Characterizing Organ Donation Awareness from Social Media

Diogo F. Pacheco*, Diego Pinheiro*, Martin Cadeiras† and Ronaldo Menezes*

*BioComplex Laboratory, School of Computing, Florida Institute of Technology, Melbourne, FL, USA
dpacheco@biocomplexlab.org, dsilva@biocomplexlab.org, rmenezes@cs.fit.edu

†Department of Medicine, University of California, Los Angeles, CA, USA
mcadeiras@mednet.ucla.edu

***Abstract*—Approximately 22 people die every day in the USA due to a lack of organs for transplant. Research suggests that the most effective solution is to increase organ donor rates; current, proposals range from expanding the donor eligibility criteria (donor pool) to performing mass media campaigns. However, little is known about the extent in which activities on social media are associated with aspects (e.g. awareness) of organ donation. Our hypothesis is that social media can be utilized as a sensor to characterize organ donation awareness and population engagement in donation for each different organ. In this sense, we collected Twitter messages (tweets) regarding organ donation, and characterized organ awareness by aggregating tweets from users who mostly mentioned that organ. Similarly, we assessed the relative risk between the cumulative incidence of organ-related conversations inside and outside geographical regions to characterize them regarding organ donation awareness. Our characterization suggests that organ-related conversations on social media seems to be indeed associated with aspects of organ donation such as the co-occurrence of organ transplantations. Also, we found variations regarding the specific organs that are prominently discussed in each geographical region, and that such variations seem to be associated with aspects of organ donation in that region; for instance, the abnormal amount of conversations about kidneys in Kansas. Our findings suggest that the proposed approach has the potential to characterize the awareness of organ donation in real-time.**

I. Introduction

Organ transplantation saves thousands of lives every year in the USA [1] and around the world. Despite starting as an experimental medical procedure when the first organ was transplanted in 1954 [2], organ transplantation has become a reliable, effective, and the preferred alternative for end-stage organ failure. Alas, organ transplantation only reaches a small fraction of transplant candidates, and nearly 22 patients die in the USA every day for not having access to a transplant organ. As an example, in 2012, although roughly 60 thousand patients were in the waiting list for a kidney transplant in the USA, only 17 thousand kidney transplants were performed; less than $1/3$ of what was needed.

One of the approaches that have been proposed to improve the shortage of organs focuses on the expansion of the criteria for becoming an eligible donor [3] by increasing the use of higher risk grafts; however, this approach presents side effects such as delayed graft function of the transplanted organ with potential compromise of the graft function both short- and long-term [4]. To better inform policy makers, many past research efforts have focused on the assessment of the organ allocation process. Some works investigated the major factors associated with the survivability of transplanted organs such as the ischemic time [5], and found that the allocation needs to be tailored for each organ. Other works attempted to understand the complex network structure of organ transplantation using a geographic social networks [6], and found some geographical disproportion between donors and recipients as well as some anomalies regarding different organs in the allocation process. Ultimately, the change in allocation policies was debated aiming at reducing regional organs accessibility disparities [7].

Although the aforementioned research efforts attained significant results, they also point to a research agenda focusing on raising the number of donors [8]. In this sense, conversation is a particularly important issue to organ donation awareness [9]. Establishing an effective conversation with families of donor-eligible patients may improve families' consent rates [10] specially when families are aware of organ donation. Commonly, families are approached near the death of their loved ones and approximately half of the families tend to refuse the request for donation [11]. Besides, families are also more likely to authorize donation if they had previously discussed organ donation with the deceased, and they knew the deceased's wishes regarding organ donation [11].

This paper aims at exploring the extent in which social media, such as Twitter, may be used to sense the population regarding organ donation awareness. Then, we adapted the method we previously proposed [12] to characterize entities in social media to organ-related conversations on Twitter. Our characterization demonstrates that social media has sufficient information regarding organ donation awareness and has the potential to be employed as a social sensor for organ donation campaigns.

II. Related Work

The understanding of population awareness regarding organ donation is important to raise rate of donors. Social media has evolved as a new tool to deal with the organ donation issue because it is cost-effective, and it has the potential to attain a higher population outreach [8]. In fact, there has been evidence that people look for social media as a way to create support groups, and that their conversations may lead to a structured social network [13].

Facebook and Twitter are by far the most common social media sites. They have been used in diverse contexts of physical and social phenomena ranging from disaster management [14] to interventions promoting health behavior change [15]. For instance, Facebook has been used in interventions involving sexual health [16], physical activity [17], [18] and food safety [19]. Similarly, Twitter has also been used to carry out interventions related to weight-loss [20] and smoking cessation [21].

Although Facebook and Twitter provide a rich source of information, most of it comes as unstructured text and needs to be understood and characterized in order for us to get useful information. Some previous work proposed a model to extract the dialogue structure from these conversations on Twitter [22], and also characterized conversations of organ donation on Facebook aiming at understanding how organ donation advocacy agencies can influence social media users to share messages to their personal network of contacts [23].

In the context of organ donation, social media applications range from identifying potential kidney donors [24], helping organ donation advocacy agencies to increase online social network engagement [23], and increasing donor registration rates [8]. Yet, despite considerable research on organ donation using social media [9], [23]–[25], little has been done to associate activities in social media with real-world aspects (e.g. statistics) regarding organ donation.

In this work, we tailor a method we previously proposed [12] to characterize conversations on Twitter according to the amount of mention to different solid-organs. We found that users and states can be clustered based on the way they discusses issues related to organ donation.

