Identifying Virtual Tribes by Their Language in Enterprise E-Mail Archives

Lee Morgan[a], **Peter A. Gloor**[a]

[a] MIT Center for Collective Intelligence, 245 First Street, Cambridge, MA 02142, USA
pgloor@mit.edu

Abstract The rise of online social networks has created novel opportunities to analyze people by their hidden "honest" traits. In this paper we suggest automatic grouping of employees into virtual tribes based on their language and values. Tribes are groups of people homogenous within themselves and heterogenous to other groups. In this project we identify members of digital virtual tribes by the words they use in their everyday language, characterizing e-mail users by applying four macro-categories based on their belief systems (Alternative Realities, Personality, Recreation, and Ideology) developed in earlier research. Each macro-category is divided into four orthogonal categories, for instance "Alternative Realities" includes the categories "Fatherlanders", "Treehuggers", "Nerds", and "Spiritualists". We use the *Tribefinder* tool to analyze two e-mail archives, the individual mailbox of an active academic and corporate consultant, and the Enron E-Mail archive. We found tribes for each user and analyzed the communication habits of each tribe, showing that members of different tribes significantly differ in how they communicate by e-mail. This demonstrates the validity of our approach to distinguish members of different virtual tribes by either language used or e-mail communication structure and dynamics.

1 Introduction

In today's age of alternative realities, different groups in society look at the same underlying evidence as either fact or fiction. In this paper we apply a system we developed earlier to find these groups – virtual tribes – in the corporate world. Our goal is to identify virtual tribes among employees of a company to better understand the different value systems motivating the members of the organization.

Tribes are groups of people that share common ideas, thoughts, and emotions (Cova & Cova, 2002). In other words, they are people who have strong cultural, emotional, and ideological links to each other, creating a sense of community (Cova & Cova, 2002). There are many types of tribes, and they can vary in size and role (Cova, 1996). However, most of the literature surrounding tribes has seen their use in marketing (Gloor & Colladon, 2019). Consumer tribes have emerged as an important part of a firm's success. There is limited use of tribes being used for managing human resources, as we will be proposing in this paper.

Human resource management has been taking strides towards the use of analytics for better understanding the wants and needs of employees. In the past years, human resource research based on mining data has become a notable field, with dozens of studies emerging (Stroheimer & Piazza, 2013). This has permeated into industry as well, with multiple firms employing these techniques (Marr, 2018), for instance using Email for data driven human resources management (Marr, 2018).

Until now, the concept of virtual digital tribes has yet to be used. In earlier work, strategic benefits from using the concept have been shown, like increasing the happiness of workers using virtual mirroring and identifying different emotions through tribes (Gloor & Colladon, 2019). Until now it has been difficult to automatically identify tribes instantly, without using surveys or other manual tools, but based on systematically identifying tribes based on their activity online. As a result, it has been difficult to identify and classify membership in tribes on a large scale. Due to the rise of data driven human resource management, social media, and Email, a different, digitally based, tribe identification method has been developed. Online communities can easily form based on a common idea or interest, and can have the same positive and negative implications as tribes that are not based on the Internet, though over the Internet they might have a greater impact due to the ease of access and spreading of information (Adams. & Smith, 2008). We call these new tribes virtual, electronic, or E-tribes. (Gloor et al., 2019).

Tribefinder is a novel system that uses Artificial Intelligence and Machine Learning to identify the tribes of users based on social media data (Gloor et al., 2019). While originally created to be used with data from Twitter, it can also operate on other forms of media, like Email (Gloor et al., 2019). Tribefinder works through the use of word embeddings and long short term memory (LSTM) (Hochreiter et al. 1997). It currently determines tribes using the words in their messages. More specifically, it finds the different types of tribes and their leaders on Wikipedia, then looks at the language of the leaders on Twitter (Gloor et al., 2019). People are assigned to tribes if their word usage is like that of the aggregate of all "leaders" of a tribe. (Gloor et al., 2019). As a proof of concept for applying Tribefinder for Email, this paper uses Tribefinder to determine the tribes in a personal inbox and the Enron Dataset (Klimt et al., 2004). Tribefinder works with multiple macro-categories of tribes, with users fitting into a specific tribe under each macro-category. This paper will work with the *Alternative Realities, Personality, Recreation,* and *Ideology* macro-categories (Gloor et al. 2019). Each tribe has specific traits, and this paper will look at and compare them between the different tribes. The traits, or honest signals, that this paper uses are related to productivity, connectivity, complexity, and communication habits of each tribe (Gloor 2017).

This paper advances current research as it applies *Tribefinder* to human resources management and Email. It will also show the differences in the traits of each Email tribe, in order to tease out the characteristics of each tribe. This adds onto data driven management, as it offers a novel way to analyze the data of employees and boost their productivity. This can be done through *virtual mirroring*, which mirrors back the communication habits of persons, causing internal reflection (Gloor, 2017). As a result, there can be an increase in customer satisfaction and overall productivity (Gloor, 2017).

