Equality and Social Mobility in Twitter Discussion Groups

Katherine Ellis
Department of Electrical and Computer Engineering
UC San Diego
La Jolla, CA
kellis@ucsd.edu *

Moises Goldszmidt
Samsung Research America (SRA)
Mountain View, CA
moises.goldszmidt@gmail.com

Gert Lanckriet
Department of Electrical and Computer Engineering
UC San Diego
La Jolla, CA
glanckriet@ucsd.edu

Nina Mishra
Computer Science Dept.
Stanford University
Palo Alto, CA
nmishra@gmail.com

Omer Reingold
Samsung Research America (SRA)
Mountain View, CA
omer.reingold@gmail.com

ABSTRACT

Online groups, including chat groups and forums, are becoming important avenues for gathering and exchanging information ranging from troubleshooting devices, to sharing experiences, to finding medical information and advice. Thus, issues about the *health* and *stability* of these groups are of particular interest to both industry and academia. In this paper we conduct a large scale study with the objectives of first, characterizing essential aspects of the interactions between the participants of such groups and second, characterizing how the nature of these interactions relate to the health of the groups. Specifically, we concentrate on Twitter Discussion Groups (TDGs), self-organized groups that meet on Twitter by agreeing on a hashtag, date and time. These groups have repeated, real-time meetings and are a rising phenomenon on Twitter. We examine the interactions in these groups in terms of the *social equality* and *mobility* of the exchange of attention between participants, according to the @mention convention on Twitter. We estimate the health of a group by measuring the retention rate of participants and the change in the number of meetings over time. We find that social equality and mobility are correlated, and that equality and mobility are related to a group's health. In fact, equality and mobility are as predictive of a group's health as some prior characteristics used to predict health of other online groups. Our findings are based on studying 100 thousand sessions of over two thousand discussion groups over the period of June 2012 to June 2013. These finding are not only relevant to stakeholders interested in maintaining these groups, but to researchers and academics interested in understanding the behavior of participants in online discussions. We also find the parallel with findings on the relationship between economic mobility and equality and health indicators in real-world nations striking and thought-provoking.

Categories and Subject Descriptors

I.5 [**Pattern Recognition**]: General

Keywords

Social Networks; Generative Models; Graph Analysis

1. INTRODUCTION

Online groups are of interest to researchers for a variety of reasons. First, there is a wealth of information and expertise on a variety of topics being discussed in online groups. In some cases groups cover solutions on how to deal with technical issues that are otherwise hard to find. Some discussions can also provide a sense of the pulse and sentiment on different socio-political issues. Similarly, information and opinions on books, music and movies are often shared. Second (and the motivation behind our work), these groups provide an incredible new telescope to study the nature of human interaction and behavior, albeit in a particular environment with all the biases inherent to the media (online interactions), at scale. From a pragmatic point of view, stakeholders would like to understand how to keep the groups active and healthy. From a scientific point of view, answers to these questions provide some insight into the nature of human interactions.

Online communities manifest in a number of forms — some examples that have been studied in the literature include Internet Relay Chats [25], Yahoo groups [3] and Ning groups [17]. In this work we focus on Twitter discussion groups (TDGs) [9]. TDGs are synchronized, real-time conversations that occur repeatedly (sometimes periodically), and with sessions usually lasting 1-2 hours. Participants agree on a hashtag, date and time, and then append the hashtag to their tweet so others in the group can follow the conversation at that date and time using the standard Twitter interface. Topics of these groups are numerous and varied. For example, there are support groups for postpartum depression and borderline personality disorder. Other groups enable participants to connect with others with similar hobbies and interests like skiing, photography, movies,

*Work done while Katherine, Moises, Nina, and Omer were at Microsoft Research, Silicon Valley.

WSDM 2016, February 22-25, 2016, San Francisco, CA, USA

wines and food. Some of these groups have a moderator (which may change with meetings), some invite medical experts and other personalities, and some are organized around sport team events. The richness of subject matter and popularity of Twitter makes it a compelling social network to study. Additionally, the real-time nature and limited time window of each TDG session is more similar to real-world group meetings than some other online forms of discussions such as forums and message boards.

Our first goal in this paper is to characterize the interactions of participants in a TDG, in terms of the attention exchange through the "@mention" mechanism. In particular, we characterize the interaction between the participants in terms of social *equality* and *mobility*, where the analogous currency is the accumulation and distribution of attention. Mechanistically, participants in Twitter give and receive attention via an "@mention." A user directs a tweet to a specific individual by @mentioning them to strike a conversation, to reply to a question or comment, or by retweeting the content of that user. In all these cases a user is explicitly signaling the importance of the other user and/or the content (s)he has generated. In groups with high equality the attention is divided more evenly between participants. In groups with bigger degrees of social mobility, members are able to move between different levels of attention from session to session. These concepts will be more formally defined in section 5.

Our second goal is to establish the relationship between these social equality and mobility characteristics and the *health* of a TDG. Specifically, we measure group health by the retention rate of participants over time and stability in terms of the number of sessions per period. Intuitively, a group is healthy if many people continue to attend sessions and sessions continue to take place regularly. We take inspiration from the socioeconomics literature where research has demonstrated that countries with more egalitarian income distribution and with better possibilities of economic mobility tend to be healthier [29, 14, 12, 26]. Health in these studies is measured in a variety of ways including life expectancy, infant mortality, and homicide rates.

In this paper we present the results of a large scale study of over two thousand Twitter discussion groups with over 100 thousand sessions. To the best of our knowledge this is the first large scale study to quantify the relationship between social equality and mobility to the health of online groups at scale. The main contributions of our study are:

1. A probabilistic model that captures interaction patterns between users (based on the models introduced in [1]). We establish that each participant in the TDG can be profiled by a probability distribution on three archetypical patterns of interaction. Archetype A is the one that receives the most attention (mentions), archetype B is the one that gives the most attention, and archetype C is the less active. See Section 4 for further discussion and definitions.[1]

2. We establish that mobility and inequality are statistically correlated and we find that (a) there are high degrees of inequality in TDGs (the top 10% gets 40% of the attention) but there are high degrees of mobility (so it is not always the same 10% that receives the bulk of the attention).

3. Higher mobility and lower inequality correlate to and predict healthier groups. In fact, equality and mobility are so powerful that they are as predictive of a group's health as other graph-based statistics used to predict health. We observe similar results even when concentrating on natural subsets of TDGs such as cohesive groups (where there are high levels of interconnectivity between frequent participants).