III. Methods

A. Dataset

We collected data from Twitter[1] because it is one of the most popular Social Media in the United States and it is commonly used by researchers in social experiments. More importantly, Twitter allows data collection for virtually any of its users given that users tend to leave their profile public, i.e., everyone can read their posts. Our processing pipeline has three steps. First, tweets are collected using a *filter* based on our predefined organ donation predicates (keywords). Then, the collected tweets are *augmented* to include their location; this can be done using the tweet geo-tag or the user location found in his profile. Finally, the augmented tweets are *filtered again* to retain only those belonging to USA users.

In order to focus on conversations regarding organ donation, we constrained our search with a set of keywords Q that are used to filter tweets using the *Twitter Stream API*. Figure 1 shows Q as the Cartesian product of a set of *Context* words (limited to organ donation terms) and a set of *Subject* words (limited to organs of interest). This approach guarantees every collected tweet in our dataset contains at least one word from *Context* and at least one of the words from *Subject*. Therefore, our dataset is conceived in the context of organ donation.

[1]We collected our tweets using the public *Twitter Stream API* available at https://dev.twitter.com/streaming/overview.

$$Q = \left\{ \overbrace{\begin{array}{c} \text{transplant} \\ \text{transplantation} \\ \text{donor} \\ \text{donation} \\ \text{donate} \end{array}}^{Context} \right\} \times \left\{ \overbrace{\begin{array}{c} \text{heart} \\ \text{kidney} \\ \text{liver} \\ \text{lung} \\ \text{pancreas} \\ \text{intestine} \end{array}}^{Subject} \right\}$$

Fig. 1. The set of keywords used to collect tweets related to organ donation awareness is the Cartesian product of *Context* and *Subject* words.

TABLE I. Statistics of the dataset used in this paper. The dataset contains collected tweets from users regarding organ donation.

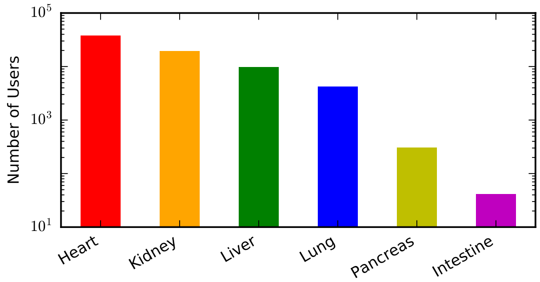
Statistic	Value
Start Data Collection	Apr 22^{th} 2015
Finish Data Collection	May 11^{th} 2016
Number of Days	385
Tweets collected	134,986
Number of Users	71,947
Avg. Tweets / Day	350
Avg. Tweets / User	1.88
Organs mentioned / Tweet	1.03
Organs mentioned / User	1.13

*134,986 out of 975,021 tweets could be identified as from USA users.

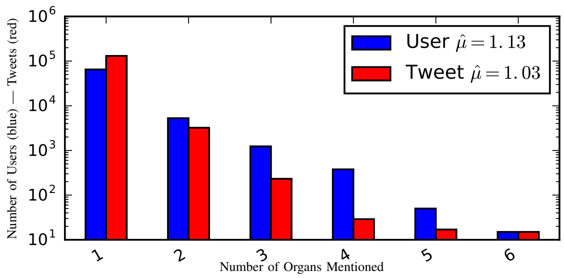
This work focuses on characterizations within the USA, so we only kept the tweets from users located in the USA. The most common options to identify the geographical location of a Twitter user is to use the GPS coordinates (included in some tweets), or to use the self-reported location field in the user profile. The GPS coordinates are more precise and dynamic, but much rarer (about 1.4% [26]). The user profile info is more static and abundant, but requires extra computation and is less precise. In this work, we locate users (country/state) augmenting their self-reported location using *OpenStreetMap*[2]. This method has been shown to be reliable even at the county level [27].

Finally, our dataset comprises of one year of data and represents more than 70 thousands users in the USA; Table I summarizes the statistics. Figure 2(a) shows the number of users mentioning each organ. For instance, heart and intestine are, respectively, the most and least mentioned organs. The attention given to organs on Twitter highly correlates with the number of transplants in the USA (Spearman correlation, $r = .84$, $p < .05$); except for heart, first in popularity on Twitter, but third on number of transplants. Figure 2(b) shows the comparison between the number of tweets and the number of users mentioning multiple organs. The number of tweets is greater than the number of users only for single mentions.

[2]www.openstreetmap.org



(a) Number of Twitter users per organ



(b) Number of organs mentioned per user (blue) and per tweet (red)

Fig. 2. Dataset information shown as histograms in log scale. (a) The number of users per organ is a proxy for the "popularity" of each organ in Twitter conversations. (b) The number of users and tweets mentioning multiple organs. Organs are more likely to be mentioned when tweets are aggregated by user than on solo tweets.

B. Characterization

To build our social sensor for organ donation, we need to capture singularities not only among different organs, but also among distinct places. The first step is to define a structured approach to characterize the continuous and seemingly random stream of information from Twitter.

A straightforward approach is to build a characterization model based on single messages. Despite its intuitiveness, such characterization may be biased by the existence of a few heavily-active users. Twitter has a very heterogeneous tweeting rates per user, ranging from hundreds tweets per day to a handful in months. In addition, users are more likely to mention multiple organs (with several tweets) than a single tweet contain multiple mentions (see Figure 2(b). Thus, to better represent the population, a characterization based on users is more appropriated. Ultimately, a user will be a representation of a collection of tweets.