2 Theoretical Background

Virtual Tribes

Tribes are affectual groups that are not held together by formal societal constructs, but instead a common emotion, belief, or ideology glues their members to each other (Cova & Cova, 2001). These tribes can be heterogeneous, meaning that members have differences in their ages, incomes, genders, races, and social status; the most important factor about deciding a tribal affiliation is a common belief (Cova & Cova 2002). The postmodern society contains a large amount of these invisible micro-groups, which all share strong emotional links (Cova 1996). Moreover, 'tribe,' as a word, hints at seemingly ancient and archaic values, like "a local sense of identification, religiosity, syncretism, group narcissism etc." (Cova & Cova, 2001).

Humans typically choose which tribes to associate themselves with through their actions and behaviors (Holzweber et al., 2015). This is called the self-categorization theory, it occurs due to one's access and fit to a tribe (Turner et al.,1991). Fit is the extent to which tribes reflect realistic societal groups and statuses. A high fit would indicate minimal intra category differences and minimize inter category similarities (Hornsey, 2008). Access is simply the ease of joining a tribe and its proximity to an individual. If one has more access to a tribe, they are more likely to categorize with it (Hornsey, 2008). It should be noted that self-categorization theory states that the process of finding a tribe can change based on the situation and is always based on the perspective of the perceiver (Hornsey, 2008). Other factors that would influence one's social categorization would be benefits to one's identity, place in society, a stronger sense of community, emotional links, and ethnic partiality (Ellemrs, Kortekaas, & Ouwerkerk 1999; Garry et al. 2008; Maffesoli 1995). Additionally, people can identify with one or more tribes (Mitchell & Imrie, 2011). This is because humans need to express separate parts of their identities, and one tribe alone cannot typically do this; humans need multiple tribes to accommodate for different aspects of their identities (Mitchell & Imrie, 2011). An example of the behaviors of tribes is an "anchoring event," where tribe members meet in public areas and perform ritual acts (Aubert-Gamet and Cova 1999). These anchoring events are essential for tribes to have consistent and sustained membership as they enforce the key ideals and values of tribes (Cova, 1999) However, it should be noted that there is a spectrum when it comes to engagement in "anchoring events" (Cova & Cova, 2002). On one side, there are sympathizers, who have a limited amount of interest in the tribe, and on the other side, there are practitioners, whose identities are based on the tribe and who engage with it daily (Cova & Cova 2002). For these reasons, tribalism is emerging in our society, today's tribes are highlighted by an important duality: the tribe influences its members, but at the same time, the members define their tribe (Bauman, 1990; Maffesoli, 1995). Moreover, traditional tribes have also shifted to virtual tribes or E-tribes (Wright et al., 2006; Hamilton & Hewer, 2010). This is due to the rise of social media and the Internet as a whole: there are many new forums and sites for tribes to be fostered and created (Adams & Smith, 2008). Tribes have been found and researched on sites like Twitter, due to the volume and availability of public messages on the site (Gloor et al., 2019).

Tribes in Human Resources Management

Data driven human resource management can be defined as the use and mining of data coming from employees and customers of a firm, and implementing models and solutions based on the data in the company (Murphy & Zandvakili, 2000). Indeed, many companies are now utilizing data instead of a manager's "gut instinct" when it comes to making decisions (Brynjolfsson et al., 2011). This is accompanied by a data revolution, where companies can now easily collect and aggregate data (Brynjolfsson et al., 2011). This practice has brought in substantial benefits as the decision making that comes from mining data can cause a 5-6% increase in firm productivity (Brynjolfsson et al., 2011).

Despite the rise of data driven HR, tribes and community organization are rarely utilized when it comes to HR management (Marr, 2018; Dealtry & Smith, 2005). They have, however, seen their use in the marketing world (Gloor et al., 2019). Tribes for specific firms, or *brand communities* have been used to easily and quickly spread information about products to consumers (Gloor et al., 2019). Because of this, the formation of consumer tribes has been identified as critical to the survival of firms at any stage of their development (Holzweber et al., 2015).

Though Emails may seem less like a social network and more of a communication tool, research based on mining data from Email databases proves the opposite (Bird et al., 2006). Emails have also been used as a source to mine data (Bird et al., 2006). In professional environments, Emails represent a typical social network and exhibit "long-tailed, small-world" traits (Bird et al., 2006). There are varying levels of participation and leadership within this space as well: a few members send the vast majority of the emails, and there seems to be a hierarchy in the social network (Bird et al. 2006). Thus, Email provides a useful substrate for discovering the tribal affiliations of individuals and groups. For example, a study looked at a firm's emotional tribes and found conclusive results which could be virtually mirrored back to the employees to increase the productivity and happiness of the workers (Gloor & Colladon, 2019; Gloor, 2017).

3 Methods

Challenges in finding tribes

Tribes can be compared to the elementary particles of quantum physics as they are difficult to pinpoint due to their fuzzy and ever changing nature (Cova, 1999). Thus, even though there have been many methods of identifying tribes, like interviews, focus groups, surveys, ethnographic and netnographic approaches, there has been no way to rapidly and automatically identify tribes based on their traits (Gloor, 2019). These manual methods provide a deep understanding of tribal characteristics (Gloor, 2019). In the past decades, tribal studies have used limited surveying methods, like making surfers fill out questionnaires and studying small groups of adult record collectors (Moutinho et al., 2007; Mitchell & Irmie, 2011). For analyzing E-tribes with millions of members, these methods are impractical and cumbersome (Gloor, 2019). Rather, for a large company that wanted to find tribal attributes of its employees, it would be easier to analyze Email databases.