[1]One may try to measure attention by simply counting the number of @mentions received by a participant (which can lead to definitions of equality and mobility). Such a "first approximation" is limited in that it fails to take into account the source of each @mention (*e.g.*, being mentioned by a central participant may be more indicative that the participant being mentioned is more central too).

2. RELATED WORK

We next describe how our work compares to prior research in machine learning, social networks, and sociology.

Role Identification in Networks.

A variety of techniques aim to discover core and periphery structures in networks. Borgatti and Everett [6] investigate techniques for identifying core and periphery nodes. Subsequent work [23] assigned nodes along a continuous spectrum from those that lie in the heart of a core to those that lie in the outer reaches of the periphery. Other methods core/periphery identification methods include [11, 13].

Another line of work that identifies roles in networks is the Mixed-Membership Stochastic Block model [1]. In this model, each node is modeled by a distribution over k types — these types are defined by a square $k \times k$ matrix indicating the likelihood that type i communicates with type j. One of the benefits of this model is that a user can probabilistically take on every role (as opposed to being assigned exactly one), and it seamlessly weights the importance of each role in the communication. We adapt this model for our setting by using it to identify the different roles of participants in a session (we expand upon this in Section 4). Note that there is extensive literature in the sociology community that describes the functional roles that individuals play in groups [4]. For example, there are task roles (*e.g.*, opinion seeker, evaluator), social and maintenance roles (*e.g.*, supporter, harmonizer) and individual Roles (*e.g.*, aggressor). While these definitions are quite compelling, due to the scale of data that we process we require automated techniques for role identification.

Group Growth.

A large-scale study of 44K Yahoo groups was conducted in [3]. Yahoo groups differ from Twitter discussion groups in that the former are more akin to a forum without predesignated meetings, while the latter are real-time meetings. One interesting discovery made in [3] is that people who eventually become heavily-engaged receive preferential treatment from the first message that they post (see also [16]). This finding may be an early indication that mobility, or ability to reach the core, is signaled in the first communication with a group.

Kairam et al [17] study Ning groups which are more akin to forums where people post content at will and not within predesignated windows of time. The decline of a Twitter discussion group is thus more readily apparent. In contrast, [17] define group death as a group that stops adding new members. In addition, while the object of study is friendship links in the Ning graph, in our work, the object of study is @mention, a finer granularity interaction that can encourage people to remain in the conversation.

Other factors are also known to affect why people join and leave online communities. A person with more friends in a group is more likely to join [18] and the more interconnected their friends are, the more this likelihood increases [2]. This finding may also be viewed as a signal of the importance of mobility in that if a person has many friends already in the core of a group, it may increase their chances of becoming core members. On the opposite side, infrequent communication hampers group responsiveness [19] while excessive communication is perceived as message overload [7, 15].

Twitter Discussion Groups.

The initial discovery of Twitter groups was reported in [10]. In that work, a *group* was also represented as a hashtag, but the definition was more strict. In addition to requiring that the group met in a synchronized window of time, their work also required that meetings occur with predictable regularity, *i.e.*, every t days, and that the group was cohesive in the sense that the people who attended the most meetings communicated with each other. We investigate synchronized and cohesive groups, but drop the requirement that groups meet with strict regularity. Discussion groups were defined in [9], where the goal was to rank chats according to a specified query. This paper focuses on modeling social structure and predicting group health.

Equality and Mobility.

Prior work has investigated the differences between more and less equal societies [22, 20, 28]. For example, countries with more egalitarian income distributions tend to have higher life expectancy rates [27], lower infant mortality rates [29], lower homicide rates [14, 12, 26] and more social mobility [5, 30]. These correlations were discovered at national scales, *e.g.*, the US has very inequitable income differences and also lower life expectancy. While Twitter discussion groups may lack the diversity and intricacies of real-world economies, we can see them as a microcosm of society in which people form relationships, exchange ideas and participate in a community. The economic constructs in these works have inspired us to measure the exchange of attention in Twitter (via the @mention) as a sort of currency. In our work, we ask whether the correlations found at national levels are mirrored in the Twitter equivalents of these measures. (No causal claims are made.)

3. REPRESENTATION OF TWITTER DATA

In this section we describe the process of taking raw tweets and (a) identifying the sessions that constitute discussion groups, (b) representing the exchange of attention amongst the participants in each session using the @mention graph, and (c) modeling archetypical patterns of interaction and participant profiles.

3.1 Identifying TDG sessions

In this study we follow [9] and regard TDGs as self-organized and synchronized group discussions focused on specific topics that use Twitter as a platform to meet with some regularity. Logistically, members agree on a hashtag and a meeting time (*e.g.*, 3pm Pacific Time on Sundays) to discuss a subject of interest. A participant indicates his or her involvement in the discussion by including the designated hashtag in each of their tweets. We call a specific instance, or meeting, of a given group a *session*.

Operationally, we define a session as a two-hour window in which at least 20% of the week's tweets using the hashtag occur within the window [9]. For our study, we restricted our analyses to TDGs that had at least 12 sessions in a 12-month period. In addition, TDGs should contain between 25 and 200 participants [2] in each session and at least 5 participants who attended at least 5 sessions.

Given the above criteria, we conducted a search on the complete corpus of Twitter for a 12 month period (June 2012 - June 2013) and identified a total of 2418 different TDGs consisting of over 113,000 individual sessions. This task required the analysis of petabytes of data which we achieved using the SCOPE language [8] on a large distributed computing cluster.

3.2 Representing sessions

For our analysis, we strip away the semantics and content of the conversation and focus on the structure of interactions by representing each session by an @mention graph. In Twitter, a user explicitly *mentions* another user by appending an "@" symbol to the name of the other user in the text of the tweet (For example, a tweet mentioning user Y will include the text "@Y"). We separate mentions into three types: replies, retweets, and other tweets. When user X replies to a tweet created by user Y, their new tweet will mention user Y. Similarly, when user X retweets (an action similar to forwarding an email) a tweet created by user Y, X's new tweet will mention user Y. In general, the author of a tweet can mention user Y by adding the text "@Y" anywhere in the tweet. In each of these situations, we view user X mentioning user Y as user X giving attention to user Y — either by responding to something user Y said, promoting something user Y said, or specifically naming user Y.

