Chapter 2
Implied Role Mining

2.1 Introduction

In an implied role analysis, a researcher defines the set of roles that users of a system are expected to exhibit *before* extensive data or structural analysis commences. It is a more kind of qualitative, rather than quantitative, way to evaluate and extract the latent roles in a system. Based on the analysts perception of a system's context, observed interactions, and qualitative analysis of network positions, the definitions of the roles are further refined. Once the roles are established, users are assigned to roles according to a dataset of information about a system. This assignment is often done heuristically with rules defined directly by the analyst.

In some contexts one may think that implied role analysis is not a great idea. This is because the roles defined and the assignment of users to roles are solely based on analyst-defined definitions of the roles that should exist in a system. But to do this accurately, the analyst must fully and precisely understand the behaviors and responsibilities of system entities. For example, consider an analyst who is studying data about friendships among children in a school. She is likely to rely on her preconceived notions of how and why friendships emerge among school children to define an initial set of roles in the system, prior to any kind of data analysis. But a different analyst, with a separate set of notions about friendship formations in schools, may define a completely different set of roles according to different hypotheses about how and why friendships emerge. Because of this challenge, implied role mining is not a suitable approach for systems where entities can interact and develop relationships in an innumerable number of ways. This often occurs in "offline" social contexts, where social relations emerge as a result of face-to-face interactions with others. Implicit social cues or norms that a group follows, the context under which the face-to-face interactions are held in, and the environment or location of interactions are but some of the difficult to account for aspects that may drive one person to interact with another.

However, constraints on the way people can interact and communicate with others exist on *online social systems* [1, 2]. An online social system is an online service that

integrates mechanisms for users to directly interact, participate, communicate, or respond to according to the actions of another. Any kind of social media service (e.g. Twitter, Instagram, Snapchat, Weibo) are kinds of online social services. For example, users connect or follow one another if they are interested in the information they post or if they have an offline relationship, users interact with each other by responding to posts made by others, and users participate in widespread conversations by tagging their posts with topical hashtags. Less obvious examples of offline social systems are collaborative news sites (e.g. Reddit or Digg) where users interact with each other by supporting (upvoting) content by others and by discussing articles that were posted, and an e-commerce site where users can rate products, post reviews, and answer questions about a product that others have posted. Older but still relevant examples of online social systems are IRC channels whereby users share information and content with each other in chat rooms, and online forums where users participate in conversations under a subject thread within a forum board.

Settings where the ways in which interactions occur follow strict rules and other constraints make it easier for an analyst to formulate hypotheses about how and why social relations emerge. This is especially true in online social systems because: (i) the context under which interactions are performed in is well defined by the functionality of the social system; (ii) the limited modes of communication define specific interaction mechanisms that limits implicit social cues; and (iii) interactions are mediated through an interface that limits expressiveness, rather than in a face-to-face setting where non-verbal cues can be transmitted. Thus, while implied role mining carries some disadvantages, researchers have adopted the practice to unearth and understand the roles of users on online social systems and the interactions among them.

In an online social system, the role of a user is often defined by the functionalities of the system and the ways it allows users to interact with each other. As a concrete example, consider an online forum consisting of a number of "boards", which are collections of discussions about a broad topic. Discussions within a board are organized into "threads", with a thread containing a discussion about a specific subject. One role on the system may be a *moderator*, that is, someone who is responsible for making sure the conversations in the threads of a board stay on topic, and when inappropriate content is posted, deleting it and possibly reprimanding a user. An *administrator* may be responsible for promoting board members to be moderators and decide if users should be blocked or banned from board access. A *lurker* may be a user who only views conversations without ever contributing to a conversation, and a *respected user* may have a large post count with 'upvoted' or positively rated comments in a thread. Further kinds of roles can be defined according to an analyst's perception about the kinds of users and interactions occurring on a forum. With these roles assumed to exist, the analyst studies the actions of users and their relations with others. The initial definitions of the roles are then iteratively refined as evidence from the social system is collected.

This chapter presents an overview of the implied role mining process and gives examples of its use to discover roles in the context of an online social system. The review of the process is meant to serve as a guide for social computing researchers who have already a clear, well established picture of the roles emerging in an online social system. The studies that are reviewed are interesting in their own right, but also highlight the fact that implied role mining may lead to separate and potentially conflicting analyses over the same social system by different researchers.

