Subjectivity in Social Networks

Ammar Raşid
Özyeğin University
ammar.rasid@ozu.edu.tr

Abstract

In this paper we study the properties of subjectivity in the context of online social networks. We explore the effect of subjectivity levels in appealing to audiences. We examine the relation between subjectivity and homophily within a contextual social network. We study the polarization effect of subjectivity as a measure of stubbornness in information diffusion models. We explore the patterns of subjectivity within communities.
Keywords: Subjectivity, Polarization, Homophily

1 Introduction

According to the oxford dictionary, subjectivity is the quality of being based on or influenced by personal feelings, tastes, or opinions. In philosophy, a subjective experience, or often referred to as *qualia*, is the unique experience an observer goes through upon encountering its objective analogue. For example, a red surface is an objective analogue. However, different observers can have different experiences when they examine the color of the same surface. In social sciences, subjectivity is often observed as the property of being a subject. Similar social configurations create similar interpretations of the world. While objectivity is about the objective analogue of the observed entity, subjectivity is about the plethora of experiences of different observers. Any disagreement, therefore, is rooted by the subjective nature of personal viewpoints. People who fail to appreciate their human inability to encompass all and every interpretation of a subject may tend to display biased attitude toward standpoints different from their own. This is referred to in social psychology as biased assimilation, where individuals readily accept confirming evidence, but are rather critical when provided with disconfirming evidence. Naturally, this would not be the case if arguments are expressed in objective manner, where there is no room for the subjective nature of personal opinions to manifest in disagreements. The objective analogue of a subject is shattered into a myriad of subjective interpretations. An individual's variance of assimilated viewpoints is a strong motive for that individual to appreciate the error-proneness of his subjective views, and, in turn, make him express his interpretations in less biased and a rather more objective manner. Social media platforms have been criticized for fostering echo-chamber effect. This goes in conjunction with homophily [5], which has been empirically verified by [4] in the context of online social networks. The echo-chamber effect traps individuals into a limited set of ideas, with monotonously increasing adherence to their dogmas. As a group of individuals get more narrowly focused on a subject, the more likely they will base their standpoints on personal interpretations, which implies a high level of subjectivity. We examine the potential empirical evidence of the mentioned hypothesis, which is stated in Hypothesis 1. In this paper, we explore various properties of subjectivity in the context of online social networks. To the best of our knowledge, this is the first study on subjectivity in that context. The goal of this paper is four-fold;

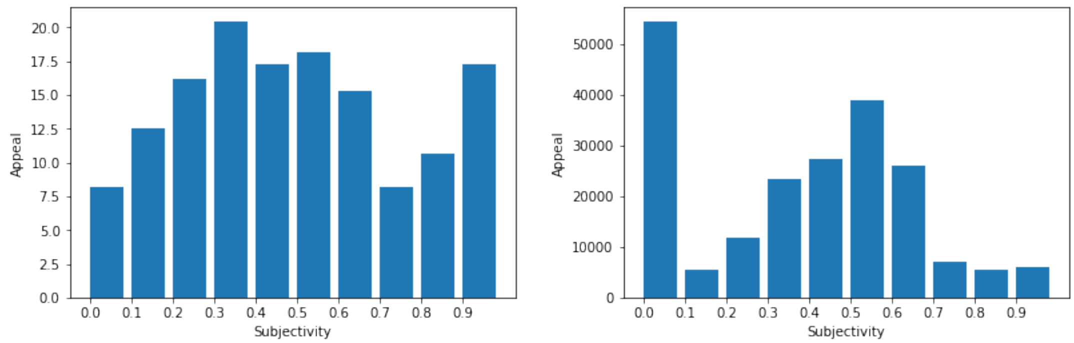
- We study the appeal effect of subjectivity.
- We examine the polarization-related effects of subjectivity in information diffusion models.
- We examine the relation between homophily and subjectivity.
- We study the subjectivity patterns within network-structure based communities, see hypothesis 2.

2 Methodology

2.1 Network Configuration

Similarity in interests is best modeled as an affiliation network. However, affiliations are often represented with high-level interaction elements, like a college or a region. In our paper, we model the affiliation as a reddit link. Reddit [1] is an online social media and opinion sharing platform. It is broken up into over a million *subreddits*, each covering a different topic. Users can post recursive comments on the same post. For example, a user u can comment on a post, then another user v can comment on user u's comment, and user w in turn can comment on user v's

[1] https://www.reddit.com/



(a) With normalization

(b) Without normalization

Figure 1: Appeal vs Subjectivity

comment on user u's comment, and so on. Each level of posts is identified by a unique *link* $l \in L$, where L is the set of interaction links. The set $C_u \subset C$ denotes the comments posted by user, and is a subset of the global set of all comments C.