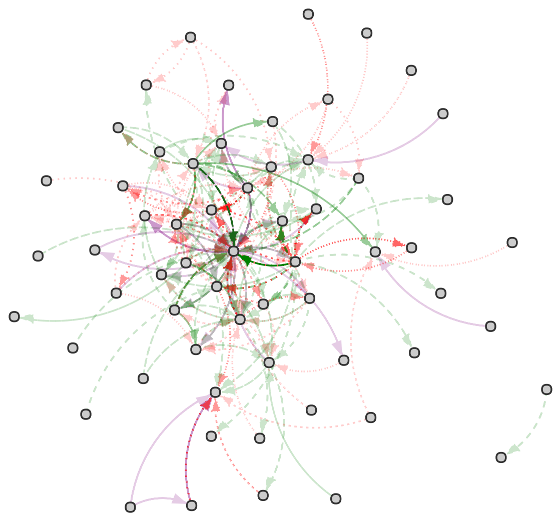
An @mention graph is defined as a graph $\mathcal{G} = (\mathcal{N}, \mathcal{E})$ where the nodes $\mathcal{N}$ are participants in the session and the edges $\mathcal{E}$ are @mentions between participants. For example, if participant X mentions participant Y, there will be an edge from X to Y. If participant X mentions multiple participants, there will be an edge from X to each participant mentioned. In our analyses we differentiate between replies, retweets and other tweets. We decompose the set of edges into $\mathcal{E} = \{\mathcal{E}_{RT}, \mathcal{E}_{Re}, \mathcal{E}_O\}$, where $\mathcal{E}_{RT}$ are retweets, $\mathcal{E}_{Re}$ are replies and $\mathcal{E}_O$ are all other tweets containing an @mention. In addition, each edge $e \in \mathcal{E}$ has an integer weight corresponding to the number of times the mention occurred during the session. A visualization of the @mention graph for an example TDG session is shown in Figure 1 (a).

4. MODELING SESSIONS

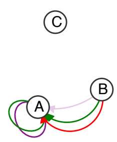
To analyze equality and mobility in terms of attention (@mention) exchanges, we generate a latent model of interactions based on the explicit evidence in the @mention graph. Our model is a generalization of a well known generative probabilistic model called the *mixed membership stochas-*

[2]We focused on groups small enough to allow interactions between participants – too large of a group inhibits real discussion.

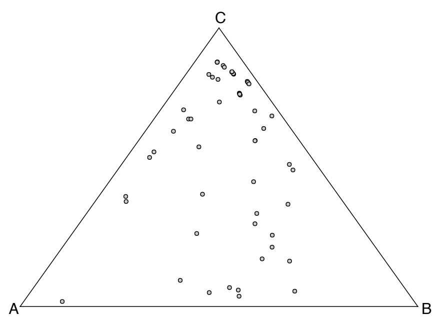
Figure 1: The @mention graph and learned model for an example session of the TDG #mtos, "Movie Talk on Sunday".



(a) The @mention graph. Green (dashed) edges are retweets, red (dotted) edges are replies, purple (solid) edges are normal @mentions.



(b) The archetype matrices Λ.



(c) The membership vectors π on the simplex.

tic blockmodel [1, 24]. The next few subsections describe the model and its output in the form of a set of roles for the participants in the TDG sessions. These roles form the basis for the characterization of mobility and inequality.

4.1 The interaction model

For modeling the interactions between the participants, given the @mention graph, we extend the generative model in [1] as follows: each participant is characterized by a latent vector π that represents a distribution over the different *roles* that the participant can take in terms of his/her actions (e.g., initiating or receiving an @mention) in the session. We call these roles *archetypes*. In addition, we want to capture (a) the different types of edges that @mention graph exhibits (i.e., replies, retweets and other tweets) and (b) the number of times specific mentions are made. To this end we generalize the model to contain 3 matrices corresponding to the 3 modes of interaction. Each entry in each of one of the matrices Λ_i represents the probability that a participant of archetype X interacts (in a specific way) with a participant of archetype Y. Furthermore, the entries of the matrix contain the intensity parameter of a Poisson distribution, which encodes the expected number of times that the different archetypes interact. Thus the activity in an @mention graph $\mathcal{G} = (\mathcal{N}, \mathcal{E})$ where $\mathcal{E} = \{\mathcal{E}_{RT}, \mathcal{E}_{Re}, \mathcal{E}_O\}$ is modeled according to the following generative process:

- For each node $i \in \mathcal{N}$
  - Draw a K dimensional latent archetype vector: $\pi_i \sim \text{Dirichlet}(\alpha)$
- For each pair of nodes $(i, j) \in \mathcal{N} \times \mathcal{N}$:
  - Draw archetype indicator for the initiator, $z_{i\to j} \sim \text{Multinomial}(\pi_i)$.
  - Draw archetype indicator for the receiver, $z_{i\leftarrow j} \sim \text{Multinomial}(\pi_j)$.
  - For each edge type $\mathcal{E}_m \in \mathcal{E}$:
    * Sample the weight of the edge, $e_{ij,m} \sim \text{Poisson}(z_{i\to j}\Lambda_m z_{i\leftarrow j})$.

where $\Lambda_m(i,j)$ is the Poisson intensity parameter, *i.e.*, the expected number of type m mentions of archetype j by archetype i. Our use of a generative model to uncover the latent roles of the participants, as opposed as for example, counting directed edges in the @mention graph, provides the ability to automatically take into account the importance of the edge in relation to the role of both the participants (nodes) at the extremes of each edge.

4.2 Fitting session models

We used variational message passing [31] in the Infer.NET package [21] to fit the parameters of our model for each TDG session (*i.e.*, the type vector π for each participant and the interaction matrices Λ). Although there are a large number of sessions (>113,000), parallelization was straightforward as we fit a model for each session independently.[3] We used a combination of Dryad/Linq [32] for managing the distributed (parallel) computation of the models on a cluster of 20 servers.

The original model as well as our generalized version has a free parameter: the number of archetypes K. This is a model selection problem, and it is usually resolved using a penalized likelihood approach such as the Bayesian Information Criterion (BIC) [1]. The tension is that if K is too small, we may be unable to express important aspects of the interaction, while if it is too large we may be overfitting. For our analysis, we also need the same K in all models to enable the computation of statistics of groups (across sessions) and to be able to compare and contrast among the different groups.[4] we performed an exploratory analysis over a sample covering over 10% of the total number of sessions. Our observations were that the majority of the models maximized BIC for $K = 3$, and over 75% of the models maximized BIC for $K < 4$. We therefore settled on $K = 3$.We observed in our explorations that increasing K by 1 has the effect of identifying subgroups inside the session rendering two (essentially) indistinguishable archetypes in terms of the mentioning action. Thus, we don't lose any information relevant to our analysis by forcing $K = 3$ in these cases. Similarly, the addition of an archetype in those

[3]Note that fitting a model per TDG (as opposed to by session) would remove traces of the evolution of the archetypes over time.
[4]This need is different, for example, from trying to analyze in more detail the roles of participants in an isolated session.

sessions where BIC was maximized by $K = 2$, forces the split of one archetype into two very similar archetypes in terms of the mentioning action.