2.2 The Implied Role Mining Process

Implied role mining studies generally follow the workflow illustrated in Fig. 2.1. The figure shows a qualitative, iterative process [3]. Initially, the analyst collects information about the context of the online social system under investigation. The context is defined by the kinds of interactions that are allowed and what the analyst understands is (not) meant by the presence of an interaction. For example, a 'follow' relationship on Twitter from a user to the account of a news agency indicates that the user may use the social service as a news source [6], while a 'follow' relationship between two people whose location is noted as the same high school could suggest a more social kind of connection. On LinkedIn, a 'follow' relationship may imply that someone is interested in the content posted by another (hence a user with many 'followers' may be an *influencer*), while a connection may be indicative of a close business relationship. On an emotional support system like 7 Cups of Tea [2] or CrisisChat, interactions imply that one user cares about the emotional well-being of another user. This initial analysis establishes a 'baseline' set of social roles and hypotheses describing why and how users fall into them.

Data collected from a social system is collected after contextual information and baseline roles are established. This data could consist of information about individual user (e.g. user id, location information, profile description, messages posted),
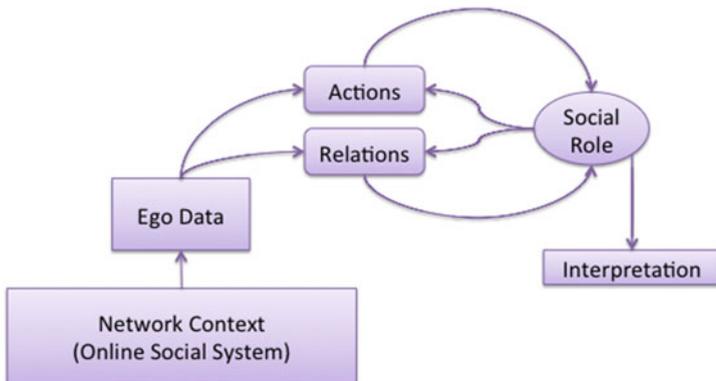

Fig. 2.1 Workflow for implied-role analysis

information about who the user holds various types of relationships with, and information about the actions they partake in. The analyst defines heuristic rules that classify users into roles according to a hypothesis. In this process, it may be the case that an analyst cannot neatly classify users into the hypothesized roles because their actions and relations do not reflect the baseline hypothesis of a role. This necessitates the need to iterate on the definition of the baseline roles first defined and the heuristics used to classify users into them. With the roles and heuristics refined, actions and relations are reevaluated, roles and heuristics are further refined, and the process continues until the analyst is satisfied with the roles defined and the classification of users into them. Finally, interpretation of the roles can then begin.

2.3   Illustrations with Usenet

Examples of implied role analysis to a particular online social system, namely *Usenet*, is discussed next. The presentation is meant to not only explain actual implied role mining processes, but to also demonstrate the kinds of insights implied role mining can reveal.

Usenet is a worldwide distributed system hosting discussion threads about a variety of topics. Users log in to read and post messages (hereafter called a post) into a particular discussion (hereafter called a newsgroup). Newsgroups are themed around a topic that can be gleamed from the name of the newsgroup (e.g. `sci.physics.research`). A Usenet client called a newsreader creates a conversation topic in a newsgroup or adds a post to an existing topic. Moderators are users able to administer a newsgroup with actions approving a topic or post before it is published. Usenet remains an extraordinarily popular way to communicate on the Internet because of its distributed architecture: Usenet content is mirrored across an ever-changing network of Usenet servers, and its protocol uses packets with mutable fields that make it difficult to find the identity of a computer who uploaded a particular message. Its distributed architecture also makes it robust to physical hardware failures. The history of Usenet, dating back to the early days of computer networking, may be another alluring quality to some of its users.

All newsgroups allow users to interact with one another in the same way, but each develop unique kinds of behavioral norms. The norms might moderate the kinds of content that users post, whether or not new topics can be created by new users, and if the newsgroup is open to users asking (simple) questions about the subject of the newsgroup. The norms may reveal how open a newsgroup is to new users, or if the newsgroup fosters a closed community. Distinct newsgroups are thus thought of as different types of *speech communities* [5], which is a group of people that define their own rules for how speech should be conducted and interpreted. Besides the universal role of a "moderator" as defined by the Usenet protocol, each newsgroup may thus have its own set of applicable social roles. For example, users who scarcely post in a closed community, or who get "flamed" or admonished for posting content in a closed community, may take on the social role of an "outsider". A user who

often creates topics that become very active may be seen as a "discussion leader" in supportive newsgroups.

2.3.1   *Golder et al.'s Taxonomy*

Golder et al. performed an implied role analysis on sixteen different Usenet newsgroups [4] to build a taxonomy of social roles. They chose newsgroups that are unmoderated, have a large number of topics that are active, and cover a wide variety of subject matter. Newsgroups in scientific domains (e.g. physics, computer science), hobbies (e.g. gambling, music), and societal issues (e.g. religion) were included in the analysis. For each newsgroup, the set of posts made by a user, the threads the user posted in, and the content of their posts were collected over a one month period. In this context, relationships among users may be defined by whether they participate in the same newsgroup or thread, and even if one user directly responds or reacts to a post made by another.

