

A Corpus of Participant Roles in Contentious Discussions

Siddharth Jain, Angelique Rein, Archana Bhatia, Eduard Hovy

University of Southern California, Carnegie Mellon University

USA

siddhajj@usc.edu, {arein, archnab, ehovy}@cs.cmu.edu

Abstract

Participants in online social environments adopt various roles. The concept of *social role* has long been used in social science to describe the intersection of behavioural, meaningful, and structural attributes that emerge regularly in particular settings. In this paper, we explore various behavioural attributes such as stubbornness, sensibility, influence, and ignorance to create a model of the social roles that participants assume. We annotate discussions drawn from two different corpora in order to validate our model of social roles and their signals. We discuss various criteria for deciding values for the behavioural attributes that define the roles. We deliver a new corpus of online contentious discussions with annotated social roles.

Keywords: Social roles, Contentious discussions, New corpus

1. Introduction

People assume certain distinct roles when participating in discussions: we all recognize discussion leaders, gadflies, and so on. According to the Dictionary of Psychology (Corsini, 2002), a role is “the set of behaviors expected of a person possessing a certain social status” (p. 850). Accordingly, a role is an upper-level concept that can comprise a set of behaviors.

As social life becomes increasingly embedded in online systems, the concept of social role becomes increasingly valuable as a tool for simplifying patterns of action, recognizing distinct participant types, and cultivating and managing communities. These roles organize behavior and give structure to positions in local networks. In decision-making discussions, the roles may exert a major influence on the outcome of the discussions. It is therefore helpful to be able to identify automatically the role someone is playing; this helps explain interaction patterns and allows one to cultivate and manage communities. For example, the roles can be used to identify authoritative sources in social media, find influential people in the community, detect subgroup and subgroup organizations, mine attitude towards events and topics, measure public opinion and the controversy of different topics, summarize different viewpoints with respect to some topic or entity, and many other similar applications.

The number of such roles, and what exactly defines a discussion leader or any of the other roles, is however not easy to specify. Although the automated analysis of social roles in online discussions has attracted the NLP community recently, very few benchmark corpora are available to researchers active in the domain. This paper presents a new, publicly available, corpus of online contentious discussions, including a rich set of annotations of behavioral attributes such as stubbornness, sensibility, influence, and ignorance, which we believe all contribute in the identification of roles played by participants.

We have been developing a classification of the various roles adopted by participants in contentious discussions.

Our investigation shows that certain behavioral features reliably reflect the role a participant has adopted. These features pertain to various observables, including patterns of interaction between participants, individuals’ own contributions (of facts, opinions, interpersonal remarks, etc.), patterns of silence, and so on. We have been annotating a set of contentious discussions for these observables. In Section 2 we describe the corpus, and in Section 3 the annotation procedure that eventually produced our model. In Section 4 we outline the model of behavioral attributes and various criteria for deciding values for each attribute. We define the social roles in Section 5.

2. Corpus

The corpus presented in this paper consists of two sets of online contentious discussions. The first set consists of 80 discussions from the *Wikipedia: Article for Deletion (AfD)*¹ forum. Wikipedia, being a very large peer production system, has its own decentralized governance system to maintain the quality of articles created by users. Any Wikipedia user can nominate any article on Wikipedia for consideration for deletion. Any interested participants are then welcome to state their stance (usually in their first contribution to the discussion), after which an open discussion ensues. Participants may change their stance during the discussion. The discussion continues for at least 7 days. When it has died down, a Wikipedia administrator officially assigned to the discussion declares the final consensus, which may or not reflect the numerical majority of the participants’ opinions.

The second set consists of 10 discussions from *4forums.com*² forum. These discussions are political debates on controversial topics. Unlike the AfD corpus, participants do not have to state their stances explicitly, there is no time limit for how long the discussion can

¹ Wikipedia: Articles for Deletion (AfD): http://en.wikipedia.org/wiki/Wikipedia:Articles_for_deletion

² 4forums.com: <http://www.4forums.com/political/>

continue, and there is no official moderator or facilitator. Participants on this forum express their views and argue with others without any visible goals to achieve.