Tribefinder

In order to solve these issues around manual identification of tribes, *Tribefinder* discovers tribes based on text data (Gloor et al., 2019). *Tribefinder*, which until now has been mainly been used on Twitter, categorizes people into one tribe for each specific macro-category, since there are multiple macro-categories, people can belong to multiple tribes (Gloor et al., 2019). *Tribefinder's* analysis of Tweets and Emails extracts data on multiple ideas and leaders. It outputs tribal affiliations for each specific user allowing researchers or managers to find typically unnoticeable traits that distinguish individuals. It should be noted that the macro-categories that *Tribefinder* can output are not rigid. In previous instances, its outputs were macro-categories like *Alternative Realities*, *Ideologies*, and *Personalities*, but the user of *Tribefinder* can create different tribes based on their own specifications (Gloor et al., 2019). For instance, *Tribefinder* has been used to create a "Bernie Sanders Tribe" and a "Donald Trump Tribe" and sort Twitter users into them (Gloor, 2017).

Tribefinder includes two functions: tribe allocation and tribe creation. With the tribe creation function the user can create macro-categories and specific tribes within them. Tribe allocation assigns tribal affiliations to people based on their characteristics. This paper uses the tribe allocation process as the macro-categories are already determined.

In order to create a new tribal macro-category, the user first has to find a group of key individuals that are representative of each tribe within the tribal macro-category, the "tribe leaders" (Gloor et al., 2019). For instance, the "Bernie Sanders Tribe" could have Bernie Sanders and some of his most ardent supporters and campaign managers as its leaders. After this, *Tribefinder* would find a large sample of individuals similar to the "tribe leaders" based on automatically extracted keywords that would be associated with a certain tribe. After the user would identify key leaders, dozens of similarly self-identified members of the tribe will be proposed as additional tribe leaders. For instance, if the user wanted to find individuals that were part of the "Arts" tribe on Twitter, it would search profiles for biographies, Tweets, friends, and followers related to art in order to find these new tribe members. After this, they are shown to the user, who can choose to include these people as tribe leaders or not.

These results can be demonstrated to the user in two types of charts. The first is a word cloud which shows the most common concepts for a certain tribe and can act as a suggestion for new keywords. The second is a drawn-out network of members to demonstrate the most connected and well-known members of tribes.

After this, tribe allocation occurs. This begins with TensorFlow deep learning being used to find the key patterns and ideas in the tweets of tribe leaders (De Oliveira & Gloor, 2018). This is used to identify textual patterns for each tribe and create a specific set of words for each tribe as well. Then, more deep learning is used to analyze the vocabulary and syntax of tribe leaders in order to be able to connect unconnected individuals to a tribe (De Oliveira & Gloor, 2018). Then, using long short term memory and word embeddings, classifications for specific users can be created. Specifically, one's words in Emails or Tweets are converted into vectors which are then inputted into the long short term memory models (Greff et al., 2017).

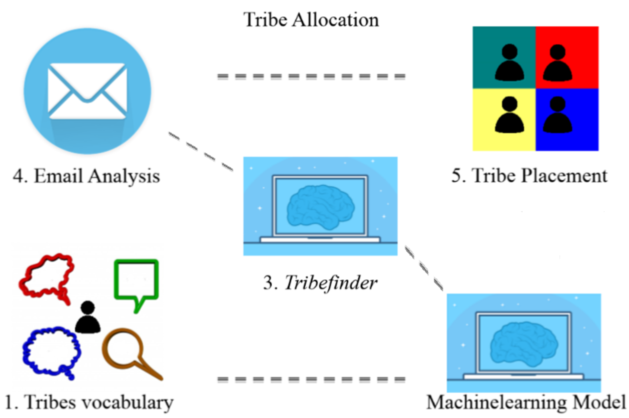

Fig 1. Tribe Allocation Diagram

Although it has been difficult to run models on sites like Twitter due to a lack of long messages (Vo & Ock, 2015), the word embeddings and LSTM used proved to be sufficient. One limitation is that *Tribefinder* needs a large amount of Tweets or Emails in order to work. It should be noted that *Tribefinder* does not depend on deep learning models in order to work, as it can use different methods for short text analysis, but it is most accurate when LSTM and word embeddings are used (Gloor et al., 2019).

Tribe categories

This paper does not focus on developing new tribes, but rather focuses on analyzing the traits of pre-existing macro-categories. The macro-categories focused on in this paper are *Alternative Realities, Personalities, Recreations,* and *Ideologies*. These have been created and utilized in previous research on tribes (Gloor et al., 2019).

The *Alternative Realities* macro-category is broken into four groups. The first are fatherlanders, which can be described as extremely patriotic. Their main vision would be a recreation of the national states from the 1900s. The spiritualism tribe unsurprisingly has a focus on all things spiritual. The nerds are people who believe in seeing advances in technology and strides to the future. The tree huggers are environmentalists and strive to protect nature from phenomena like global warming (Gloor et al., 2019).