Evidence from Usenet posts lend support to the six roles presented in Fig. 2.2. Its hierarchy conveys the fact that the flamer and ranter roles are a particular type of troll role, and a lurker is a particular type of newbie. The red dashed outline and shaded boxes of the of troll, flamer, and ranter roles indicate that they try to intentionally disrupt a Usenet community. Two of these roles, namely celebrity and ranter, were devised to satisfy a specific set of user characteristics that may exist in any kind of speech community. For example, celebrities are central figures that emerge in most

Fig. 2.2 Golder et al.'s taxonomy of roles on Usenet. Celebrities, Newbies, and Trolls are observed in all newsgroups, although the behaviors indicative each role may be newsgroup specific. Flamers and Ranters are particular types of Trolls. The *dotted* outline and shaded boxes of the Troll, Flamer, and Ranter roles indicate that they try to intentionally disrupt a newsgroup's user community
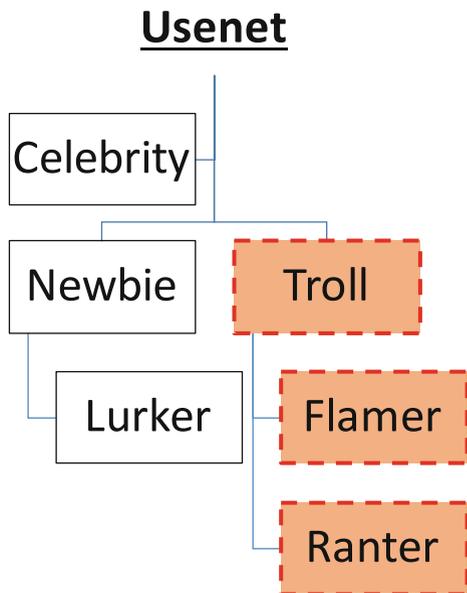
online communities: a prolific poster who devotes a significant amount of energy towards their contributions. Ranters are also prolific posters, but they are transfixed on a specific topic and may not have the interest of the online community at heart. The remaining three roles, namely newbie, troll and flamer, were derived by examining the interactions of specific newsgroups. The authors identified the roles by generalizing the individual kinds of interactions seen in each newsgroup. For example, a newbie in a newsgroup about mathematics may ask a question about a high school homework problem, which community norms may discourage. A newbie in a newsgroup about religion may not ask simple or taboo questions, but could unwittingly violate a unique norm of that community by, say, admonishing the virtues of a religion.

Celebrities. Across every newsgroup, the celebrity role is represented by the small percentage of users who submit huge numbers of high quality, competent posts. The phenomenon of a celebrity is not distinct to newsgroups: most online social systems feature a small percentage of users participating in the majority of interactions or posting the majority of content [7]. Their frequent postings make them well known in a newsgroup community, and their posts may elicit positive feedback and respectful responses. Usenet celebrities also often submit high quality, on topic posts and have viewpoints that are aligned with most other users in the newsgroup. Golder et al. notes that this viewpoint alignment is due to the fact that celebrities typically define and then articulate the zeitgeist of a community, including its most popular beliefs and norms. Celebrities may thus participate for egotistical reasons, as they seek to accumulate social power in a community and improve their own self-image. Positive and frequent contributions help to establish the user as an authoritative figure that further contributes to his or her self-image.

It is important to note that the quality of a post is essential for a user to take on the celebrity role. For example, a frequent poster who does not submit quality content will be admonished by other users, and ultimately be held in lower regard. Moreover, while all newsgroups have celebrity roles, the meaning of 'quality' varies. For example, a gambling community may require a high quality post about a strategy or optimal move in a given scenario to be backed up with a probabilistic analysis. Golder et al. give an interesting example from the `alt.anagrams` newsgroup where quality posts need to give praise to anagrams that are shared prior to the one being presented.

Newbies and Lurkers. Newbies are new users to a newsgroup. They are identified by posts that do not covey a newsgroup's standard of quality, or by posts that do not follow newsgroup norms. They may also not be aware of the prevailing viewpoints and opinions of users in a newsgroup, often leading to conflict. For example, consider a newbie to a newsgroup about personal computers. The user knows very little, and may post a basic question like "Should I purchase a Windows PC or a Mac?". The reader may be able to appreciate how a question like this, presented to a group of

devoted technology experts, could quickly spiral into an argumentative and unhelpful discussion. It may also simply be ignored by the community as a passé question that has been discussed too often and in too many places for it to be worthy of a response.