We selected these rather different corpora in order to ensure that our model of social roles and their signals holds up in general. Although both sets consist of contentious discussions, the nature of the discussion is very different. Where the AfD discussions have a measured and polite tone, the 4forums discussions can become quite heated and ad hominem. Participants on the Wiki forum are goal-oriented—they want their stance to be the final consensus of the discussion—while participants on the 4forums.com forum are opinion-oriented—they are primarily focused on presenting their own viewpoint. This is a classic example of argumentation by reason vs. argumentation by insistence.

We present some stats for both Wikipedia: AfD and 4forums.com corpus in Table 1.

	Wikipedia: AfD	4forums.com
#DISCUSSIONS	80	10
#PARTICIPANTS	500	107
#TURNS	1487	624
#WORDS	96138	51695

Table 1: Stats for both corpus

3. Annotation Procedure

The annotations began with a training annotation set consisting of 10 AfD articles. Three annotators were asked to identify two basic social roles performed by participants (*Leaders* and *Rebels*, where *Rebels* were described as the participants who have enough contribution but are unable to exercise any kind of influence). Also, they were asked to assign any other role that would identify a participant with characteristics different from the two given roles.

After the completion of the initial task, the annotators agreed upon a set of social roles for the initial coding manual. The annotators came up with a set of characteristics that define each role and also the criteria to assign values to each characteristics for each participant. Using the initial coding manual, the annotators were asked to annotate 8 more sets, each consisting of 10 AfD articles. After annotating each set, the annotators discussed the annotations and refined and/or added any roles, characteristics, or criteria that were agreed to be helpful. After completing all 8 sets, the annotators re-annotated all the discussions again using the final coding manual.

The annotations for the 10 discussions from the 4forums.com forum started using the same coding manual. Some of the criteria for assigning values to characteristics were modified in order to adapt to the different style of 4forums.com discussions but the set of characteristics and social roles remained the same. The final coding manual included 4 characteristics and 8 social roles, described in Sections 4 and 5 respectively. In the annotation process, each annotator determines the value of each characteristic for each participant in the discussion, and subsequently

assigns the corresponding role.

We compute inter-annotator agreement for annotated social roles using Cohen's Kappa³ coefficient.

	Kappa coefficient
Annotator1-Annotator2	0.74
Annotator2-Annotator3	0.71
Annotator3-Annotator1	0.69

Table 2: Cohen's Kappa score for roles' annotation

4. The Model

As annotation progressed, we increased the number of roles and refined their definitions and characteristics they are comprised of. Eventually, we produced the following principal roles: *Leader*, *Follower*, *Rebel*, *Idiot*, *Voice in Wilderness*, *Nothing Sensible*, *Nothing*, and *Other*. We define these roles in Section 5 below.

Digging deeper, we identified three aspects of participants' overall contribution, further subdivided into four identifiable characteristics, which in various combinations reflect the behaviours of these social roles: *Participation Type* (*Stubbornness*(St), *Sensibility*(Se)), *Attendedness Value* (*Ignored-ness*(Ig)), and *Influence Value* (*Influence*(In)).

We discuss the three aspects and their corresponding characteristics next.

4.1 Participation Type

This aspect helps us determine the type and amount of contribution the participants make and what their style of participation is.

4.1.1. Stubbornness

Stubbornness captures the intransigence of a participant in the discussion. This characteristic has two components: the amount of participation and the degree of unwillingness to change opinion or stance. This characteristic differentiates between participants who form the heart of the discussion from participants who may comment only once or twice, or in minor ways only. A combination of the following criteria determine participation: the number of comments by the participant, the arguments/claims presented by the participant, and the level of engagement with other participants. We compare the number of comments by each participant against the average number of comments for each discussion, if the number of comments is higher than the average, the participant is considered stubborn. However, note that while calculating the average, we do not consider the outliers. For example, if most of the participants comment between 1–5 times, but there is a participant who comments 15 times, then while calculating the average, the participant with 15 comments is considered an outlier. When counting the number of comments by a participant, we do not account for comments that are not considered stubborn, i.e., the comments that do not present/strengthen their arguments. Some examples of these may be a query asking for information, notes stating

³ Cohen's Kappa: http://en.wikipedia.org/wiki/Cohen's_kappa

changes in the Wikipedia article, a comment just endorsing some other participant, or a comment having no arguments. The first two comments in Figure 1 show such examples from the Wikipedia corpus and the third comment shows the same from 4forums.