The *Personalities* macro-category has four parts as well: stock-traders, politicians, journalists, and risk-takers. Stock-traders have a focus on capital and the economy. “Politicians” are representative of people who use “political language” instead of simply saying the truth. Risk-takers like to make daring decisions (this category has been trained with wingsuit flyers and cave divers), and journalists, other than politicians, use direct language to report actual events.

The *Recreation* macro-category is composed of the fashion, art, travel, and sport tribes. Fashion tribe members focus on the new styles of clothes; the arts tribe has an interest of all types of art like music and painting; the travel tribe enjoys travelling around the world; and the sport tribe enjoys actively engaging in sports (Gloor et al., 2019).

Finally, the *Ideologies* macro-category is made of the liberalism, socialism, capitalism, and complainers tribes. The liberalism tribe focuses on enhancing and protecting the freedom of individuals. The socialism tribe advocates for more government control and intervention in economies. The capitalism tribe is practically the opposite of socialism—it argues for minimal government intervention in markets. The complainers tribe frequently voices their protests to problems they see.

Utilizing Tribefinder: Honest Signals

Tribefinder proved to be powerful as it can be used to discover non-obvious characteristics of employees in a firm. In previous work, it was tested with Twitter with an accuracy rate of 81.2% in the best case and 68.8% in the worst case (Gloor et al., 2019). It has been used to identify customer's tribal affiliations to see which tribes are more likely to have interest in certain brands, and it has seen use in identifying the traits of customer tribes (Gloor et al., 2019). In this paper, employee tribes will be analyzed through the use of the honest signals (Gloor, 2017; Pentland, 2010). These honest signals identify differences in the activity and language of tribes, which firms can use to see which tribes are the most positive and active in the workplace.

Honest signals are part of network science, which is a way of seeing individuals as part of a group instead of isolated people (Pentland, 2010). They can be seen as seemingly unnoticeable patterns which reveal the goals and key ideals of people (Pentland, 2010). They are honest because they are uncontrollable due to them being processed unconsciously. These honest signals can be extremely effective as they can predict outcomes in seemingly random situations like dates and job interviews (Pentland, 2010).

The interconnectedness of a tribe is measured through its social network's network centrality (Freeman, 1978). There are two measures of centrality used in this paper: betweenness and degree centrality. Degree centrality is simply the amount of people a user sends and receives Emails from. Betweenness centrality is the frequency in which a user appears in a path connecting other users. This is computed through finding the shortest paths in a network that connects all network's users to each other, then counting the amount of times one appears in a path connecting two other users (Wasserman & Faust, 1994).

The activity of users was measured through messages sent, contribution index, and rotating leadership. Messages sent is simply the amount of Emails sent by an individual. Rotating leadership is the oscillations in betweenness centrality in a specified time period (15 days). This is calculated by finding the "number of local maxima and minima in the betweenness curve of an actor" (Gloor, 2017). A rotating leader is someone who alternates between being a leader and follower in groups. For example, they would start discourses then allow the other members of the network to carry them on (Kidane & Gloor, 2007). The contribution index measures the balance of messages sent and received for a user. It is calculated by subtracting messages received from messages sent then dividing the result by messages sent added to messages received (Gloor, 2017).

The last characteristic analyzed was the language of the tribe members. This was done by finding the average sentiments, emotionalities, and complexities. Average sentiment is the measure of the positivity and negativity of a user's Emails. It was calculated using a classifier algorithm and varies from 0 and 1, with 0 being the most negative and 1 the most positive (Gloor et al., 2019). Average emotionality is the measure of user's deviation from the usual sentiment and is measured as the standard deviation from the mean sentiment (Gloor, 2017). Finally, average complexity measures the complexity of a user's vocabulary. The more varied words one uses, the higher their complexity (Gloor, 2017). All of the honest signals were calculated with Condor (Gloor, 2017).

4 Results

Though *Tribefinder* can be used to create new tribes, this paper works under the framework using the predefined tribes provided by Gloor et al. (2019). These tribes have their notable traits identified through data mining Emails and social network analysis. This can be impactful as firms can identify employee tribes that need an increase in sentiment and those that speak with the highest complexity, meaning new ideas are coined. This analysis uses the Enron Large Dataset (Gloor, 2017), which had 1738 users that were placed into tribes. The results from Enron are compared to those of a private Email inbox, which had its 20 most active participants placed in tribes. The charts and tables below demonstrate the significant differences within tribes and compare the results for the Enron and private Emails. The bar charts have error bars of the 95% confidence intervals.