Usenet newbies are also characterized by users who are simply unfamiliar with Usenet as a technology. For example, newbie users are known to post empty messages or post the same messages many time in short order because they do not understand how to use the newsreader. Usenet newbies may also be unfamiliar with a newsgroup that would serve as a better place to ask a question or to contribute to discussions about a specific topic. Finally, Usenet newbies may not understand how the technology works at a system level. For example, a newbie may be surprised to find that posts cannot be deleted once sent, as deletion commands are ignored by some servers. They may also not be aware that Usenet servers replicate posts across themselves, but this could take time depending on network conditions, thus causing a delay before some other newsgroup participants see a post.

Lurkers are particular type of newbie. Rather than making low-quality posts or posts that do not follow newsgroup norms, lurkers silently read conversations without participating. This practice is meant to teach the lurker about the social norms of the group, and through observation, how to create content that is acceptable and of high quality. Some newsgroups may require new users to lurk, or to read FAQs that discuss rules about what content is acceptable. Lurking is meant to be a temporary role, but a user who cannot understand the norms, lack confidence in their ability to generate quality posts, or are fearful of becoming a newbie to a newsgroup may become a permanent lurker.

Trolls, Flamers, and Ranters. Usenet trolls are infamous for exhibiting negative behaviors in a newsgroup. Their goal is to try to disrupt the experience of others, perhaps by hijacking a conversation to take it off topic, to aggressively attack another post, or to fake honest participation. Examples of faking participation include posting a sarcastic response to a post that, to the uninitiated or to a newbie, may appear to be legitimate or worthy of a response. In Fig. 2.2, the negative behaviors of a troll and its sub-roles are highlighted by a faded background and dashed outline. The flamer is a particular type of troll that responds to others aggressively, is overt in his attacks, and disrupts just for the sake of doing so. Ranters also disrupt, but they do so with a particular agenda or point of view they must express at any cost.

Trolls that are not flamers or ranters are skilled in identity deception. In other words, a troll needs to develop an understanding of the conversational norms and prevailing thoughts of a newsgroup not unlike a newbie. In doing so, a troll learns how to subtly provoke other members to respond, to effectively post sarcastic messages that will lure responses, and to initiate threads that generate pointless responses. Newbies are especially vulnerable to trolls because their lack of understanding about a newsgroup and its norms make it more difficult for them to ascertain if a user is 'trolling' or making a legitimate contribution. Trolls can thus be very dangerous to a newsgroup because, by confusing and victimizing newbies and by eroding the trust of long-standing users to newer members, they diminish its growth and reputation.

Flamers are not interested in deception. Rather, they try to integrate themselves into a newsgroup through a study of the popular topics and users and then post

content that will surely be insulting or controversial. Their speech is often hateful, aggressive, and promotes intolerance for the interests of the particular newsgroup's interest. Flamers are not part of a community, although newsgroups where self-proclaimed flamers congregate does exist. Flamers are fought by strong celebrities and newsgroup leaders who promote norms discouraging flaming behaviors. Posts by flamers could thus go ignored or quickly disarmed by users who quickly criticize the flamer and redirect the conversation to a different topic.

Ranters are not interested in deception, nor do they troll with the only intention of attacking and disrupting conversations. Instead, ranters exhibit negative behaviors, perhaps unbeknownst to themselves, in order to promote an idea or viewpoint at any cost. Their unwillingness to consider alternative ideas or views lead them to exhibit some troll like characteristics, like aggressively attacking or writing lengthly posts to counter the points of an opposing viewpoint. Users who argue with a ranter often find themselves in an endless argument, with the ranter unwilling to secede from any of his thoughts.

2.3.2   Nolker et al.'s Hierarchy

Nolker et al. introduced a different collection of Usenet roles as shown in Fig. 2.3. Their implied role mining approach focuses only on users who make large contributions to a newsgroup through frequent posting. They define two broad roles: key members and chatters. Key members are defined as users who positively contribute to a number of discussions on a newsgroup, while chatters focus their attention on a small number of threads. Key members are further refined into leaders, who create many threads and provides cohesive thoughts across newsgroup threads, and motivators who contribute to existing threads more so than striking up new ones. Like Golder et al.'s taxonomy, the broad roles were defined through iterative qualitative evaluations of use behaviors, conversations, relationships, and ego-network structures. The evaluations occurred over a database of 25,737 postings over a one year period from a single news group `alt.support.loneliness`. Although fewer roles are defined, the taxonomy comes paired with a set of metrics for classifying a user into one of the three roles according to the structure of their ego-network and the frequency of their posts to newsgroup threads.