Consider a network graph $G = (U, V)$ defined on a set of nodes V, and a set of edges E. An edge $e \in E$ between two nodes $u, v \in V$ is denoted as $e = \{u, v\}$. The set of links a user u has posted in is denoted as L_u. An edge $\{u, v\}$ exists if and only if $\Psi(u, v) \geq 0.25 and (u, v) \geq 0.5$, where Ψ is the *Bipartite Graph Reinforcement* similarity function, discussed in 2.2.

We use TextBlob [2] to calculate the subjectivity of each comment. The subjectivity score of a comment c is denoted by $\omega(c)$, whereas the subjectivity score of a node u is denoted by $\Omega(u)$

Definition 2.1 *The subjectivity of a node $u \in V$ in the network is the average subjectivity scores of comments posted by that user u:*

$$\Omega(u) = \frac{1}{|C_u|} \sum_{\forall c \in C_u} \omega(c)$$

2.2 Edges weights

We use a similar approach to the one used by [2], where they computed a user similarity space for stance detection. We compute the similarity between two users as the normalized summation of the two conditional probabilities $p(u|v)$ and $p(v|u)$. We model the similarity space as a bipartite graph in conjunction with graph reinforcement. Given a user u. We traverse to interaction link l to estimate the conditional probability $p(l|u)$, which is defined formally in equation 1. We then traverse the graph from l to the target user v to estimate the conditional probability $p(v|l)$, which is formally defined in equation 2. Using the maximum likelihood estimate, if there is one interaction link l between users u and v, then $p(v|u)$ the probability that v *maps* to u is $p(v|u) = p(l|u) \times p(v|l)$. However, since it is typically the case that two users are linked via multiple interaction links, we want to reinforce the similarity link between such two users. That is done in equation 3. The conditional probabilities $p(v|u)$ and $p(u|v)$ are then summed up and normalized by the number of mutual interaction links.

$$p(l|u) = \frac{\text{count of posts by } u \text{ in } l}{\text{count of posts by } u} \tag{1}$$

$$p(v|l) = \frac{\text{count of posts by } v \text{ in } l}{\text{count of posts in } l} \tag{2}$$

$$p(v|u) = 1 - \prod_{\forall u, v \in V, \forall l \in L} (1 - p(l|u)p(v|l)) \tag{3}$$

$$\Psi(u, v) = [p(v|u) + p(u|v)]^{\frac{1}{L_u \cap L_v}} \tag{4}$$

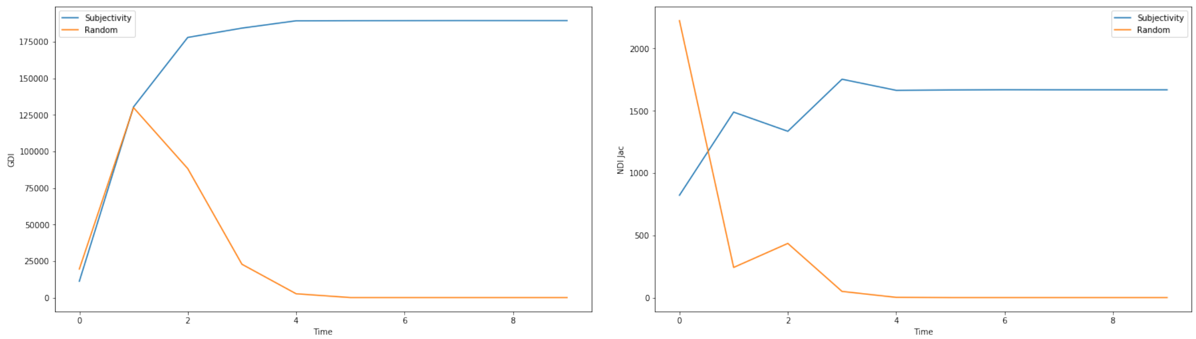
Any edge $e \in E$ that has less than 0.25 similarity according to this metric, is filtered out of the network. Only $6,883$ edges satisfied the BGR threshold. Whereas, when we tried only jaccard similarity, defined in equation 5, nearly $227k$ edges satisfied the threshold. However, upon filtering with *BGR* we

$$J(u, v) = \frac{L_u \cap L_v}{L_u \cup L_v} \tag{5}$$

3 Experiment

We collected $100k$ reddit comments from February, 2013. The largest subreddit in our data was *AskReddit*, with $16,143$ comments. We considered only the comments from the largest subreddit for network construction. The largest of 59 component in the network contained 871