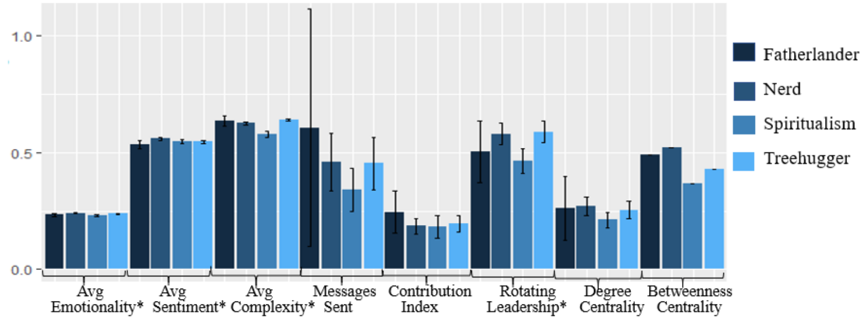

Fig. 2. Text and network metrics for the *Alternative Realities* macro-category (sig. differences marked by asterisk)

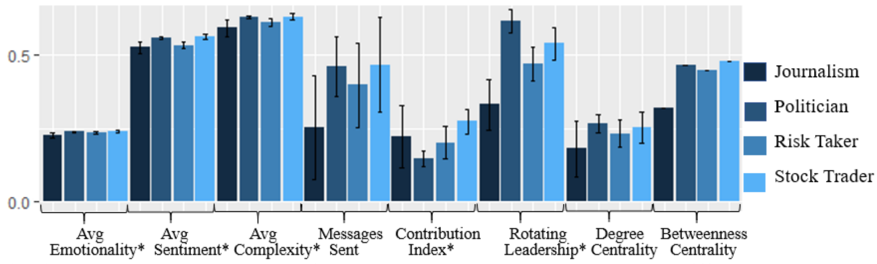

Fig. 3. Text and network metrics for the *Personality* macro-category (sig. differences marked by asterisk)

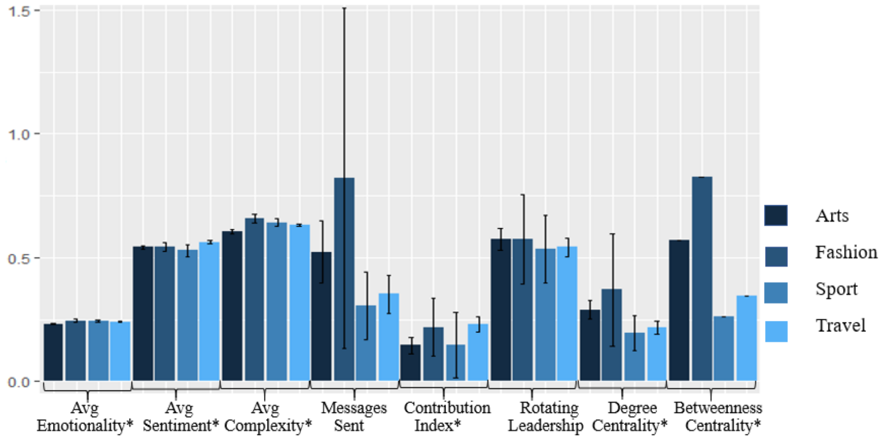

Fig. 4. Text and network metrics for the *Recreation* macro-category (sig. differences marked by asterisk)

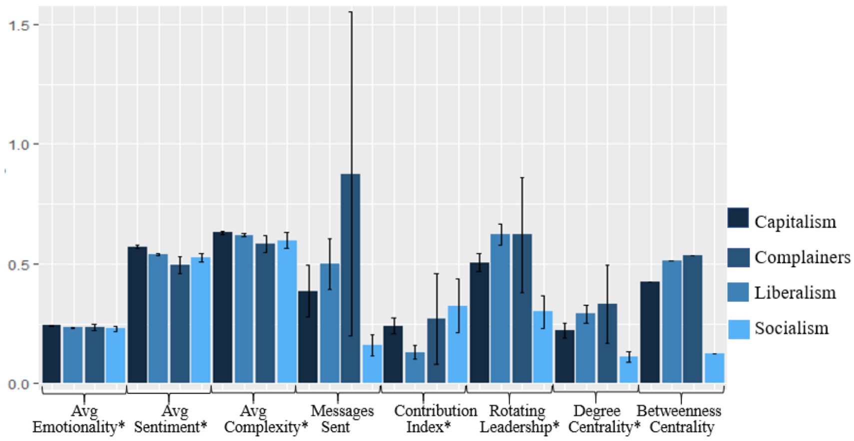

Fig. 5. Text and network metrics for the *Ideology* macro-category (sig. differences marked by asterisk)

***In these charts, Betweenness Centrality was divided by 500,000, Messages Sent by 200, Average Complexity by 10, Rotating Leadership by 100, and Degree Centrality by 100 in order to compare the data in one graph.**

Significant Comparisons for Email Data, N=20

Honest Signal	Group	Group	Mean Difference	P-Value
Average Complexity	Spiritualism	Nerd	-1.814	0.0331496
		Treehugger	-2.623	0.0018502

Significant Comparisons for Enron Data, N=1738

Honest Signal	Group	Group	Mean Difference	P-Value
Rotating Leadership	Spiritualism	Nerd	-11.67	0.0103851
		Treehugger	-12.47	0.015948
Average Sentiment	Nerd	Fatherlander	0.0270	0.0246773
		Treehugger	0.0129	0.0402434
Average Complexity	Spiritualism	Treehugger	-0.615	0
		Nerd	-0.474	0
		Fatherlander	-0.582	0.0000001
Average Emotionality	Spiritualism	Nerd	-0.010	0.0000259
		Treehugger	-0.0077	0.003112

Table 1: Differences in Honest Signals among Alternative Reality Tribes for Email Inbox and Enron Inbox