Key Members. Key members may be thought of as the 'core' set of users that contribute across the entire newsgroup, is willing to strike up new threads, and participates often. In other words, key members are essential for maintaining a baseline level of activity in the newsgroup. A leader is distinguished as a popular user who contributes to a number of threads, irrespective of the subject, and is not afraid to initiative new threads. Leaders thus have wide exposure to other users and make a total contribution to a newsgroup. The contribution of a motivator is only smaller in the sense that they would rather contribute to existing threads than initiate new ones.

Nolker et al. also examine specific attributes and measures of users falling into each role. They are defined as follows: Consider a social network where user *A* has
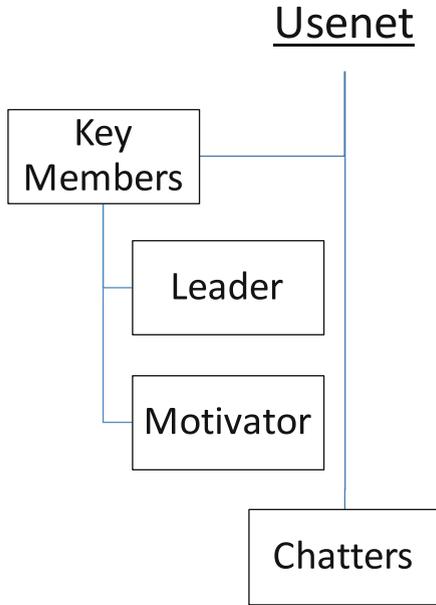
Usenet

Key
Members

Leader

Motivator

Chatters

Fig. 2.3 Nolker et al.'s taxonomy of social roles on Usenet. Key members and chatters are broadly defined roles with leaders and motivators as types of key members. Leaders spread knowledge and provide cohesiveness and consistency to a conversation, Motivators are users who contribute to conversations. Chatters are users who post often, but only in a small number of threads

a directed relationship to user B if A responds to a post made by B in a discussion thread. When a directed edge exists from $A \rightarrow B$, we say that A is engaged in a *one way conversation*. A *two way conversation* is one where there is an edge from A to B and from B to A in the social network. A two way conversation is *direct* when these edges are due to responses within the same discussion thread, otherwise it is *indirect*. The fraction of two way conversations that are direct is the *discussion ratio* of a user. For example, a user with a high discussion ratio indicates that the user has a number of back and forth conversations with others within the same thread, and hence is engaged in a concentrated discussion about a particular subject. Furthermore, we say the *degree* of a user is the number of conversations he is engaged in, the *betweenness* of a user is the number of shortest paths between any pair of others he is a part of, and the *closeness* of a user as the average distance between him and all other users in the social network. The 1-TFIDF of a user is given by:

$$1\text{-TFIDF} = \log(O + 1) \log D \tag{2.1}$$

where O is the total number of one way conversations of a user and D is the fraction of users in the network with which a conversation is held, respectively. 2-TFIDF and T-TFIDF are defined analogously for two way conversations and for threads.

Table 2.1 Attributes and corresponding measures for a leader

Attribute	Measure
Contributes to many conversations	High node in-degree
Receives many responses	High node out-degree
Conversational conduit	High betweenness
Responds to many posters a similar number of times	Low 1-TFIDF and Low 2-TFIDF
Has a mix of direct and indirect two-way responses	Moderate discussion ratio

Table 2.1 presents the behavioral attributes of a leader that Nolker et al. expects one to exhibit. It also gives a measure related to the social network or posting frequency of a user for that behavior. Because a leader is expected to participate in many different conversation threads, he will be responding to (and his posts will be responded by) a large and diverse set of other users. These users should thus have very high in- and out-degree in the network. Leaders may also be a conversational conduit, in the sense that his response to a post by one user encourages or inspires another user to post in the thread. For example, a leader might respond to a post with a unique point of view that resonates with another user, encouraging a response. Leaders should thus have high network betweenness, as his posts form bridges for other users. To make contributions across an entire newsgroup, leaders should be posting in a variety of discussion threads, and hence, will submit posts that respond to a large number of others. However, keeping with the idea that the leader makes contributions across the newsgroup, we do not expect him to fixate most of his replies to a single other. This implies that the leader should have a low 1-TFIDF and 2-TFIDF. Finally, following the fact that this user is a key member, a leader should have a mixture of indirect (engaging with as many users on the newsgroup) and direct (engaging with many users on a specific topic) conversations, suggesting a moderate discussion ratio.