[2]TextBlob: Simplified Text Processing, https://textblob.readthedocs.io/en/dev/



(a) Global Disagreement Index

(b) Network Disagreement Index with Jaccard

Figure 2: Subjectivity as stubbornness measure in information diffusion. Orange curve is the baseline uniform stubbornness 0.5

nodes, totalling 82.25% of the entire users V set, and $1,648$ edges, which form 92.58% of the entire edges E set. We considered only the largest component of the network for analysis.

3.1 Appeal effect of subjectivity

We observe the *appeal* scores reported by reddit users. Each post accumulates a score through prolonged public voting. We notice that moderate subjectivity levels, between $0.4 \geq \omega \leq 0.7$, harness the highest appeal scores. We also notice that high subjectivity, $\omega > 0.7$, are least common, yet their average appeal scores get higher as they increase. Whereas, low subjectivity scores, $\Omega < 0.1$, are most common, yet their average appeal score is low. The subjectivity scores, $0.7 \geq \omega \leq 0.8$, although significantly less common than subjectivity scores, $0.0 \geq \omega \leq 0.1$, they harness similar average appeal scores. These observations suggest that;

- The significant minority of opinions are expressed in highly subjective manner.
- People tend to like highly subjective, moderately subjective, and objective opinions, in that order.
- High level of subjectivity can be interpreted as assertiveness, which is a crucial element in mass influence.

3.2 Subjectivity in information diffusion

Dandekar et al. [1] examined the effect of stubbornness in biased assimilation studies in a social network configuration. We examine the validity of using subjectivity as a stubbornness measure in an information diffusion model. We use Network Disagreement Index (NDI), equation 3.1, weighted by Jaccard similarity 5, and Global Disagreement Index (GDI), equation 3.2, to quantify polarization level over a timeline.

Definition 3.1 *Given a graph G, and a vector of opinions $x \in [0,1]^{|V|}$ of individuals in V, the network disagreement index $\eta(G,x)$ is defined as:*

$$\eta(G,x) = \sum_{(u,v)\in V} J(u,v)(x_u, x_v)$$

Definition 3.2 *Given a vector of opinions $x \in [0,1]^{|V|}$ of individuals in V, the global disagreement index $\gamma(x)$ is defined as:*

$$\gamma(x) = \sum_{i<j} (x_i - x_j)^2$$

The seed probability for the information diffusion model was empirically chosen to be $p = 0.001$. At time t, a node v is activated if and only if at least one incoming edge u, v, coming from an active node u, and has a weight $J(u,v)$ larger than the *stubbornness* of v. Figure 2 shows that when subjectivity is considered as a stubbornness measure, the information diffusion stops at a significantly larger disagreement index, than the baseline of 0.5 stubbornness. This suggests that subjectivity can be a hindrance against diffusing new ideas within a community. This observation emanates from the fact that people tend to adopt ideas preached to them by trusted connections. Consider a directed link u, v, where node u is trying to activate node v. The more subjective the user v is, the more trust it requires from u to adopt its standpoint.

3.3 Homophily and subjectivity

Social media platforms have been criticized for fostering echo-chamber effect. This goes in conjunction with homophily [5], which has been empirically verified by [4] in

	min	max	mean	std	r	p-value
Sizes	25	434	233.92	130.87		
W CC.	0.00013	0.004	0.00176	0.0012	**0.84**	**3.9 e-28**
BGR CC.	0.004	0.082	0.04	0.027	**0.82**	**6.9 e-26**
Jac CC.	0.0013	0.035	0.018	0.011	**0.85**	**1.6 e-29**

Table 1: 1) Pearson correlation coefficients of clustering coefficients weighted with different similarity metrics, and average subjectivity scores. 2) Stats of average weighted clustering coefficients of subgraphs sample. *Sizes* refer to the sizes of subgraphs in the sample. W CC. is clustering coefficient weighted by number of mutual interaction elements $W = |L_u \cap L_v|$. BGR CC. is clustering coefficient weighted by Bipartite Graph Reinforcement. Jac CC. is the clustering coefficient weighted by Jaccard similarity.

the context of online social networks. The echo-chamber effect traps individuals into a limited set of ideas, with monotonously increasing adherence to their dogmas. As a group of individuals get more narrowly focused on a subject, the more likely they will base their standpoints on personal interpretations, which implies a high level of subjectivity. We examine the potential empirical evidence of the mentioned hypothesis, which is stated in Hypothesis 1.