4.3 Naming the archetypes

In addition, we must address the identifiability issue in comparing latent models between different sessions and groups. To this end we examined the expected number of mentions for each archetype (according to the Λ matrices), and devised a naming convention in order to match archetypes between sessions. We use the letters A, B, and C to denote the three archetypes. Formally, the expected number of mentions received by archetype k is computed as:

$$E[k \text{ is mentioned}] = \sum_m \sum_{l \neq k} \Pi(l)\Lambda_m(l,k),$$

where $\Pi(l)$ is the prior probability of archetype l, *i.e.*, $\Pi(l) = \frac{1}{N}\sum_{i=1}^{N} \pi_i(l)$. Similarly, the expected number of mentions sent by archetype k is

$$E[k \text{ mentions}] = \sum_m \sum_{l \neq k} \Pi(l)\Lambda_m(k,l).$$

We define A to be the archetype that has the highest expected number of mentions in a session. On average, A displays a significantly higher probability of being mentioned than any other archetype. The two other archetypes are not as well distinguished by their expected number of times being mentioned, but are quite different on average when considering their expected number of times mentioning others. We thus define B to be the archetype with the higher expected number of mentions (among the two remaining). The remaining archetype is named C.

In our data analysis over the 113K sessions, we found that 100% of groups have E[A is mentioned] more than double E[B is mentioned], while 57% of groups had E[B mentions] more than double E[C mentions] and 73% of groups had E[B mentions] at least 50% more than E[C mentions].

4.4 A qualitative look at session models

The sessions models learn abstractions describing prototypical patterns of interaction that are directly interpretable. Figure 1 shows an example session for the TDG "#MTOS" (Movie Talk on Sunday). Each session of Movie Talk on Sunday is centered on a set of movie-related questions. Throughout the session, a host or moderator tweets questions and participants in the chat answer the questions and discuss their answers. In this session of "mtos", there was a guest host who wrote the list of questions. He is introduced at the beginning of the session by the group's founder. Participants often reply to him when answering questions (although answers are often provided using only the hashtag #mtos, without mentioning anyone), and thank him at the end of the session by mentioning him. As participants respond to questions throughout the session, they often retweet answers they agree with, or reply to a user to engage more about an answer they made. In part (b) of Figure 1 the archetypes A, B and C are represented graphically in terms of the probabilities of mentioning other archetypes. We see that the most common modes of interaction are archetype B mentioning archetype A, or archetype A mentioning another archetype A. The most likely type of mention is a reply from B to A. In part (c) the latent profiles π of each participant in the session are visualized on the simplex. The guest host is located at the bottom left corner of the simplex, indicating that he is the most dominant archetype A participant in the session. This is expected, as the host he sends and receives the most attention. Participants along the left edge of the simplex have low weight on archetype B — they may be mentioned frequently but rarely mention others. For example, one participant that falls in this location provided some answers that others agreed with and retweeted. Participants along the right edge of the simplex have low weight on archetype A — they may mention others often but are rarely mentioned. One participant that falls near the middle of this area is a user who often retweeted answers to signify agreement. Notice that the bottom right corner is fairly empty of participants, indicating that in this session, there were no participants who mentioned others often but were never mentioned themselves. This seems to indicate that if you initiate discussion in this community you will receive some attention in exchange.

5. STATISTICS

In order to examine concepts of social mobility and inequality in TDGs, we define a set of statistics based on the session models learned in the previous section.

Equality.

We propose a set of statistics characterizing whether attention is distributed equally amongst participants in the discussion (or to what extent it is concentrated in the hands of the privileged few). The intent is to mirror notions of economic inequality, namely, to what extent the total wealth in a country is controlled by the wealthiest X% of people (commonly "the 1%"). We propose three sets of statistics with the objective of obtaining analogous metrics of inequality over TDG participants. These statistics are computed on the session models described in the previous section and averaged over all the sessions of a TDG.

The first statistic consists of the proportion of the total weight on archetype k that is held by the 10% of participants who have the largest weight on archetype k:

$$\text{Top10}_k = \frac{\sum_{i \in S} \pi_i(k)}{\sum_{i=1}^{N} \pi_i(k)},$$

where $S =$ is the set of $N/10$ participants with largest π_k. Note that as Top10_k approaches 1, more attention is focused on a select few. A small value suggests that attention is spread more evenly.

We explore two other ways to measure disparity in attention. For each archetype, we can first order participants according to their weight on the archetype, and then identify the top fraction of participants required to reach 50% of the weight on the archetype. A third simple set of statistics is the average weight on each archetype. A complete list of inequality statistics can be found in Table 1.

Mobility.

Similarly to inequality, we are interested in characterizing the ability of participants to change their dominant archetype over time. For example, a change in proportion from archetype C to archetype A in the profile of a participant will signal a change in the status of that participant. This also mirrors the notion of mobility in economic terms that characterize ability to move (usually upward) in income over time. Mobility statistics are computed over a sequence of session models, and thus can only be measured for participants that attend multiple sessions. We compute mobility statistics over the set N_a of participants who attended at least five sessions. As with the equality statistics, we explore three sets of measures of mobility.

The first statistic is the average range of participants' weight on a specific archetype, computed as,

$$\text{AvgRange}_k = \frac{1}{N_a}\sum_{i=1}^{N_a} \max_t \pi_{i,t}(k) - \min_t \pi_{i,t}(k).$$

It follows that larger ranges signal greater mobility in participants' archetype over time.

The second statistic is the probability that a given participant will have at least half their membership in a specific archetype (most notably the archetype A),

$$\text{Half}_k = Pr(\pi(k) >= 0.5) = \frac{1}{N_a}\sum_{i=1}^{N_a} \mathbb{I}[\pi_i(k) >= 0.5]$$

Since A is the archetype that receives the most attention, Half_A measures the fraction of participants that receive a lot of attention in at least one session.