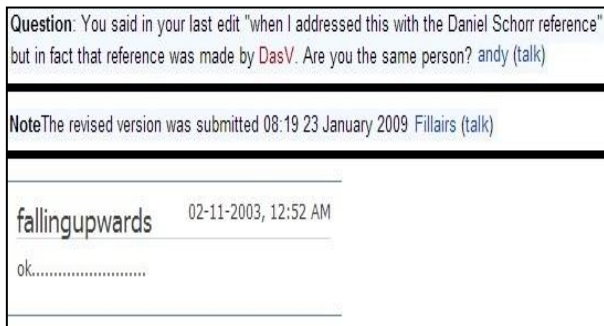


Figure 1: Non-stubborn comments

If a participant's number of comments is lower than the average, but the length of the comment(s) is large enough and their arguments/claims either strengthen the case for their own stance or weaken the case of the opposite stance, then they are considered stubborn. This is because this captures the level of commitment from the participant towards his/her own stance. Alternatively, if most of the comments made by a participant are just responses to other participants' questions and do not really support their own stance, then those comments are not considered as contributing towards the stubbornness of the participant. We also consider whether the participant only presents his/her own arguments or he/she also replies to others' comments to further support their stance.

The possible values for stubbornness are +1, 0 or -1, where +1 means that the participant is stubborn, -1 means that the participant is not stubborn and 0 means that the stubbornness level for the participant cannot be determined.

4.1.2. Sensibility

The level of sensibility of the arguments/claims presented by a participant is very important for measuring his or her impact on the discussion. Sensibility analysis is highly dependent on the domain or nature of the discussion. Therefore we define somewhat different criteria for assigning sensibility values in the two corpora. As mentioned, the AfD discussions are goal-oriented: each participant tries to sway the decision of the discussion in favor of their own stance. Also, since Wikipedia pages should meet the requirements stated in *Wikipedia policies*⁴, as one would expect discussions on this forum sometimes revolve around such policies. Therefore the main criterion for someone to be sensible in such discussions is that they appeal to authority in support of their arguments/claims. Examples of such authority can be Wikipedia policies, links to external sources of recognized expertise, etc. The first comment in the figure refers to a Wikipedia policy WP:PROF which contributes to sensibility whereas the

second comment is considered sensible because it questions the sources of the article and notability which in turn refers to Wikipedia policies.

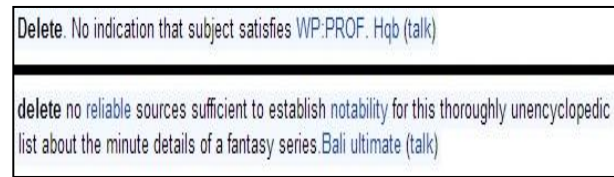


Figure 2: Sensible comments including Wikipedia policies

Another criterion is to evaluate whether the type of argumentation used is reasonable, logical, expands or develops on the previous arguments, reflects a vision; or whether it is a non sequitur or starts a tangential discussion. We also try to decide whether they know the domain: do they seem knowledgeable? Also, since the discussions on AfD include a wide range of topics, many of which the annotator may not be familiar with, participants' analysis of arguments/claims presented by other participants, i.e., peer review, also plays an important role in determining the sensibility level. Further, if it seems that the participant's contribution is guided more by his or her emotional and often inappropriate response to others, e.g., if they directly attack a participant or show that they are hurt by others not adopting their position or agreeing with them, these participants are considered not sensible.