Table 2.2 gives behavioral attributes for a motivator. The motivator has a large number of posts but well distributed across many different threads. Owing to the wide exposure of the motivator, he will respond and be responded to by many others. This translates into a motivator as having a high closeness value in the social network. That a motivator posts over many threads, rather than concentrating his posts to a

Table 2.2 Attributes and corresponding measures for a motivator

Attribute	Measure
High posting count spread over many threads	High closeness
High posting count spread over many threads	Low T-TFIDF and 1-TFIDF
Has a mix of direct and indirect two-way responses	Moderate discussion ratio

Table 2.3 Attributes and
corresponding measures for a
chatter

Attribute	Measure
Posts often, but are exposed to a small set of others	High 2-TFIDF
Majority of responses are limited to a thread and a specific set of users	High discussion ratio

single thread, also implies a low T-TFIDF and 1-TFIDF value. Like the leader, as a key member, we also expect the motivator to have a moderate discussion ratio.

Chatters. In contrast to key members, chatters post legitimate content and contribute in a limited context (e.g. one discussion thread). They are not key members because they may stop contributing to a newsgroup after posts in threads they are interested in cease. Chatters may make interesting and meaningful posts to a thread, but their lack of interest in the entire newsgroup is underscored by the fact that they do not post about other subjects.

Table 2.3 gives the behavioral attributes and measures for a chatter. Since they limit themselves to a small number of threads, the posts of a chatter will only be exposed to the small set of users subscribed to the particular threads. Thus, one expects chatters to exhibit large 2-TFIDF. They will also post in response to others, and because their posts are substantive and on-topic, should elicit responses back. Thus, a chatter should have a high discussion ratio as he participates in a larger number of two way conversations.

2.4 Analysis of Implied Role Mining

In light of the above review of separate implied role mining studies over the same social system, it is informative to further examine important aspects of this approach.

2.4.1 Qualitative Nature

First, the qualitative aspects of implied role mining should now be very clear. Before examining data in detail, implied role mining requires the analyst to consider what they know about the domain of a system and hypothesize about the kinds of user roles that could exist on it. While the functionalities of a system offer guideposts, the selection of role labels and their meaning ultimately lay on the analyst. Domain knowledge is surely necessary to find these initial roles. Examples of these roles on Usenet included newbie, troll, and key member. Looking at the data *after-the-fact* then helps the analyst refine the definition of roles, and even define new, more specific ones according to interesting trends and behaviors that were observed. The process

of this iteration is very loose, and to some extent, finding more specific types of roles to define is a bit of an art. Neither Godler et al. nor Nolker et al. define a specific algorithm or procedure for iterating and converging on roles in their data analyses, which is understandable given the informal way we think about what a social role means and how we can identify them. Converged upon social roles are also highly dependent on the set of initial roles that are defined. This is because the initial roles essentially define how the analysts will cluster users together when they examine the data. More refined roles are then defined according to behavioral patterns common within the pre-defined clusters of users.

The qualitative nature of implied role mining can be seen in the two Usenet studies. Note how Golder et al. defined the broader roles celebrity, newbie, and troll with some behaviors that could be applied to most kinds of online social systems. For example, a celebrity is a popular user, a newbie is someone unfamiliar with the system's norms, and a troll as a user with ill intentions. Nolker et al. define the broad roles key member as anyone vested in the system and chatter as someone who contributes often but is less vested. Even the evaluation of the social roles is a qualitative one. For example, Golder et al. identified specific posts and conversations indicative of roles, while Nolker et al. defined metrics mentioned in Sect. 2.3.2 reflecting various behaviors and then found users whose metrics fall into relative ranges ('low', 'moderate', 'high') expected for each role.

2.4.2 Compatibility

The two Usenet implied role analyses were correctly executed, well designed, and provided ample information about the social roles that were unearthed, despite the fact that they defined different sets of social roles. The two most overlapping roles are celebrities and key members, with the distinction being that celebrities submit posts that help define the norms and prevailing thoughts of a newsgroup while key members simply create and post in many threads and maintain a high level of activity. Such kinds of overlap begs the question, *are separate implied role mining studies over the same social system* **compatible**? In other words, can we synthesize together the findings from multiple role mining studies into a single one, without any contradiction? Incompatibilities, which could arise if separate social roles are defined by the same definition, or if the definition of social roles in the studies carry conflicting definitions, would encourage one to choose a single taxonomy that he or she thinks best represents the true roles of users, thus adding yet another layer of subjectivity to implied role mining. Compatibility may be assessed by checking if (i) the studies define separate, distinct sets of social roles; and (ii) when a user satisfies the definition of a role in many studies, if it is reasonable to assign all such roles to the user.