The third and final set of statistics encodes the probability of transitioning between archetypes. We construct a state space model over the archetypes and build a Markov chain to describe the probabilities that a participant will transition between archetypes in subsequent sessions. The model has four states: A, B, C and not present. The transition probabilities between each state are calculated by summing the changes between each participant's type from one session to the next (only participants who attended at least 5 meetings were included). Higher between-state transition probabilities indicate higher mobility between archetypes. For example a high $Pr(B \to A)$ indicates "upward" mobility, as a participant is more likely to move from the less attention-gathering to the more attention-gathering archetype.

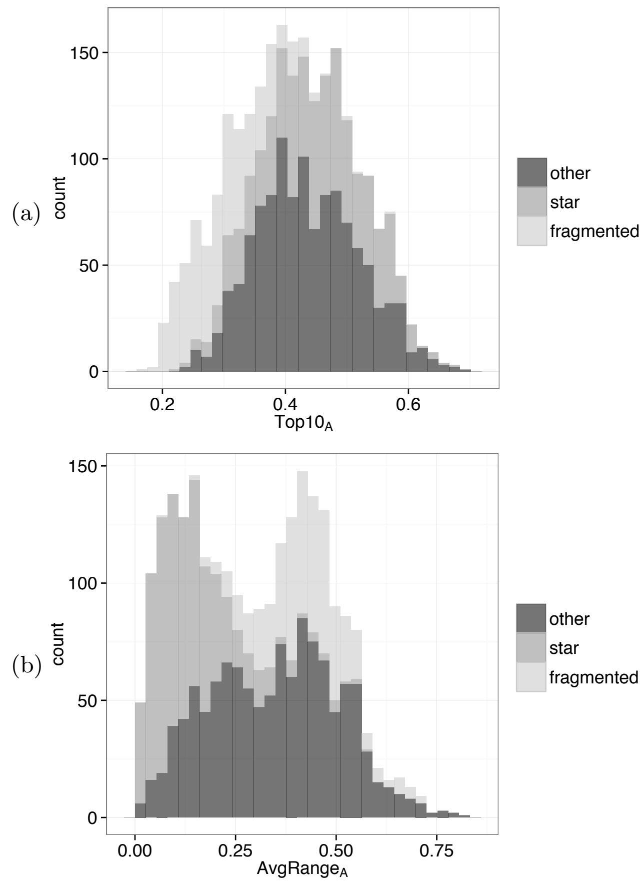
Additional Statistics.

In addition to mobility and equality, we consider other statistics derived from the session models. These include the expected number of mentions of each tweet type (reply, retweet, etc.), and the probability of each archetype being mentioned. Details can be found in Table 1. Further, we compute several statistics independently of the models. These include the number of sessions, average size of sessions, clique sizes and other graph-based statistics that have been successfully used in previous work to predict growth and decline of online groups [17]. We will compare these additional statistics to the model-based mobility and equality statistics in upcoming sections.

6. CHARACTERIZING EQUALITY AND MOBILITY IN TDGS

Figure 2 summarizes inequality and mobility statistics for the TDGs we studied. Overall, we found that groups exhibit significant levels of inequality. On average, the top 10% of archetype A held 41% of the weight on archetype A. At the same time we also see that many groups exhibit high levels of mobility. On average, the range of weight on archetype A is 0.3, and in many groups a significantly higher range is observed. In fact, the distribution is highly bimodal.

Figure 2: Distribution of (a) inequality and (b) mobility statistics for all 2418 TDGs. Groups are separated by graph structure into star, fragmented, and other groups.



Star and fragmented @mention graphs.

We find that the bimodality in the distribution of mobility across TDGs can be explained by identifying two subtypes of very different structures of @mention graphs. We define star graphs as @mention graphs in which over X% of incoming edges are directed to one or two participants. We define fragmented graphs as @mention graphs in which less than Y% of participants are in the largest connected component. For our analyses we set X to 90% and Y to 75%. With these subclasses of graph structures, the bimodality of the mobility statistic seen in Figure 2 can be explained by grouping TDGs into fragmented and star groups. We see that groups corresponding to star graphs display the highest degree of inequality and the lowest degree of mobility, while the fragmented groups display the opposite characteristics.

We notice that many star graphs correspond to groups that center around a celebrity or a television show. The celebrity might tweet something, and many people retweet or reply to them. Similarly, when the TV show airs, many people will mention the TV show handle in their tweets discussing the show. In these cases the celebrity-like participant commands an unequal share of the attention, and it is very unlikely for any other given participant to move into such a position of power. Conversely, many fragmented graphs correspond to groups that center around a common interest, such as a sports team. For example, when the team plays, many people will tweet to each other about the game, al-

Table 1: List of all statistics.

		Feature name	Description
Model Statistics	Inequality	$\Pi[k]$	Average weight on archetype k
		$Weight50_k$	The fraction of participants required to reach 50% of the weight on archetype k
		$Top10_k$	The proportion of the total weight on archetype k that is held by the top 10% of participants
	Mobility	$AvgRange_k$	The average range of weight on archetype k
		$Half_k$	The probability that a given participant will have at least 0.5 weight on archetype k
		$Pr(k \to l)$	Probability of transitioning from archetype k to archetype l from one session to the next
		$\mathbb{E}[k$ is mentioned$]$	Expected number of times archetype k is mentioned
		$\mathbb{E}[k$ mentions$]$	Expected number of mentions made by archetype k
		$\mathbb{E}[k$ recevies a mention of type $l]$	Expected number of times archetype k is mentioned with a certain tweet type (retweet, reply or normal)
		$\mathbb{E}[k$ mentions of type $l]$	Expected number of times archetype k mentions with a certain tweet type (retweet, reply or normal)
		Pr(k is mentioned)	Probability that a given tweet mentions archetype k
		Pr(k mentions)	Probability that a given tweet is sent by archetype k
Other Statistics		numSessions	Number of sessions between 25 and 200 participants in the 12 month period we observed
		avgN	Average number of participants in each session
		ratioActiveN	Ratio of the number of participants who attended at least 5 meetings over the average number of participants
		avgAttendance	Average number of meetings participants attended
		density	Average number of tweets per participant
		avgTime	Average number of days between sessions
		stdTime	Standard deviation of the number of days between sessions
		maxClique	Size of the maximal clique
		nClusters	Number of disconnected clusters
		maxCluster	Size of largest cluster
		cliqueRatio	Fraction of participants in the maximal clique
		clusterRatio	Fraction of partcipants in the largest connected cluster
		centralityDegree	Average degree of centrality
		closeness	Average closeness (transitivity)

though there is no central participant leading discussion. In these cases attention can be spread more equally among participants and upward mobility in terms of attention may be more easily achieved. Figure 3 shows a few example @mention graphs belonging to these subtypes.

