanced English *July 2010*

Invited Talks and Presentations

- Poster Presenter, Supporting Occasional Groups in Online Crowdsourcing Platforms, Doctoral Consortium In Companion of the 2019 ACM Conference on **Computer Supported Cooperative Work and Social Computing**
  Austin, TX - 2019

- Speaker, Future of Work in Virtual teams, iSchool Graduate Seminar: Organization of Tomorrow
  Syracuse, NY – 2019

- Poster Presenter, Occasional Groups in Online Communities, **WAIM Conference**
  NYC, NY – 2019

- Poster Presenter, Group learning supported by a conversational agent in online communities, **HCIC Conference**
  Watsonville, CA – 2019

- Speaker, Humans, Machines, and the Future of Citizen Science, iSchool Graduate Seminar
  Syracuse, NY – 2019

- Poster Presenter, Talking the Talk in Citizen Science, **CSCW Conference**

  Jersey City, NY – 2018

- Poster Presenter, Organizing Data in Dynamic and Flexible Tagging System, **ICCS Conference**

  Cambridge, MA – 2018

- Poster Presenter, Extracting and Presenting Different Viewpoints from News Articles, **iConference**

  Philadelphia, PA – 2016

- Paper Presenter, Survery of User Profiling in News Recommender Systems, **RecSys Conference**

  Vienna, Austria – 2015

Service

ACM Conference on Computer Supported Cooperative Work and Social Computing, Austin, TX

Student Volunteer, Nov.2019

- Supporting running different sessions and events at the conference

Syracuse University, School of Information Studies, Syracuse, NY

Reviewer

- Conference on Computer Supported Cooperative Work and Social Computing (CSCW) – 2020
- Conference on International Conference on Information Systems (ICIS) – 2020
- Conference on Computer Supported Cooperative Work and Social Computing (CSCW) – 2020
- Conference on Human Factors in Computing Systems (CHI) – 2020
- European Conference on Information Systems (ECIS) – 2020

- Journal of Electronic Markets – 2020
- Hawaii International Conference on System Sciences (HICSS) – 2020
- ACM Group Conference, 3 papers – 2020
- European Conference on Information Systems (ECIS) – 2019
- Hawaii International Conference on System Sciences (HICSS) – 2019
- European Conference on Information Systems (ECIS) – 2018
- ACM Computer Human Interaction (CHI) – 2017

Syracuse University, School of Information Studies, Syracuse, NY

Organizer, Oct.2018

- Organized the 1st Research Day at School of Information Studies, 44 research projects showcased

Syracuse University, School of Information Studies, Syracuse, NY

Committee member, Sep.2107-May2020

- Serving as a doctoral representative in department committee in hiring faculty search
- Served as a doctoral representative in department committees: faculty evaluation, research, doctoral admission

Syracuse University, School of Information Studies, Syracuse, NY

Mentor, Nov.2016-Dec.2019

- Guided two graduate students for text mining and social network analysis
- Helped four graduate students for applying different NLP and data mining techniques in four research projects
- Taught one undergrad student for content analysis

Press

- Innovations, School of Information Studies News, Syracuse University, Sharing the Wealth: iSchool Researchers Showcase the Breadth of Their investigations, 2019
- Connections, School of Information Studies News, Syracuse University, ”44 Projects Showcased at iSchool’s First Research Day”, Spring/Summer 2019, VOL. 18, NO.1
- School of Information Studies News, Syracuse University, iSchool Team Wins ASIST Social Informatics “Best Paper” Honors, 2019
- School of Information Studies News, Syracuse University, ”44 Projects Showcased at iSchool’s First Research Day”, 2018
- The Official Blog of the Syracuse University iSchool, ”Once an It Girl, Always an It Girl: 2017 It Girls Overnight Retreat”, 2017
- Daily Orange, ”Syracuse University researchers continue to further gravitational wave research”, 2016
- Symmetry Magazine, “Citizen scientists join search for gravitational waves”, 2016