This work adapts the characterization proposed by Pacheco et al. [12] in which users are defined based on the amount of *attention* they give to a set of entities—here, the set of the most common solid organs transplanted. More specifically, we measure user's attention from his/her tweets as frequencies of mention to organs in the donation context. Formally, we represent m users and their respective attention to n organs using a normalized contingency matrix $\hat{U} = [\hat{u}_{ij}]_{m \times n}$. In this matrix representation, each row fully represents a user, i.e., $\sum_{j=1}^{n} \hat{u}_{ij} = 1$.

Individual characterizations may be too specific to perform an exploratory data analysis. Therefore, to help us to learn and to gain intuitions about the data, we aggregate users to explore two perspectives:

Organs: This hypothesis is that the characterization of an organ in relation to a set of organs can detect dependencies among them, such as the co-occurrence of transplantation (two organs are transplanted at the same time) and the cascade effects in organs failure. We characterize users by the extent of their attention to different organs. We assume an organ can be represented by averaging the behavior of the subset of users who dedicate most attention to it.

Regions: This hypothesis is that the characterization of regions in relation to a set of organs may reveal differences among regions regarding health issues, local policies, or levels of engagement in the donation cause. A region is represented by the aggregation of their inhabitants.

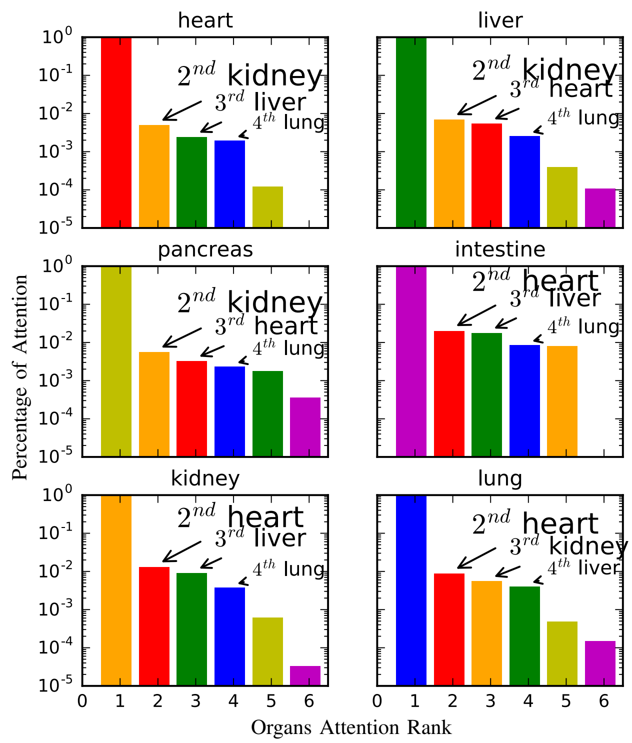


Fig. 3. Characterization of the six major solid organs based on Twitter conversations. Each plot represents an organ, i.e. a row on K (see equations 3 and 1); it conveys how the user whose primary focus is on a specific organ also mentions other organs. The information for heart, kidney, liver, lung, pancreas, and intestine is depicted in red, yellow, green, blue, olive, and magenta, respectively. Note that the histogram bars are in log scale and their values are ranked based on mentions.

To implement each characterization, we define a membership-indicator matrix L. In the characterization of organs, m users are aggregated based on their most cited organ and the membership-indicator matrix $L = [l_{ij}]_{m \times n}$ can be defined as

$$l_{ij} = \begin{cases} 1, & \text{if } j = \arg_j \max \hat{U}(i,j), \\ 0, & \text{otherwise.} \end{cases} \quad (1)$$

Similarly, in the characterization of regions, m users are aggregated based on their locations and the membership-indicator matrix $L = [l_{ij}]_{m \times r}$ can defined as

$$l_{ij} = \begin{cases} 1, & \text{if } i \text{ is inhabitants of } r, \\ 0, & \text{otherwise.} \end{cases} \tag{2}$$

Using each membership-indicator matrix L, we can finally derive our aggregation matrices K from the users matrix $\hat{U}$. In this sense, the interpretation of K depends on the definition of L. For instance, if L aggregates based on the most cited organ (Equation 1), then $K = [k_{ij}]_{n \times n}$ and rows in K contain the characterization of n organs. However, if L splits users based on regions (Equation 2); then, $K = [k_{ij}]_{r \times n}$ and rows represent the characterization of r regions. Formally, K is given by

$$K = (L^{\mathsf{T}} L)^{-1} L^{\mathsf{T}} \hat{U}. \tag{3}$$

IV. Results and Discussion

In this section, we present the results of characterizing users based on organ transplantation conversations on Twitter from different perspectives of aggregation. First, we investigate the distribution of mentions to organs in the USA in order to grasp intuitions about them. Second, we analyze the states in the USA identifying organs highlighted in conversations, and revealing underlying similarities among them. Last, we explore the possibility of grouping users as a first step in the direction to identify common topics in the conversations of organ donation.

A. Organ Perspective

The *organ characterization* is based on the aggregation of users who behave similarly, i.e. users who mostly mention the same organ (Section III-B and Equation 1). Figure 3 shows the characterization of the six major organs transplanted in the United States; each plot represents an organ.

In order to emphasize the co-occurrence differences, the amount of attention is presented in ranked bins, from left to right. For instance, liver (top-right plot in Figure 3) tend to be mentioned more frequent with kidney, heart, and lung, respectively. Kidney is the most important organ for heart, liver, and pancreas. Heart, on the other hand, is more important for intestine, kidney, and lung. Clearly, these co-occurrences are not reciprocal.