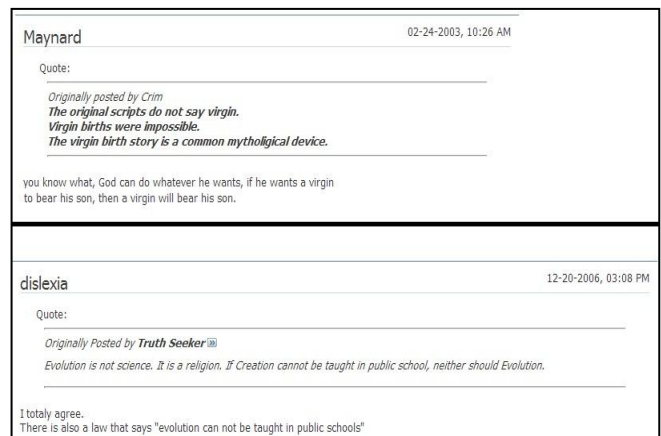


Figure 3: Non-sensible comment on 4forums.com

In contrast, the discussions on 4forums.com are opinion-oriented, where participants primarily focus on presenting their own opinions and reasoning, and do not seriously consider that of others except to dispute it. In this domain our sensibility analysis differs from the AfD forum in several ways. First, expressing emotions can be considered sensible; second, the introduction of tangential discussions, common in 4forums.com discussions, does not have an effect on our sensibility analysis; third, requiring a reasonable and logical argument does not apply in this domain. That does not mean that anything

⁴ Wikipedia policies: http://en.wikipedia.org/wiki/Wikipedia:Policies_and_guidelines

goes, however. Figure 3 presents examples of non-sensible comments in 4forums.com. The first comment seems illogical whereas the second comment is untruthful.

Since 4forums.com has no argumentation policies to refer to and the topics for the discussions are controversial socio-political issues, our sensibility analysis is heavily dependent on the type of issue. One of the criteria is to distinguish arguments that present personal preference as opposed to ones that present global effects. For example, participants' arguments may reflect local-level reasoning (e.g., personal preferences such as participants wanting their children to have Christian education) as opposed to global-level reasoning (e.g., deforestation having a major impact on global warming). This depends on the nature of the issue: local-level reasoning about local-level issues (e.g., whether home schooling is good) are considered sensible. However, participants' personal preferences stated for global-level issues (e.g., whether evolution and religion can coexist) are considered low on sensibility. Our annotation judgments in this forum are performed more intuitive than for the AfD corpus.

The possible values for sensibility are +1, 0 or -1, where +1 means that the participant is sensible, -1 means that the participant is not sensible and 0 means that the sensibility level of the participant cannot be determined.

4.2 Attendedness

This aspect is concerned with how other participants attend to a participant's contribution.

4.2.1. Ignored-ness

Ignored-ness captures the attitude of participants towards each other. This attitude can be an indicator of the importance or relevance of a comment. Generally, for example, participants ignore spammers. The main criterion for ignored-ness analysis is to see whether the participant receives any replies to his or her comments. Therefore, when someone mentions a participant in his/her comment then that participant is considered not ignored. Wikipedia:AfD discussions are stored in a structured way (replies to a comment are indented), making determination of replies easy. In contrast, in 4forums discussions we have to rely on a participant quoting arguments presented by another participant, or naming them explicitly, to decide that the latter is not ignored. Figure 4 shows some examples that indicate that a participant is not ignored: participants *Boston* and *Crim* are being replied to in the comments shown.

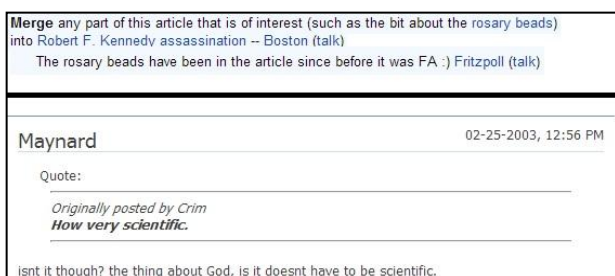


Figure 4: Not-ignored participants

The possible values for ignored-ness are +1 or -1, where +1 means that the participant is ignored and -1 means that the participant is not ignored.