For the *Alternative Realities* macro-category, the one-way ANOVA demonstrates multiple significant differences in honest signals in both Email datasets. The Spiritualism tribe has the lowest average complexity in both datasets, suggesting that they bring less new ideas that Nerds, Fatherlanders, and Treehuggers. These differences are significant (p ranges from 0 to .033). There are no other significant differences in the personal Email inbox, which may be due to the limited sample size. In the Enron dataset, the Spiritualists rotate their positions of leadership the least, which means that their betweenness centrality rarely oscillates. This suggests that in comparison to the Nerds and Treehuggers, Spiritualists rarely change their positions in a group and either stay as group leaders or followers. Nerds speak with the highest positivity out of all the other groups as they have the highest average sentiment. Spirituals also have the least variation in the positivity of their messages, as demonstrated by their low Average Emotionality. All of these differences are significant as well (p ranges from 0 to .04).

Significant Comparisons for Email Inbox, N=20

Honest Signal	Group	Group	Mean Difference	P-Value
Average Complexity	Sport	Arts	1.49	0.0002092
	Fashion	Arts	-3.37	0.0000166
		Sport	-4.86	0.0000002
		Travel	-3.54	0.0000081
Average Emotionality	Fashion	Arts	-0.154	0.0001033
		Sport	-0.142	0.0003459
		Travel	-0.146	0.0001779

Significant Comparisons for Enron Data, N=1738

Honest Signal	Group	Group	Mean Difference	P-Value
Degree Centrality	Travel	Fashion	-15.34	0.0497952
		Arts	-7.334	0.0160727
Betweenness Centrality	Travel	Fashion	-239167	0.0572116
		Arts	-111761	0.023589
Contribution Index	Travel	Arts	0.0842	0.0011708
Average Sentiment	Travel	Arts	0.0215	0.000007
		Sports	0.0333	0.0180155
Average Complexity	Arts	Fashion	-0.525	0.0000107
		Sport	-0.373	0.0080805
		Travel	-0.279	0
Average Emotionality	Arts	Fashion	-0.0134	0.0040415
		Travel	-0.0086	0.0000008

Table 2: Differences in honest signals for Recreation Tribes in Email Inbox and Enron Data

In Table 2, for the *Recreation* macro-category, there is only one similarity between the Enron and Email datasets. The Arts tribe has a lower average complexity than the sports tribe, meaning they bring less ideas to the table (p = .0002, .008). There were many differences between the data as well. Primarily, in the average complexity metric, the Fashion tribe had the lowest complexity in the personal Email inbox, and it had the highest average complexity in the Enron data. The same was true for the average emotionality metric. In the private Emails, there was relatively low variation in the positivity of Emails coming from the Fashion tribe, but there was a relatively high variation in the Enron data. There were only significant differences in average complexity and emotionality for the Email data. In the Enron data, members of the Travel tribe seems to be the most central, as they have a higher degree and betweenness centrality than members of the Fashion and Arts tribes. Moreover, the Travel tribe contributes relatively more than the Arts tribe, with a higher Contribution Index. The Travel tribe also speaks in the most positive manner with the highest average sentiment out of all the recreational tribes. These results are also statistically significant (p ranges from .000002 to .0498).

Significant Comparisons for Email Inbox, N=20				
Honest Signal	Group	Group	Mean Difference	P-Value
Average Complexity	Liberalism	Capitalism	1.237	0.0302374
Significant Comparisons for Enron Data, N=1738				
Honest Signal	Group	Group	Mean Difference	P-Value
Degree Centrality	Liberalism	Capitalism	6.83	0.0272564
		Socialism	17.93	0.0106698
Contribution Index	Liberalism	Capitalism	-0.112	0.0000035
		Socialism	-0.195	0.0012411
Rotating Leadership	Liberalism	Capitalism	11.748	0.0003462
	Socialism	Capitalism	-20.573	0.0161735
		Complainers	-32.243	0.0300972
		Liberalism	-32.322	0.0000184
Average Sentiment	Capitalism	Complainers	0.07544	0.0000014
		Liberalism	0.0327	0
		Socialism	0.0458	0.0000446
Average Complexity	Capitalism	Complainers	0.4923	0.0070199
		Liberalism	0.1172	0.0463337
		Socialism	0.3295	0.0114416
Average Emotionality	Capitalism	Liberalism	0.0074	0.0000306
		Socialism	0.0127	0.0053641

Table 3. Differences in Honest Signals among Ideology Tribes in Emails and Enron Data

In Table 3, the only significant comparison from the Email inbox was that between the complexity of Liberalism and Capitalism tribes, where the Liberalism tribe displayed a wider vocabulary (p = .03). Surprisingly, the Enron data displayed a different trend, as the Capitalism tribe had a higher average complexity than Liberalism did (p = .046). Liberals seem to have the highest connectivity of all the Ideology tribes, as they have the highest degree centrality. However, they seem to communicate less relative to the content they receive, with lower contribution indices than the rest of the tribes. Moreover, the Socialism tribe seems to have the most changes in leadership positions, followed by the Capitalism, Liberalism, and Complainers tribes. The Capitalism tribe speaks most positively in its messages, with an average sentiment higher than the rest of the tribes. Moreover, it has the most oscillations in its sentiment, with the highest average emotionality. These results are all significant, with a maximum p value of .046 overall.