6.1 Correlation of equality and mobility

We find that mobility and equality in TDGs are not independent. Using $Top10_A$ to represent inequality and $AvgRange_A$ to represent mobility[5], the coefficient of correlation between inequality and mobility is -0.512[6]. This negative correlation between inequality and mobility mirrors the relationship observed in socioeconomic studies — as inequality grows, it becomes more difficult to move between social strata. This relationship is intuitive when considering the "extreme" star and graph structures discussed in Section 6. However, concern that groups with these extreme graph structures may not represent "true discussion" (*i.e.*, groups that are so tightly organized around a central figure have limited interaction) leads us to consider another subtype of group that is more aligned with our traditional notion of a discussion-based community. Toward this end, we define a subtype of TDGs called cohesive groups.

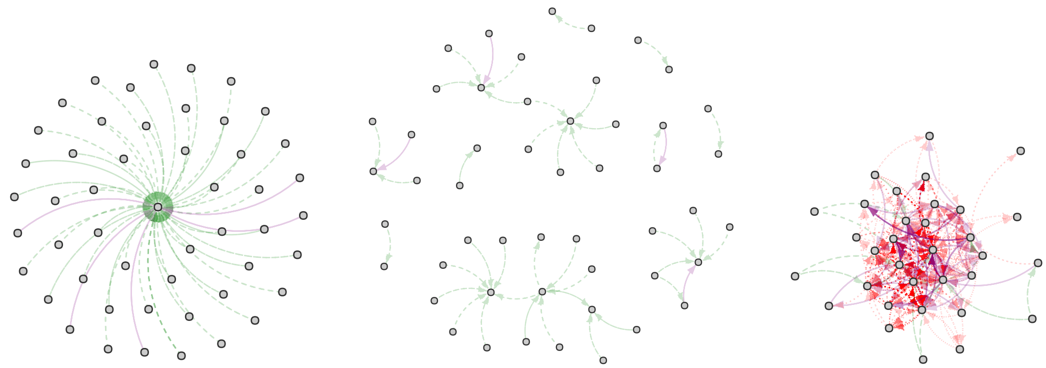
Cohesive groups.

Cohesive TDGs have a solid core of members that interact among themselves and persist between sessions. More specifically, we define a group as "cohesive" if the 5 participants who attended the most sessions (over the course of a year) have at least 4 shared edges in the @mention graph representing every session. These "cohesive" groups capture our idea of a discussion-based group, requiring consistent interaction between several key participants. This cohesion measure has been used previously to identify Twitter groups [10]. When we restrict our analysis to only cohesive TDGs, we find that the inverse relationship between inequality and mobility holds — in fact, it is strengthened. The coefficient of correlation between $Top10_A$ and $AvgRange_A$ is -0.661. Figure 4 graphically displays this relationship. In Section 7, as we investigate how mobility and equality affect group health, we will pay special attention to the cohesive groups.

[5]We chose $Top10_A$ and $AvgRange_A$ as intuitive measures of inequality and mobility; the alternative statistics listed in table 1 show similar correlations.

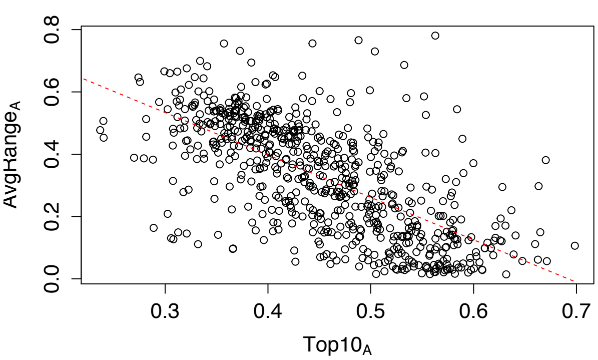
[6]Correlation for star and fragmented groups is -0.324 and -0.068, respectively.

Figure 3: Examples of @mention graphs of three subtypes of TDGs. “Star” TDGs (a) have one (or two) participant(s) mentioned far more than the rest. “Fragmented” graphs (b) represent discussions with many disjointed participants. “Cohesive” TDGs (c) have a strongly connected main component of participants that persist between multiple sessions.



(a) star (b) fragmented (c) cohesive

Figure 4: Inequality and mobility are inversely correlated. For clarity, we show the relationship only for cohesive groups (r = -0.661, MSE = 0.018).



7. GROUP HEALTH

Finally, we study the relationship between the statistics computed in Section 5 and the health of a TDG. We use two measures of health: the stability or decline of the number of sessions of the group over time, and the power of retention of participants. Since this is a complex multivariate problem, we use machine learning and statistical pattern recognition to quantify these associations.

7.1 Session decline

Problem setup.

First, we attempt to use the statistics to actually *forecast* the decline or stability of the TDGs. More specifically, we use statistics computed from one year of data (June 2012 to June 2013) to predict changes in the following six months. We measure the change in the number of sessions between the 6 months from June 2012 to November 2012 and the 6 months from June 2013 to November 2013. We selected these time periods (the same six months of the year) to avoid any confounding seasonality effect (*e.g.*, chats about football going dormant in the springtime). We define two populations: those TDGs that declined by at least 70% in the number of sessions, and an equal number of the most *stable* TDGs (this resulted in 398 stable groups, namely those with less than 5% change). We build a classifier that takes as input the statistics of interest for a TDG (*e.g.*, mobility, inequality) and as output a decision on whether the group will remain *stable* or *decline*. The accuracy of this classifier is a measure of the precision by which one can establish the future change of a TDG just by looking at these computed statistics, and quantifies the predictive power of the statistics for the decline or stability of the TDG.

After benchmarking several classifiers, including random forests and support vector machines, we settled on logistic regression (augmented with L1 and L2 regularization) as the main classifier. The estimation of accuracy is done using 10 fold cross-validation. We gather results using the mobility and equality statistics, as well as some additional statistics we computed from the model (see Section 5. We also compare some statistics that are unrelated to the model — such as the number of sessions of the group, average number of participants, and maximal clique size in the @mention graph. These features correspond to some previously used to predict online group growth in related literature [17]. A full list of statistics is shown in Table 1. It is important to note here that our intention with these experiments is not to build the optimal classifier for precisely forecasting group decline, but to understand the relationships between these statistics (particularly those with meaningful interpretations in terms of social interactions, *i.e.*, mobility and inequality) and group health.