Hypothesis 1 (H1): *Homophilous groups tend to display higher level of subjectivity than heterophilous groups.*

We use random walk with depth 3 to sample 100 subgraphs of the network. We then computed the average clustering coefficient of each sample, weighted by, edge weights (i.e. number of mutual interaction links W), Jaccard similarity, equation 5, and BGR similarity, section 2.2. We computed the pearson correlation coefficient r between each of the averages of weighted clustering coefficients, and average subjectivity scores of each subgraph. The subgraphs stats are shown in table 1.
There are three caveats to point out before we accept hypothesis 1. First, clustering coefficients reflect homophily when edges in the network are based on contextual similarity. Second, subjectivity calculation has its own share of error, as it is a computational model rather than a human-labeled score. Third, it is open for empirical analysis to tone the filter parameters while constructing the network. The sheer size and flexibility of contextual networks can hide important information.

3.4 Subjectivity within the same communities

Subjectivity levels displayed by members of the same intellectual communities can be influenced by the nature of the subjects shared by that community. We examine the hypothesis that individuals within the same intellectual community display similar levels of subjectivity compared to general levels of subjectivity of the network. We use the community detection algorithm proposed by M. Girvan and M. E. J. Newman [3] to label nodes in our graph with their corresponding community. Due to the nature of our network configuration, communities reflect topics of interests of individuals within them, which make them applicable for testing hypothesis 2.

Hypothesis 2 (H2): *Individuals in the same community tend to display similar subjectivity levels.*

The algorithm detects 11 communities in our network. We compute the standard deviation of average subjectivity scores of each community. We then run a one-tailed t-test to detect if there is a significant difference between the average of standard deviations of subjectivity scores of each community μ_σ, and the standard deviation of subjectivity scores of the entire population, which is $\sigma = 0.243$.

$$\mu_\sigma = \frac{1}{|C|} \sum_{\forall c \in C} [\frac{1}{|c|} \sum_{\forall u \in c \subset V} \Omega(u) - \bar{\Omega}_c] \qquad (6)$$

C is the set of communities, $\Omega(u)$ is the subjectivity score of user u, $\bar{\Omega}_c$ is the average of subjectivity score of all users in community c. $|S|$ is the size of a set S.

$$H_0 : \sigma = \mu_\sigma$$
$$H_a : \sigma > \mu_\sigma$$

The t-score of the test was -5.74, with a *p-value* of 0.0002. Therefore, we can reject the null hypothesis H_0. This suggests that individuals within the same community display similar levels of subjectivity than cross-community individuals.

4 Conclusion

In this paper we examined various properties of subjectivity in the context of online social networks. We showed subjectivity effect in appealing to audiences. We examined the reliability of subjectivity as a stubbornness measure in information diffusion, and highlighted its polarizing effect in conjunction with biased assimilation.

We examined the relation between subjectivity and echo-chamber effect, which suggests implications that are potentially important in spreading new ideas and fighting fanaticism. We showed that communities are the gateways to evaluating and influencing the subjectivity levels of individuals. We hope to explore the predictive power of subjectivity in applications like stance detection and false news identification.

References

[1] Pranav Dandekar, Ashish Goel, and David T. Lee. Biased assimilation, homophily, and the dynamics of polarization. *Proceedings of the National Academy of Sciences*, 110(15):5791–5796, 2013.

[2] Kareem Darwish, Walid Magdy, and Tahar Zanouda. Improved stance prediction in a user similarity feature space. pages 145–148, 07 2017.

[3] M. Girvan and M. E. J. Newman. Community structure in social and biological networks. *Proceedings of the National Academy of Sciences*, 99(12):7821–7826, 2002.

[4] Bas Hofstra, Rense Corten, Frank van Tubergen, and Nicole B. Ellison. Sources of segregation in social networks: A novel approach using facebook. *American Sociological Review*, 82(3):625–656, 2017.

[5] Miller McPherson, Lynn Smith-Lovin, and James M Cook. Birds of a feather: Homophily in social networks. *Annual Review of Sociology*, 27(1):415–444, 2001.