The organ characterization can show dependencies among them, such as dual organ transplantation. For instance, dual organ transplantation is more common among the pairs heart-kidney, liver-kidney, and kidney-pancreas [1]. Although further analysis are needed, the results in Figure 3 show how the population perceive the combined importance of organs. Another plausible explanation relates to the awareness of users who are interested in one organ transplantation with other types of transplantation. Such understanding can help us to have more effective social network intervention strategies. For instance, users who are more aware of lung transplant may be more influenced to get involved in programs related to heart transplant than kidney transplant (Figure 3, bottom right).

Furthermore, the relations shown in Figure 3 might also indicate side effects or how an organ failure can lead to other organs failures [28]. People who have heart disease can have renal dysfunction which is commonly caused by diabetes

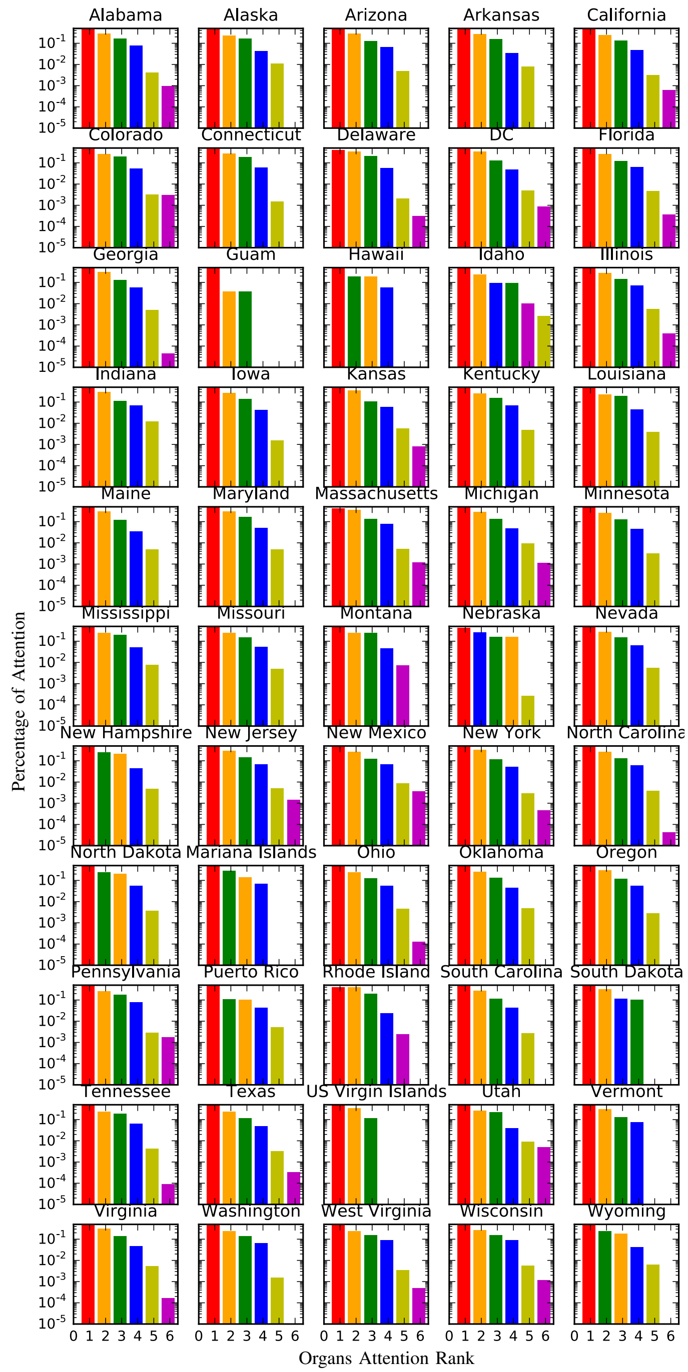


Fig. 4. Characterization of states in the USA r based on attention given to organs by Twitter users (see equations 2 and 3). The states in the USA have different distributions of mentions to organs which might indicate awareness of programs and even links between states (when they have similar characteristics). The bins in the histograms indicate the intensities of attention given to each organ. These histograms have different "shapes". For instance, most states in the USA have their first and second-most-mentioned organ as heart and kidney, which may indicate the overall "ubiquity" of these transplants [1].

and hypertension [29]. Similarly, people with heart disease develop fluid retention which damages the liver [30]. Then, a small portion of these patients needs a liver transplant and as the heart and the liver are damaged, the kidneys are usually affected. People who have liver disease tend to have renal