The first check for compatibility ensures that every social role carries a unique definition. For example, a chatter (from Nolker et al.'s taxonomy) as a user who limits himself to a small number of threads is certainly not a celebrity, newbie, or lurker (from Golder et al.'s taxonomy) because the chatter does not post widely and

is experienced enough to post often in a single thread. Nor could a chatter be a troll, flamer, or ranter, since chatters by definition offer positive contributions to a newsgroup. When there is *some* overlap between roles from different studies, such as the overlap of key members and celebrities, details in their definition can establish a distinction that separates the roles. For example, some measure of how the frequent posts of a user define the norms and steer prevailing viewpoints across a newsgroup may be used to separate those who are celebrities from those who are key members.

Whereas the first check evaluates the case when the definition of two roles overlap, the second check evaluates the case when a user satisfies the definition of multiple roles. For example, if we identify a user who submits posts that agitate others in a limited number of threads in a newsgroup, is it appropriate to call this user a chatter and a troll (a chatter-troll)? Can a new user who is overly anxious to begin posting content on threads a newbie-chatter? The answer to these questions is subjective. An analyst could argue that newbies cannot be chatters because, by definition, a newbie will not post volumes of content on threads without an understanding of the norms of the newsgroup. Another may say that yes, newbies could be chatters if the newbie was more familiar with norms from a different but related newsgroup that was acceptable.

The issue of compatibility grows even more complex as the number of implied role mining studies performed over a social system increase. If $R_1, R_2, \ldots, R_n$ are sets of roles devised from n separate implied role mining studies, sets of roles in the power set $\mathbb{P}$ of $\{R_1, R_2, \ldots, R_n\}$ need consideration. Specifically, for each set of roles $\{r_1, r_2, \ldots r_k\} \in \mathbb{P}$, one needs to check that the definitions of $r_1, r_2, \ldots r_k$ are compatible and if it is feasible to allow a user to take on these roles simultaneously. This implies that resolving many implied role analyses over the same system can quickly become difficult and time consuming.

2.4.3   Simplicity and Interpretability

Owing to its qualitative nature, a major feature of implied role mining is its simplicity and interpretability. This is emphasized in Fig. 2.4, which summarizes the differences between implied role analysis and the more quantitative methods to be discussed in subsequent chapters. The major difference is that the latter does not presume that the analyst knows the kinds of roles that users exhibit in a system, whereas the former requires a definition of some roles before extensive data analysis. As to be explained, specifying what it *means* for nodes to fall in the same position may be defined by an equivalence definition, by a statistical measure of user or node similarity, or by a model that probabilistically describes the construction of a social network as a function of role assignments. Computational machinery then maps users to positions, and the analyst defines the meaning of a position (i.e. defines a role) by common patterns among the mapped users. The position-based process is rigorous and driven by algorithms, enabling its consistent application no matter the system or study.
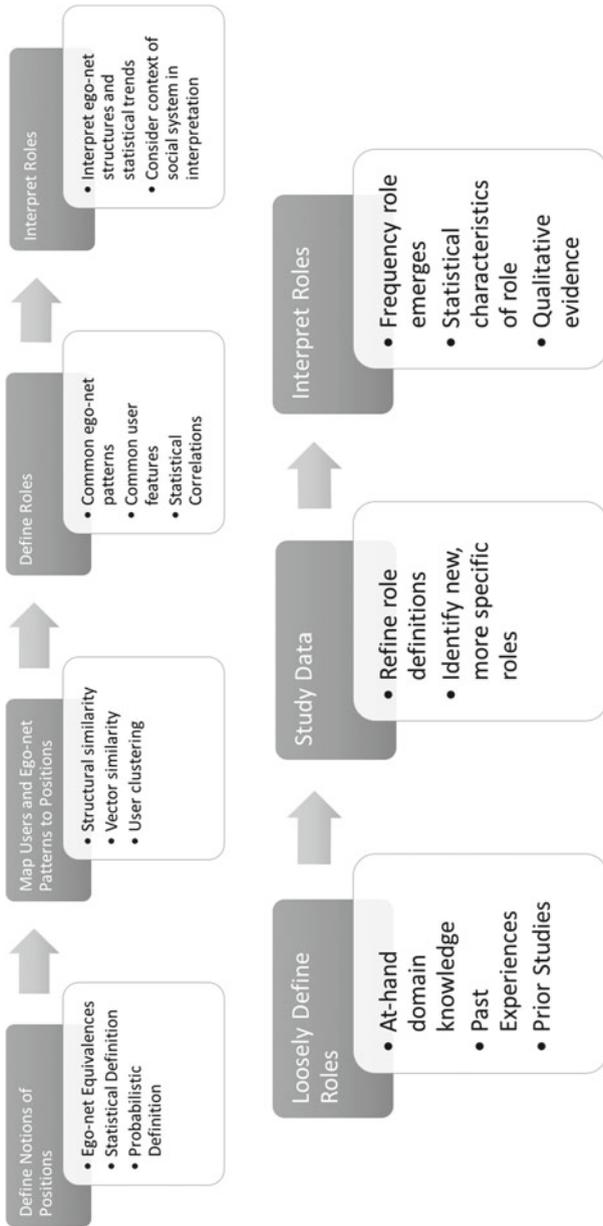
Fig. 2.4 Comparison of implied role mining (larger; below) against other role mining methods (smaller; above). Other methods initially asks an analyst to define the meaning of a position, and then uses an algorithm to define positions to users. Characteristics of users and their ego-networks in each position define a label and interpretation of the role, which can sometimes be hard to achieve. Implied role mining instead asks an analyst to define roles up front. Heuristics and qualitative thinking refines the definition of the roles and their interpretation. Implied role mining is thus an easier to perform and to interpret, but is less rigorous