No significant differences in Email Inbox, N=20

Significant Differences in Enron Data, N=1738

Honest Signal	Group	Group	Mean Difference	P-Value
Contribution Index	Stock-Trader	Politician	0.127	0.000008
Rotating Leadership	Politician	Risk-Taker	14.36	0.0015132
		Journalist	28.36	0.0001742
Average Sentiment	Journalist	Politician	-0.03	0.0050642
		Stock-Trader	-0.036	0.0037871
	Risk-taker	Politician	-0.0248	0.0001364
		Stock-Trader	-0.0272	0.000232
Average Complexity	Journalist	Politician	-0.361	0.0030641
		Stock-Trader	-0.392	0.0019163
	Risk-Taker	Politician	0.172	0.0241113
Average Emotionality	Journalist	Politician	-0.0116	0.0111244
		Stock-Trader	-0.025	0.0008865

Table 4. Differences in Honest Signals among Personality Tribes in Enron Data

In Table 4, there are no significant differences in the private Email inbox. However, there are some differences in the Enron data. The Stock-Trader tribe seems to be more productive to conversations than the Politician tribe, as it has a larger Contribution Index. However, the Politician tribe changes its position more in discussions than the Risk-Taker and Journalist tribes, with a high Rotating Leadership. The Journalists and Risk-Takers speak most positively in discussions, as they have the highest average sentiments. Moreover, the Journalists have the largest deviations in the positivity of their messages in comparison to the Politicians and Stock-traders, given that they have a high average emotionality.

5 Discussion and Implications

This paper's findings add to the theoretical and practical study of tribes. We illustrate the validity of this concept also for the analysis of organizations, extending its main use in marketing. From an academic standpoint, this paper expands the use of *Tribefinder* to the Email setting. Earlier work has been mainly focused on social network sites like Twitter (Gloor et al., 2019). Since Email databases also behave like a social network (Bird et al., 2006), the same methodology could be applied there. Moreover, this paper utilizes a new tool developed by Gloor et al. (2019) in order to identify tribes. This allows us to circumvent the traditional methods of identifying tribes, like focus groups and interviews (Mitchell & Irmie, 2011). These groups have their traits analyzed through honest signals (Gloor, 2017; Pentland, 2010), which demonstrates that there are differences among the tribes that have impacts on communication habits. Finally, this paper furthers work done in the field of data driven human resources management and decision making, which has been taking recent strides (Stroheimer & Piazza, 2013), by dividing Email users into groups based on their social network behavior and word content.

This paper is of importance in a managerial sense as well. For many companies, tribes have emerged as a critical factor of their success, especially in marketing (Gloor & Colladon, 2019). In this paper we illustrate the usefulness of this concent in HR management as well. Many human resource managers have begun to analyze the traits of their employees through Emails (Marr, 2018), but division of their users based on their traits has seen limited use. The use of digital social networks in HR is important due to the ease of access and spreading of information in the modern-day Internet (Adams & Smith, 2008).

6 Limitations and Future Work

This work clearly has some limitations. Primarily, workers do not only communicate through Email and use messaging services and social networks. It could be beneficial to also analyze these sources to identify if these results are generalizable. There are also other models that could be used to identify the tribes of certain users, and they could yield different, and potentially more accurate results. Finally, other honest signals could be used besides those in this paper. Average response time and nudges (the amount of Emails one sends in order to get a reply from another) could be used.

7 Conclusion

This paper illustrates the usefulness of the tribe concept for HR analysis. It shows the use of *Tribefinder* in a different medium and framework. It analyzes the communication habits of people in organizations through the lens of Emails, utilizing LTSMs and word embeddings, and places them into tribes that the user can flexibly create depending on the focus of analysis. Four macro-categories of tribes are employed: Alternative Realities, Ideologies, Recreation, and Personality. However, this system could easily be extended for instance to measure moral values of employees, or their attitudes towards risk by creating the appropriate tribes. By comparing the tribal affiliations with the "honest signals of communication", we illustrate the underlying traits of different groups of employees, thus providing valuable cues to managers about the characteristics of their employees. This paper is early research, but it clearly demonstrates the power of this approach to discover the underlying individual attributes and behavioral characteristics of members of an organization otherwise not accessible.