4.3 Influence Value

This aspect helps in mainly to identify leaders and followers among the other roles. This has two aspects: (i) influence on others: was a participant able to influence another participant through their contribution?, and (ii) endorsement: did a participant acknowledge another participant's influence explicitly? This aspect helps in mainly to identify leaders among the other roles. The primary characteristic of leaders in contentious discussions is that they are able to influence others by their actions or arguments/claims. Therefore influence analysis becomes a key part of identifying leaders and followers in such discussions. Below we mention the indicators of influence.

4.3.1. Influence

This characteristic deals with the question if the contributor is influential to other participants. The most observable indication of influence occurs when another participant changes his/her stance during the discussion and acknowledges the influential participant for influencing him/her for the change. In such cases, the participant who was influential in engendering the change is considered a leader and the participant who changes his/her stance or endorses other participants is considered a follower. Another example of indication of influence is when other participants acknowledge the influential participant through expressions such as "according to ...", or "as per ...", etc. Figure 5 shows some examples of influence indicators.

The possible values for influence are +1, 0, or -1, where +1 means that the participant was able to influence another participant, -1 means that the participant got influenced by another participant, and 0 means that the participant was neither influential nor got influenced by others.

An early implementation of a simplified model of Leadership appears in (Jain and Hovy, 2013).

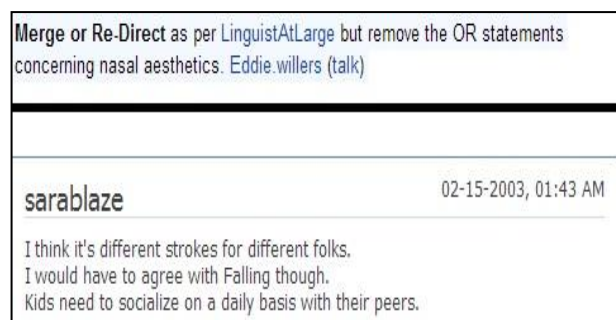


Figure 5: Examples of influence indicators

5. Social Roles

As mentioned, the final coding manual comprises 8 principal roles distinguished by 4 characteristics. Table 2

shows the relationship between the characteristics' values and the corresponding social roles.

St	Se	Ig	In	Role
x	x	x	+1	Leader
x	x	x	-1	Follower
+1	+1	-1	x	Rebel
+1	+1	+1	0/-1	Voice in Wilderness
+1	-1	x	x	Idiot
-1	+1	x	x	Nothing Sensible
-1	0/-1	x	x	Nothing

Table 3: The relation between the values of characteristics and social roles corresponding to them. Stubbornness, sensibility, and influence can take values +1, 0, or -1. Ignored-ness can take values +1 or -1. An 'x' in the table indicates that the corresponding characteristic for the role can take any possible value. As one can notice, participants can thus have multiple roles. A Rebel can become Rebel-Leader if the participant has influence value +1. Similarly, a Nothing can become Nothing-Follower if influence value is -1.

Note that the roles Leader and Follower have a special status in that all the characteristics' values except *influence* are unspecified. This means that any participant, irrespective of their sensibility, stubbornness etc., may be seen as a Leader/Follower as long as they have the appropriate influence value. Hence any other roles could additionally acquire the qualities of being a Leader or Follower.

Any combination of characteristics not specified in the table is annotated as the role Other. We discuss each role and their characteristics next.

5.1 Leader

We define a Leader as a participant who manages to influence another participant to change his/her stance or influence them to follow him by endorsing through his/her actions or arguments/claims. We believe that the defining characteristic of a Leader is the amount of influence he/she is able to induce regardless the amount or type of contribution. The characteristic combination for a Leader is (x, x, x, +1) for stubbornness, sensibility, ignored-ness, and influence respectively, where 'x' means any value for that characteristic is allowed.