Certainly, the implied role mining process is simpler to implement and understand. First, asserting the existence of some types of roles based on our knowledge about the social system is easier than choosing a mathematical definition or model of a 'position'. Second, to a novice, algorithms mapping users to position may be seen as a kind of black-box that, if not well understood, may not give an analyst confidence in its results. Algorithms may also be unable to scale well to networks that are very dense, large, or have users with a large number of attributes. Furthermore, while 'roles' extracted from position-based techniques are meaningful by mathematical definitions, they can be very difficult to interpret. The process of refining the definition of social roles in implied role mining is qualitative and may be biased to the beliefs of a particular analyst, but it is at least transparent and follows a human-driven rationale. Moreover, the derived roles have a human understandable definition tailor made to the context of the social system under study. The results of an implied role mining study are therefore simple to interpret and apply.

2.5   Conclusion

This chapter introduced implied role mining as a way to mine social roles. The implied role mining process considers contextual information about a social system to define a set of roles before extensive data analysis commences. The definitions of these roles are then refined, and new ones may be formulated, as the data goes under scrutiny. At the end of the process, a set of well defined, easy to interpret social roles have been established. The methodology is best suited for social systems where interactions are done and relationships are formed in a limited number of ways. Online social systems, with well defied functionalities and ways to express a relationship, are popular candidates for implied role analyses.

Two example implied role analyses, carried out over newsgroups on Usenet, demonstrate the kinds of well defined role taxonomies the methodology can develop. Both examples presented well defined, understandable roles that are reasonable given the context of Usenet. For example, newbies are new users that need time to learn the norms of a newsgroup before participating, trolls seek to derail and disturb newsgroup members, celebrities are popular users that steer conversations and drive the definition of norms, and chatters make positive contributions to but a limited number of threads. Imagining how users who take on these various roles interact with each other on Usenet paints an interesting picture about Usenet as a society.

The example studies demonstrated important qualities of implied role mining. Its qualitative nature can lead to soft and interpretable rules for classifying a user into a given type of role. Furthermore, the derived roles strongly depend on the initial set that the analyst believed to exist. The compatibility of different implied role mining studies over the same social system needs to be resolved. This challenge increases with the number of different studies performed. Compared to position-based approaches for role mining, which the rest of this brief is devoted to, implied role mining is simpler to perform and yields results that are easier to interpret.

References

1. Calzarossa, M.C., Massari, L., Doran, D., Yelne, S., Trivdei, N., Moriarty, G.: Measuring the users and conversations of a vibrant online emotional support system. In: Proceedings of the 2016 IEEE International Symposium on Computers and Communication (2016)
2. Doran, D., Yelne, S., Massari, L., Calzarossa, M.C., Jackson, L., Moriarty, G.: Stay awhile and listen: User interactions in a crowdsourced platform offering emotional support. In: Proceedings of the 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, pp. 667–674. ACM (2015)
3. Gleave, E., Welser, H.T., Lento, T.M., Smith, M.A.: A conceptual and operational definition of 'social role' in online community. In: 42nd Hawaii International Conference on System Sciences, pp. 1–11 (2009)
4. Golder, S.A., Donath, J.: Social roles in electronic communities. Int. Res. **5**, 19–22 (2004)
5. Hymes, D.: Foundations in Sociolinguistics: An Ethnographic Approach. Psychology Press (2003)
6. Kwak, H., Lee, C., Park, H., Moon, S.: What is Twitter, a social network or a news media? In: Proceedings of 19th International World Wide Web Conference, pp. 591–600 (2010)
7. Mislove, A., Marcon, M., Gummadi, K., Druschel, P., Bhattacharjee, B.: Measurement and analysis of online social networks. In: Proceedings of the ACM Internet Measurement Conference (2007)