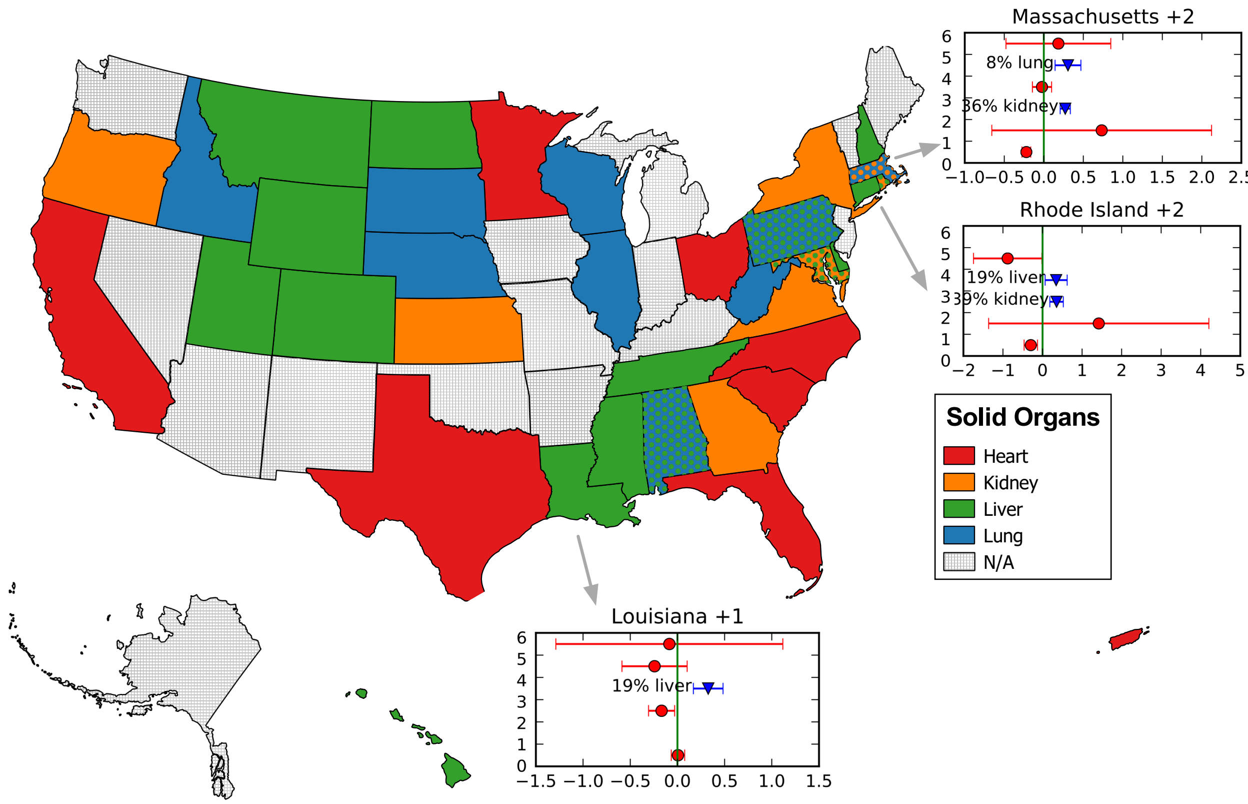


Fig. 5. The states colored according to excessive conversations about specific organs. The excess (*relative risk - RR*) is explained by deviations to other states (see Eq. 4). A state is colored by organs with significant RR (confidence interval lower limit is greater than zero). Three inset examples (Louisiana, Massachusetts, and Rhode Island) show organs' RR which is depicted in blue when they are significant.

dysfunction, some of them due to diabetes, but many others due to the liver disease dysfunction. Note that we are not arguing that the conversations indicate the co-occurrence of failures of different organs but rather that because of these co-occurrence, people may have a tendency to talk about them together in the same tweet.

Finally, the analysis of intestine is less significant, since the majority of transplants happen in pediatric patients and are only related to a small fraction of the overall organ transplants [1]. This fact leads to less reliable statistics. However, all these may reflect the coexistence of diseases and problems or the level of awareness of individual inflicted/affected by problems in one or many organs.

B. Region Perspective

In this section, we explore the geographic characterization of *regions* (as defined in Section III-B and Equation 2) where regions are the states in the USA.

Similarly to the organ characterization, a state is represented as a distribution of attention to the set of six organs. Figure 4 shows the characterization of all states and territories of the USA as histograms. Despite strong similarities, every state appears to have its own histogram shape (organ signature). We explore two aspects of this characterization:

- Since the prevalence of organs mentioned is not normally distributed, we cannot perceive the highlighted organs by comparing absolute values of mentions. For instance, from Figure 4 we tend to believe all states highlight heart.
- States seems to share underlying similarities when dealing with organs. Not only the importance rank of organs varies among states, but also in the amount of attention they give to each of them. For instance, apparently states can be split by their second most mentioned organ: kidney, liver or lung; or by the number of significant organs mentioned (3-6).

1) Identifying Highlighted Organs per State: We want to understand the impact/correlation of different states according to organ-related conversations. This might allow us to understand spatial disparities regarding organ-related conversations, identify clustering of well-defined borders of adjacent regions and geographic anomalies. For instance, is there any particular state in the USA unexpectedly associated with a specific organ-related conversation?

The simplest approach to answer this question is to count the number of users mentioning each organ and use a "winner-takes-all" strategy, i.e., the organ most cited is the one highlighted for that state. However, since some organs are much more prevalent than others, it is more likely to find a greater number of users mentioning that organ everywhere. Figure 4 shows heart as the prevalent organ in all the states in the USA. To minimize this problem, instead of using the prevalent organ in a state, we calculate the relative risk (RR) [31] of each