Results.

The results are presented in Table 2. We present classification results for all declining or stable TDGs (N = 796) and also for the subset of cohesive TDGs (N = 260). Despite the fact that forecasting is a very difficult problem, accuracies are still well above the 0.5 baseline.[7]. The accuracy using just the mobility statistics as input to the classifier is between 0.721 and 0.772. Most notably, the mobility statistics alone capture between 96% to 98% of the maximum accuracy (achieved using all the features), and comparable accuracy to the non-model features that have been successfully used in previous work to predict group growth [17]. As we relax the constraints on the definition of a stable TDG from exhibiting a change of 5% to plus/minus 14%, and sim-

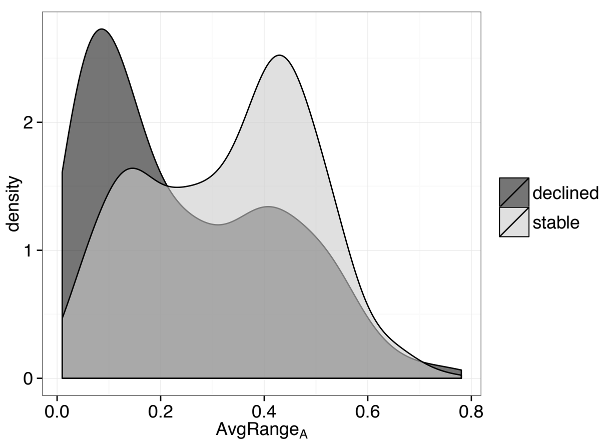
[7] The two populations of declining and stable TDGs have an equal number of groups, namely 398.

ilarly we label a TDG as declining when it presents at least a 50% reduction, the accuracy of the forecasting exercise is reduced from 0.72 to 0.63, as this is a very noisy exercise. Yet, we remark that (a) this is still significant (*i.e.*, above the baseline of 0.50 – again equal populations) and (b) the relationships between the different group of statistics remain the same.

Table 2: Accuracy of predicting groups that declined (> 70% decline in number of meetings) vs. those that stayed stable (< 5% change in number of meetings) in the next six months (based on features from the previous year).

	All	Cohesive
Inequality	0.680 (0.052)	0.760 (0.090)
Mobility	0.721 (0.047)	0.772 (0.086)
All model	0.751 (0.064)	0.779 (0.074)
Non-model	0.693 (0.061)	0.772 (0.090)
All	0.754 (0.062)	0.785 (0.070)

Figure 5: Distributions of the mobility statistic measuring the average range of π_A, grouped by declining and stable TDGs.



To understand the directionality of the relationship between equality or mobility and decline or stability, we build classifiers based on a single statistic. Using the equality statistic $Top10_A$ (see Table 1) alone, the threshold to classify a chat as "declining" is a value greater than or equal to 0.466 (0.59 accuracy[8]). This indicates that greater inequality in the distribution of attention over the participants in a TDG is associated with greater likelihood of the number of sessions declining in the future. Using the mobility statistics $AvgRange_A$ (see Table 1) alone, the threshold to classify a group as "declining" is a value less than or equal to 0.350 (0.62 accuracy). This indicates that lower mobility over the attention receiving archetype (A) between session to session is associated with greater likelihood of the number of sessions declining in the future. Figure 5 shows the distributions of the mobility statistic $AvgRange_A$ for declining and stable chats. We recall that the direction of these relationships match those found in socioeconomic research, where higher mobility and lower inequality were found to be associated with healthier nations.

[8]Performance was similar for the other features when taken individually. Higher performance with combined features indicates that they capture complementary information.

7.2 Participant retention

Problem Setup.

A second measure of a group's health is the retention rate of its participants, *i.e.*, how likely participants are to return to the group. We quantify the retention rate over 12 months of data as follows: Given that a participant attends a session in the first 10 months, what is the probability that he or she returns to a second session in those 12 months? Similar to the forecasting exercise in the previous section, we study the relationship between the group statistics and the *retention rate* using the accuracy of learned classifiers as a measure of association. We define two populations: the 300 groups that had the lowest retention rate and the 300 groups with highest retention rate. The resulting "low retention" population has retention rates below 12%, and the "high retention" population has retention rates above 42%. As before, we use logistic regression and 10 fold cross-validation to estimate performance.

Results.

Results of the retention rate classification task are presented in Table 3. Again, we show results for all groups (N = 600) and the subset of groups that are cohesive (N = 172). These results show that association between the group statistics and retention rate is even stronger than with decline (part of this is due to the fact that we are no longer predicting the future — retention rate is calculated for only the first 12 months of data). Mobility statistics alone result in accuracies between 0.830 and 0.872 — between 93 and 98 percent of the maximum accuracy. While highest classification accuracy is obtained using a combined set of statistics, comparable performance with the equality and mobility statistics alone indicates the presence of a relationship between these characteristics and the retention rate of a TDG.

Table 3: Classification accuracy distinguishing groups with high retention vs. low retention. 300 groups with lowest retention rate (< 0.12) vs. 300 groups with highest retention rate (> 0.42).

	All	Cohesive
Inequality	0.773 (0.054)	0.814 (0.081)
Mobility	0.834 (0.040)	0.872 (0.059)
All model	0.875 (0.039)	0.895 (0.061)
Non-model	0.800 (0.044)	0.855 (0.051)
All	0.887 (0.037)	0.890 (0.044)

8. CONCLUSIONS

In this paper we have proposed a framework for characterizing the interactions of participants in online discussion groups in terms of social equality and mobility. We have shown that in the context of TDGs, (a) mobility and inequality are statistically correlated, with an inverse relationship that mirrors the socioeconomic literature, and (b) higher mobility and lower inequality correlate to and predict healthier groups. In fact, equality and mobility are as predictive of a group's health as other characteristics of online groups that have been previously used to predict the health of online groups and communities. Such observations may be useful to group moderators or other stakeholders, as a more tangible metric on which to monitor the group.

These results hold even when restricting analysis to cohesive groups, which fit the natural idea of discussion-based communities with recurring members. We find these micro-level analogs in online discussion groups non-obvious and thought-provoking.

In any analysis such as this, some initial assumptions must be made. We chose to use a generative model to uncover latent spheres of attention and intuitive metrics of mobility and inequality, although there are many possible ways of defining these concepts. Future work should investigate how robust these results are to the modeling assumptions and parameter choices made. It would also be interesting to see if our findings hold for other types of online communities. In addition, while our results uncovered interesting correlations and predictive relations, further research is needed to establish causal relationship. In such a case, it could also be fascinating to uncover the social mechanisms in which equality and mobility may contribute to the health of an online community.