References

1. Adams, T. L., & Smith, S. A. (Eds.). (2009). *Electronic tribes: The virtual worlds of geeks, gamers, shamans, and scammers*. University of Texas Press.
2. Bauman, Zygmunt (1990). Thinking sociologically. B. Blackwell, Oxford ; Cambridge, Mass
3. Bird, C., Gourley, A., Devanbu, P., Gertz, M., & Swaminathan, A. (2006, May). Mining email social networks. In *Proceedings of the 2006 international workshop on Mining software repositories* (pp. 137-143). ACM.
4. Brynjolfsson, E., Hitt, L. M., & Kim, H. H. (2011). Strength in numbers: How does data-driven decisionmaking affect firm performance?. *Available at SSRN 1819486*.
5. Cova, B., & Cova, V. (2001). Tribal aspects of postmodern consumption research: the case of French in-line roller skaters. *Journal of Consumer Behaviour: An International Research Review*, *1*(1), 67-76.
6. Cova, B., & Cova, V. (2002). Tribal marketing: The tribalisation of society and its impact on the conduct of marketing. *European journal of marketing*, *36*(5/6), 595-620.
7. Cova, B. (1999). From marketing to societing: When the link is more important than the thing. *Rethinking marketing: Towards critical marketing accountings*, 64-83.
8. Cova, B. (1996). The postmodern explained to managers: Implications for marketing. *Business Horizons*, *39*(6), 15-24.
9. De Oliveira, Joao Marcos, and Peter A. Gloor. "GalaxyScope: Finding the "Truth of Tribes" on Social Media." *Collaborative Innovation Networks*. Springer, Cham, 2018. 153-164.
10. Dealtry, R., & Smith, E. A. (2005). Communities of competence: new resources in the workplace. *Journal of Workplace Learning*.
11. Ellemers, N., Kortekaas, P., & Ouwerkerk, J. W. (1999). Self-categorisation, commitment to the group and group self-esteem as related but distinct aspects of social identity. *European journal of social psychology*, *29*(2-3), 371-389.

12. Garry, T., Broderick, A. J., & Lahiffe, K. (2008). Tribal motivation in sponsorship and its influence on sponsor relationship development and corporate identity. *Journal of Marketing Management*, *24*(9-10), 959-977.
13. Gloor, P. A., & Colladon, A. F. Heart Beats Brain-Measuring Moral Beliefs through E-Mail Analysis.
14. Gloor, P. A. (2017). *Sociometrics and human relationships: Analyzing social networks to manage brands, predict trends, and improve organizational performance*. Emerald Publishing Limited.
15. Gloor, P., Colladon, A. F., de Oliveira, J. M., & Rovelli, P. (2019). Put your money where your mouth is: Using deep learning to identify consumer tribes from word usage. *International Journal of Information Management*.
16. Greff, K., Srivastava, R. K., Koutník, J., Steunebrink, B. R., & Schmidhuber, J. (2016). LSTM: A search space odyssey. *IEEE transactions on neural networks and learning systems*, *28*(10), 2222-2232.
17. Hamilton, K., & Hewer, P. (2010). Tribal mattering spaces: Social-networking sites, celebrity affiliations, and tribal innovations. *Journal of Marketing Management*, *26*(3-4), 271-289.
18. Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural computation*, *9*(8), 1735-1780.
19. Holzweber, M., Mattsson, J., & Standing, C. (2015). Entrepreneurial business development through building tribes. *Journal of Strategic Marketing*, *23*(7), 563-578.
20. Hornsey, M. J. (2008). Social identity theory and self-categorization theory: A historical review. *Social and personality psychology compass*, *2*(1), 204-222.
21. Kidane, Y. H., & Gloor, P. A. (2007). Correlating temporal communication patterns of the Eclipse open source community with performance and creativity. *Computational and mathematical organization theory*, *13*(1), 17-27.
22. Klimt, B., & Yang, Y. (2004, September). The enron corpus: A new dataset for email classification research. In *European Conference on Machine Learning* (pp. 217-226). Springer, Berlin, Heidelberg.
23. Maffesoli, M. (1995). *The time of the tribes: The decline of individualism in mass society* (Vol. 41). Sage.
24. Marr, B. (2018). *Data-driven HR: how to use analytics and metrics to drive performance*. Kogan Page Publishers.
25. Mitchell, C., & Imrie, B. C. (2011). Consumer tribes: membership, consumption and building loyalty. *Asia Pacific Journal of Marketing and Logistics*, *23*(1), 39-56.
26. Moutinho, L., Dionísio, P., & Leal, C. (2007). Surf tribal behaviour: a sports marketing application. *Marketing Intelligence & Planning*, *25*(7), 668-690.
27. Murphy, T. E., & Zandvakili, S. (2000). Data-and metrics-driven approach to human resource practices: Using customers, employees, and financial metrics. *Human Resource Management: Published in Cooperation with the School of Business Administration, The University of Michigan and in alliance with the Society of Human Resources Management*, *39*(1), 93-105.
28. Pentland, A. (2010). *Honest signals: how they shape our world*. MIT press.
29. Strohmeier, S., & Piazza, F. (2013). Domain driven data mining in human resource management: A review of current research. *Expert Systems with Applications*, *40*(7), 2410-2420.
30. Turner, J., Hogg, M. A., Oakes, P. J., Reicher, S. D., & Wetherell, M. S. (1987). Rediscovering the social group: A social categorization theory. *Oxford, UK: B. Blackwell. van Knip-*

penberg, D., & Schippers, MC (2007). Work group diversity. Annual Review of Psychology, *58*, 515-541.

31. Vo, D. T., & Ock, C. Y. (2015). Learning to classify short text from scientific documents using topic models with various types of knowledge. *Expert Systems with Applications*, *42*(3), 1684-1698.
32. Wasserman, S., & Faust, K. (1994). Wasserman, Stanley, and Katherine Faust, Social Network Analysis: Methods and Applications. New York: Cambridge University Press, 1994.
33. Wright, L. T., Cova, B., & Pace, S. (2006). Brand community of convenience products: new forms of customer empowerment–the case "my Nutella The Community". *European journal of marketing*