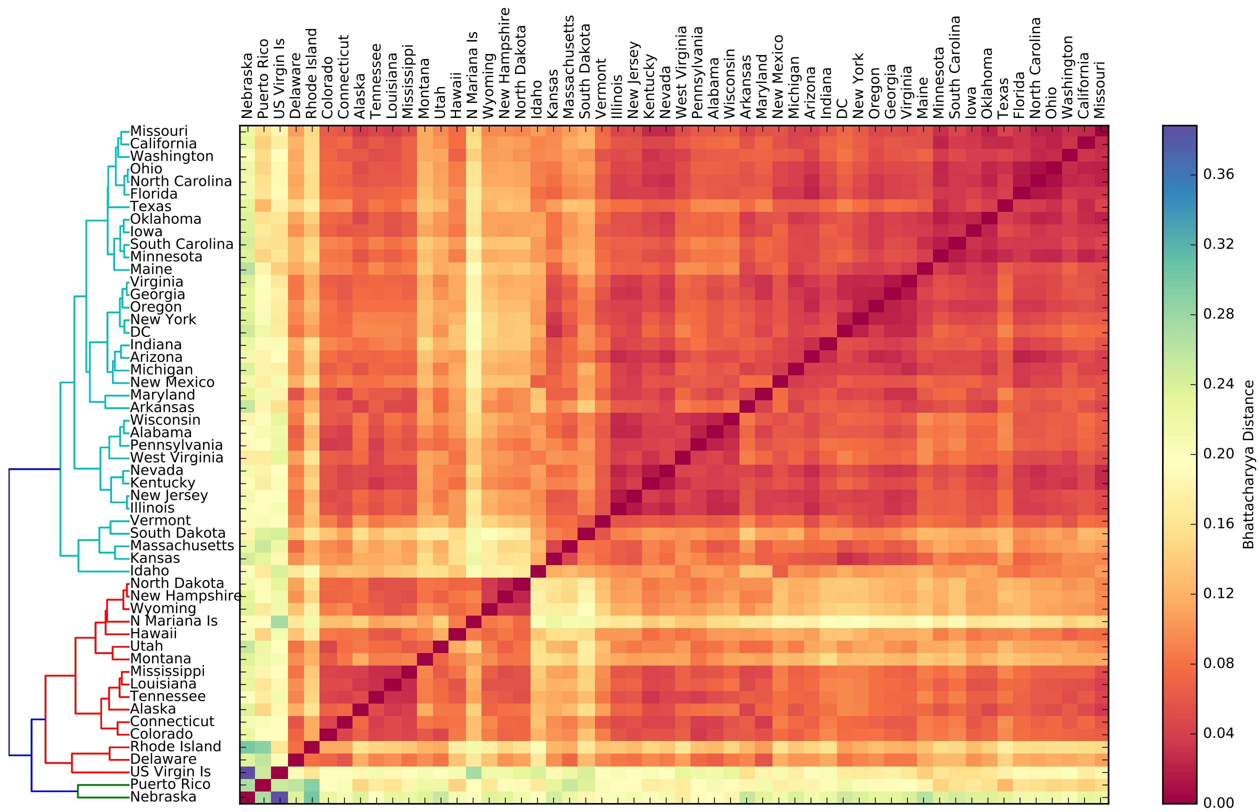


Fig. 6. Hierarchical clustering of states based on their similarity with regards to the extent of incidence of specific organ-related conversations. States are outlining zones of organ-related conversation. For instance, the states belonging to the cluster depicted in red are mostly associated with liver conversations.

organ in each state as

$$RR_{ir} = \frac{\rho_{ir}}{\rho_{in}} \ , \tag{4}$$

where ρ_{ir} and ρ_{in} are the prevalence of mention of organ i inside and outside the state r, respectively.

In this sense, the RR gives the excessive incidence of an organ in a state *relative* to the overall incidence in the rest of the USA. Since the distribution of $\log(RR_{ir})$ is approximately normal, an organ significantly exceed their expected national proportion in a state (i.e., it is highlighted), if $\log(RR_{ir}) - z_\alpha \times \sigma_{\log(RR_{ir})} > 0$. We chose $\alpha = 0.05$ for which $z_\alpha = 1.96$. Figure 5 shows the highlighted organs in each state. Although most of them have at least one organ highlighted, for some states there are no significant excess for any organ and, thus, no organ is emphasized there, while other states have more than one highlighted organ.

It is common to analyze the correlations between health-related traits geographically. For instance, the higher risk of hypertension observed in the so-called Stroke Belt in Southern USA which is associated with diet. Similarly, the increase amount of liver disease in the Western United States due to fatty liver probably associated with diet and genetic traits. Our results, for instance, show that Louisiana is associated with excess of kidney conversations while Massachusetts with both kidney and lung. Similarly, previous work analyzing geographic patterns of end-stage renal disease, kidney transplantation and deceased donors, found Kansas as the only state with a surplus of deceased kidney donors [32] in Midwestern USA. Interestingly, Kansas is also the only state in the Midwestern USA for which conversations of kidney is highly exceeding the national expectation.

2) Clustering States in the USA: In addition to identifying highlighted organs, one might be interested to investigate similarities between states considering all organs. For instance, states can be similar not only based on organs that exceed national expectation, but they can also be similar according to the organs that are unexpectedly less mentioned.

The organ distributions reveal more details about states, showing some states to be more similar between each other. We investigated the extent of similarity between the states by using the Agglomerative Clustering algorithm [33]. The hierarchical clusters can provide additional information beyond organs highlight. The elements to be clustered are states (rows of matrix K), where each component represents the probability of mentioning an organ in that state. We used the Bhattacharyya distance as the affinity (distance) metric, since it is more suitable for discrete probability distribution in comparison than other metrics, such as Euclidean distance [34].