This work also spawns other interesting avenues for future research. Other measures of group health may be interesting, including: the quality of the exchanged content or diversity in terms of the demographic characteristics of participants. Also, while we shed some light on the demise of a group, there is much to be understood about their birth. If group formation is itself sparked on Twitter then there may be electronic trails that help us understand the different ways that groups form and there is an opportunity to characterize how the interactions between participants determine this birth. Additionally, our work is based on "machine-perceived" measures of equality and mobility. This gives us the advantage of quantifying at scale the connection to group health. Future work might explore how human-perceived notions of equality or mobility (which can perceive more than changes in the @mention graph) differs from these automatic measures. Finally, the level of attention that different participants receive within a typical session exhibit significant levels of inequality. Yet, over many sessions typical discussion groups display high levels of mobility (it is not the case that the same participants receive the attention). This raises the question of whether online communities (or at least those we examined) are closer to meritocracies than "real-word" communities.

9. REFERENCES

[1] E. M. Airoldi, D. M. Blei, S. E. Fienberg, and E. P. Xing. Mixed membership stochastic blockmodels. In *NIPS*, pages 33–40, 2009.

[2] L. Backstrom, D. Huttenlocher, J. Kleinberg, and X. Lan. Group formation in large social networks: Membership, growth, and evolution. In *KDD*, 2006.

[3] L. Backstrom, R. Kumar, C. Marlow, J. Novak, and A. Tomkins. Preferential behavior in online groups. In *WSDM*, 2008.

[4] K. D. Benne and P. Sheats. Functional roles of group members. *Journal of social issues*, 4(2):41–49, 1948.

[5] J. Blanden, P. Gregg, and S. Machin. Intergenerational mobility in europe and north america. *Report supported by the Sutton Trust, Centre for Economic Performance, London School of Economics*, 2005.

[6] S. P. Borgatti and M. G. Everett. Models of core/periphery structures. *Social networks*, 21(4):375–395, 2000.

[7] B. S. Butler. Membership size, communication activity, and sustainability: A resource-based model of online social structures. *Info. Sys. Research*, 12(4), Dec. 2001.

[8] R. Chaiken, B. Jenkins, P.-Å. Larson, B. Ramsey, D. Shakib, S. Weaver, and J. Zhou. Scope: easy and efficient parallel processing of massive data sets. *Proceedings of the VLDB Endowment*, 1(2):1265–1276, 2008.

[9] J. Cook, A. Das, K. Kenthapadi, and N. Mishra. Ranking twitter discussion groups. In *COSN*, 2014.

[10] J. Cook, K. Kenthapadi, and N. Mishra. Group chats on Twitter. In *WWW*, 2013.

[11] P. Csermely, A. London, L.-Y. Wu, and B. Uzzi. Structure and dynamics of core/periphery networks. *Journal of Complex Networks*, 1(2):93–123, 2013.

[12] P. Fajnzlber, D. Lederman, and N. Loayza. Inequality and violent crime. *JL & Econ.*, 45:1, 2002.

[13] P. Holme. Core-periphery organization of complex networks. *Physical Review E*, 72(4):046111, 2005.

[14] C.-C. Hsieh and M. D. Pugh. Poverty, income inequality, and violent crime: a meta-analysis of recent aggregate data studies. *Criminal Justice Review*, 18(2):182–202, 1993.

[15] Q. Jones, G. Ravid, and S. Rafaeli. Information overload and the message dynamics of online interaction spaces: A theoretical model and empirical exploration. *Info. Sys. Research*, 15(2), 2004.

[16] E. Joyce and R. Kraut. Predicting continued participation in newsgroups. *J. Comput. Mediat. Comm*, 11(3), 2006.

[17] S. R. Kairam, D. J. Wang, and J. Leskovec. The life and death of online groups: Predicting group growth and longevity. In *WSDM*, pages 673–682, 2012.

[18] G. Kossinets and D. J. Watts. Empirical analysis of an evolving social network. *Science*, 311(5757), 2006.

[19] M. L. Markus. Toward a "critical mass" theory of interactive media universal access, interdependence and diffusion. *Communication research*, 14(5), 1987.

[20] M. Marmot and R. Wilkinson. *Social determinants of health*. Oxford University Press, 2005.

[21] T. Minka, J. Winn, J. Guiver, S. Webster, Y. Zaykov, B. Yangel, A. Spengler, and J. Bronskill. Infer.NET 2.6, 2014. Microsoft Research Cambridge. http://research.microsoft.com/infernet.

[22] K. Pickett and R. Wilkinson. *The Spirit Level: Why Greater Equality Makes Societies Stronger*. Bloomsbury Publishing, 2010.

[23] M. P. Rombach, M. A. Porter, J. H. Fowler, and P. J. Mucha. Core-periphery structure in networks. *SIAM Journal on Applied mathematics*, 74(1):167–190, 2014.

[24] M. Shafiei and H. Chipman. Mixed-membership stochastic block-models for transactional networks. In *Data Mining (ICDM), 2010 IEEE 10th International Conference on*, pages 1019–1024, Dec 2010.

[25] C. C. Werry. Internet relay chat. *Computer-mediated communication: Linguistic, social and cross-cultural perspectives*, pages 47–63, 1996.

[26] R. Wilkinson. Why is violence more common where inequality is greater? *Annals of the New York Academy of Sciences*, 1036(1):1–12, 2004.

[27] R. G. Wilkinson. Income distribution and life expectancy. *BMJ: British Medical Journal*, 304(6820):165, 1992.

[28] R. G. Wilkinson. *Unhealthy societies: the afflictions of inequality*. Routledge, 2002.

[29] R. G. Wilkinson and K. E. Pickett. Income inequality and population health: a review and explanation of the evidence. *Social science & medicine*, 62(7):1768–1784, 2006.

[30] R. G. Wilkinson and K. E. Pickett. The problems of relative deprivation: why some societies do better than others. *Social science & medicine*, 65(9):1965–1978, 2007.

[31] J. Winn and C. Bishop. Variational message passing. *Journal of Machine Learning Research*, 6:661–694, 2005.

[32] Y. Yu, M. Isard, D. Fetterly, M. Budiu, U. Erlingsson, P. K. Gunda, and J. Currey. DryadLINQ: A system for general-purpose distributed data-parallel computing using a high-level language. In *OSDI*, 2008.