5.2 Follower

As opposed to a Leader, a Follower is defined as a participant who is influenced by other participants and changes his/her stance or endorses other participants for their actions or arguments/claims. A Follower is also a participant who doesn't have his/her own arguments/claims, but instead re-states arguments/ claims made by other participants. These participants provide support to leaders by endorsing them or by re-stating the

same arguments/claims. Therefore, the contribution amount or type doesn't matter for such participants to define them as a Follower, making the characteristic combination for a Follower (x, x, x, -1) for stubbornness, sensibility, ignored-ness, and influence respectively.

5.3 Rebel

A Rebel is a participant who forms the heart of the discussion and drives it in some direction. One of the main characteristics of a rebel is his/her devotion to the discussion, based on the amount of contribution and their level of engagement with other participants. The arguments/claims presented by a rebel are sensible and he/she is not ignored by other participants, which provides justification for the importance of his/her presence in the discussion. The characteristic combination for a rebel is (+1, +1, -1, x) for stubbornness, sensibility, ignored-ness, and influence respectively.

5.4 Voice in Wilderness

A Voice in the Wilderness is very similar to a Rebel in the amount and type of contribution which forms the heart of the discussion. The arguments/claims presented by a voice in wilderness are sensible as well. The only difference between a Voice in the Wilderness and a Rebel is that the former is ignored by other participants for some reason. Therefore the sensibility value for a Voice in the Wilderness is important to distinguish them from potential spammers, since spammers are never regarded as sensible contributors. The characteristic combination for a Voice in the Wilderness is (+1, +1, +1, x) for stubbornness, sensibility, ignored-ness, and influence respectively.

5.5 Idiot

An idiot is a participant whose contribution to the discussion is not towards serving the purpose of it. He/she does participate a lot in the discussion but the content may be either emotional, illogical, or completely off-topic. The ignorance characteristic has no significance in defining the role of an idiot since the main criterion for labeling a participant an idiot is the non-sensible contribution by himself/herself. The characteristic combination for an idiot is (+1, +1, +1, x) for stubbornness, sensibility, ignored-ness, and influence respectively.

5.6 Nothing and Nothing Sensible

These are participants who make minimal contribution to the discussion and hence cannot be considered stubborn enough to stick to their arguments/claims. As a result, they may not have a major influence on the course or the outcome of the discussion. We distinguish between Nothing and Nothing-Sensible based on the number of sensible arguments/claims, in order to distinguish potential spammers from those who may have minimal but legitimate contribution. The latter may or may not be ignored based on the type of contribution. The characteristic combination for a nothing is (-1, -1, x, x) for stubbornness, sensibility, ignored-ness, and influence respectively, and the characteristic combination for a nothing sensible is (-1, +1, x, x) for stubbornness, sensibility, ignored-ness and influence respectively.

5.7 Other

The Other role is defined for participants for whom values of certain characteristics cannot be determined. These include combinations of characteristics where either the stubbornness value is 0 or the sensibility value is 0. The most common case where stubbornness cannot be determined occurs when a participant changes his/her stance during the discussion. The most common case where sensibility cannot be determined occurs when there is not enough information in the contribution of the participant to perform a sensibility analysis.

6. Conclusion

While we are continually refining the model, our annotation of contentious discussions over the past year has stabilized to the degree where we do not expect many significant changes if we were to address a new contentious corpus. While we do not expect to find many additional roles, we continue actively to discuss the nature and organization of their underlying characteristics. We are building automated classifiers to recognize the characteristics and roles.

We plan to make the annotated corpus publicly available by the time of the LREC conference.

7. References

- Crosini, R.J. (2002). *The Dictionary of Psychology*. pp. 805.
- Jain, S; Hovy E.H. (2013). Determining Leadership in Contentious Discussions. *Proceedings of the International Workshop on Social Multimedia Research (SMMR 2013)* at the IEEE International Conference on Multimedia & Expo (ICME 2013). San Jose, CA.