Figure 6 shows the similarity matrix of states as a heatmap for which the lower values are associated with higher similarity. Using the dendrogram, the hierarchical clusters can be analyzed at any location on the hierarchy. Such clusters present some degree of consistence with the aforementioned results regarding the organs that are highlighted at each state (see Figure 5). For instance, Delaware, Rhode Island, and Colorado for liver as well as Oregon, Georgia and Virginia

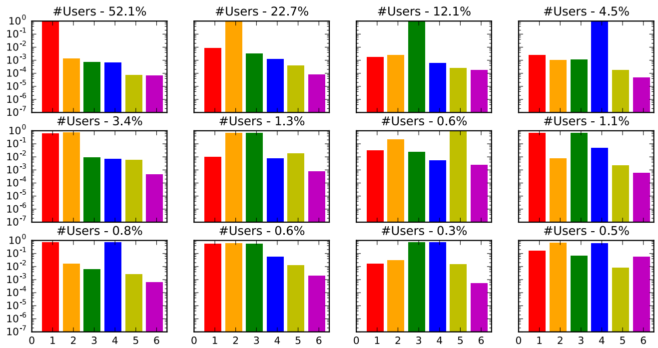


Fig. 7. Cluster of users based on their conversations on Twitter using K-Means and their relative size. We chose $k = 12$ clusters based on the silhouette coefficient, average cluster size and inertia which were 0.953, 31697.42 and 2512.27, respectively. These clusters shows possible classification of users and might be related to different users' roles in the organ donation environment.

for lung. Indeed, from the leftmost state to the rightmost state in the similarity matrix, Nebraska to Missouri, respectively, the states are outlining zones of organ-related conversation in the following order: liver (from Delaware to North Dakota), lung (from Massachusetts to Wisconsin), kidney (from New York to Virginia) and heart (from Minnesota to California). Similarly, states without a highlighted organ tend to cluster, for instance, in the zone between New Mexico and Indiana.

C. User Perspective

So far we explored the relations among organs and regions strictly according to the maximum attention and state borders, respectively. This first two characterizations can be seen as a validation phase where we could detect the richness and accuracy of the information hidden on tweets. Since, individual user characterization does encode valuable information, we can also learn from an aggregated characterization of them.

In this sense, as a preliminary investigation, we used K-Means to cluster users by their full behavior; not only based on the most-cited organ. After some empirical analysis comparing the inertia, the average cluster size, and the silhouette coefficient, we chose $k = 12$. Since we are characterizing six organs, k must be at least six in order to allow at least one cluster for each organ. Indeed, since we have approximately 72 thousand users, even the smaller cluster which is associated with 0.3% of users would still be related to roughly 2 thousand users. Therefore, a greater number of clusters could still be used.

Figure 7 brings the characterization of each cluster, as well as their relative size. Although these clusters still demand more investigations, they already seem to reveal some interesting information allowing us to identify which users present the general patterns already identified for organs (see Figure 3). We can identify the subset of users focusing on a single organ, and also users focusing on two and three organs. These clusters might even represent organ-related users with different attitudes towards organ donation. For instance, the bottom-rightmost cluster of user (see Figure 7) mention virtually all organs specially when compared with the other clusters. This information can be used again in conjunction with region to investigate possible further correlations.

V. Conclusion and Future Work

In this work, we characterized social media users and states in the USA based on their attention to different solid organs; we use markers (i.e. indicators) of awareness, norms, and behaviors towards organ donation. This characterization might lead to a better understanding of these users and their geographic variations. For instance, the geographic characterization of organ-related conversation at the state level can help us identify patterns of awareness from the angle the states in the USA. Similarly, our characterization might be used to differentiate classes of users such as health care practitioners, donors, waiting-list candidates, organ donation advocacy agencies, or simply demonstrate that different users have different behaviors towards organ donation.

The potential impact of this characterization is that it can improve the assessment of organ donation awareness approaches in the United States but also derive social intervention approaches that better fit the cultural, religious, and educational differences between states. Ultimately, this characterization can inform models of social influence to be employed in the context of organ donation aiming at designing interventions that effectively target specific groups of users.

Possible limitations of our work regards to bias in the collected data. The population of the United States is underrepresented by Twitter users since they are a highly non-uniform sample of the USA population especially with regards to geography, gender and race/ethnicity [27]. Twitter users are biased towards highly populated counties and male users. Also, depending on the region, different race/ethnicity (i.e Caucasian, African-American, Asia and Hispanic) can be over-sampled or under-sampled. For instance, the Midwestern population of United States is underrepresented among Twitter users.

Acknowledgments

Diogo Pacheco and Diego Pinheiro would like to thank the Science Without Borders program (CAPES, Brazil) for financial support under grants 0544/13-2 and 0624/14-4.

References

[1] Organ Procurement and Transplantation Network (OPTN) and Scientific Registry of Transplant Recipients (SRTR), "OPTN/SRTR 2012 Annual Data Report," Rockville, MD: Department of Health and Human Services, Health Resources and Services Administration, Tech. Rep., 2014.

[2] J. P. Merrill, J. E. Murray, J. H. Harrison, and W. R. Guild, "Successful Homotransplantation of the Kidney in an Identical Twin," *Journal of the American Medical Association*, vol. 160, no. 4, pp. 277–282, jan 1956.

[3] B. Nemes, G. Gámán, W. G. Polak, F. Gelley, T. Hara, S. Ono, Z. Baimakhanov, L. Piros, and S. Eguchi, "Extended criteria donors in liver transplantation Part I: reviewing the impact of determining factors," *Expert Review of Gastroenterology & Hepatology*, pp. 1–13, mar 2016.

[4] J. A. Leithead, N. Rajoriya, B. K. Gunson, P. Muiesan, and J. W. Ferguson, "The evolving use of higher risk grafts is associated with an increased incidence of acute kidney injury after liver transplantation," *Journal of Hepatology*, vol. 60, no. 6, pp. 1180–1186, jun 2014.

[5] D. Pinheiro, F. Hamad, M. Cadeiras, R. Menezes, and N. Nezamoddini-Kachouie, "A data science approach for quantifying spatio-temporal effects to graft failures in organ transplantation," In Press.

[6] S. Venugopal, E. Stoner, M. Cadeiras, and R. Menezes, "Understanding organ transplantation in the USA using geographical social networks," *Social Network Analysis and Mining*, vol. 3, no. 3, pp. 457–473, sep 2013.

[7] D. A. DuBay, P. A. MacLennan, R. D. Reed, M. Fouad, M. Martin, C. B. Meeks, G. Taylor, M. L. Kilgore, M. Tankersley, S. H. Gray, J. A. White, D. E. Eckhoff, and J. E. Locke, "The Impact of Proposed Changes in Liver Allocation Policy on Cold Ischemia Times and Organ Transportation Costs," *American Journal of Transplantation*, vol. 15, no. 2, pp. 541–546, feb 2015.

[8] A. M. Cameron, A. B. Massie, C. E. Alexander, B. Stewart, R. A. Montgomery, N. R. Benavides, G. D. Fleming, and D. L. Segev, "Social Media and Organ Donor Registration: The Facebook Effect," *American Journal of Transplantation*, vol. 13, no. 8, pp. 2059–2065, aug 2013.

[9] M. Cadeiras, M. Soroya, S. Patel, C. Starling, G. Lim, K. Dong, R. Menezes, and M. Deng, "Using social media as a method of increasing organ donation," *Circulation*, vol. 130, no. Suppl 2, pp. A18 386–A18 386, 2014.

[10] L. A. Siminoff, "Factors Influencing Families' Consent for Donation of Solid Organs for Transplantation," *JAMA*, vol. 286, no. 1, p. 71, jul 2001.

[11] L. A. Siminoff, H. M. Traino, and M. W. Genderson, "Communicating Effectively About Organ Donation," *Transplantation Direct*, vol. 1, no. 2, pp. 1–9, mar 2015.

[12] D. F. Pacheco, D. Pinheiro, F. B. Lima-Neto, E. Ribeiro, and R. Menezes, "Characterization of Football Supporters from Twitter Conversations," in *2016 IEEE/WIC/ACM International Conference on Web Intelligence (WI' 16)*, 2016, pp. 169–176.

[13] A. Al-rubaye, "Extracting Social Structures from Conversations in Twitter A Case Study on Health-Related Posts," in *Proceedings of the 27th ACM Conference on Hypertext and Social Media*. New York, New York, USA: ACM Press, 2016, pp. 5–13.

[14] Y. Kryvasheyeu, H. Chen, N. Obradovich, E. Moro, P. Van Hentenryck, J. Fowler, and M. Cebrian, "Rapid assessment of disaster damage using social media activity," *Science Advances*, vol. 2, no. 3, pp. e1 500 779–e1 500 779, mar 2016.

[15] L. Laranjo, A. Arguel, A. L. Neves, A. M. Gallagher, R. Kaplan, N. Mortimer, G. A. Mendes, and A. Y. S. Lau, "The influence of social networking sites on health behavior change: a systematic review and meta-analysis," *Journal of the American Medical Informatics Association*, jul 2014.

[16] S. S. Bull, D. K. Levine, S. R. Black, S. J. Schmiege, and J. Santelli, "Social MediaDelivered Sexual Health Intervention," *American Journal of Preventive Medicine*, vol. 43, no. 5, pp. 467–474, nov 2012.

[17] D. N. Cavallo, D. F. Tate, A. V. Ries, J. D. Brown, R. F. DeVellis, and A. S. Ammerman, "A Social MediaBased Physical Activity Intervention," *American Journal of Preventive Medicine*, vol. 43, no. 5, pp. 527–532, nov 2012.

[18] D. Foster, C. Linehan, B. Kirman, S. Lawson, and G. James, "Motivating physical activity at work," in *Proceedings of the 14th International Academic MindTrek Conference on Envisioning Future Media Environments - MindTrek '10*. New York, New York, USA: ACM Press, 2010, p. 111.

[19] A. B. Mayer and J. A. Harrison, "Safe Eats: An Evaluation of the Use of Social Media for Food Safety Education," *Journal of Food Protection*, vol. 75, no. 8, pp. 1453–1463, aug 2012.

[20] G. Turner-McGrievy and D. Tate, "Tweets, Apps, and Pods: Results of the 6-Month Mobile Pounds Off Digitally (Mobile POD) Randomized Weight-Loss Intervention Among Adults," *Journal of Medical Internet Research*, vol. 13, no. 4, p. e120, dec 2011.

[21] C. Pechmann, L. Pan, K. Delucchi, C. M. Lakon, and J. J. Prochaska, "Development of a Twitter-Based Intervention for Smoking Cessation that Encourages High-Quality Social Media Interactions via Automessages," *Journal of Medical Internet Research*, vol. 17, no. 2, p. e50, feb 2015.

[22] A. Ritter, C. Cherry, and B. Dolan, "Unsupervised modeling of twitter conversations," *The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, no. June, pp. 172–180, 2010.

[23] C. A. Bail, "Cultural carrying capacity: Organ donation advocacy, discursive framing, and social media engagement," *Social Science & Medicine*, jan 2016.

[24] A. Chang, E. E. Anderson, H. T. Turner, D. Shoham, S. H. Hou, and M. Grams, "Identifying potential kidney donors using social networking web sites," *Clinical Transplantation*, vol. 27, no. 3, pp. E320–E326, may 2013.

[25] A. D'Alessandro, J. Peltier, and A. Dahl, "Use of social media and college student organizations to increase support for organ donation and advocacy: a case report," *Progress in Transplantation*, vol. 22, no. 4, pp. 436–441, dec 2012.

[26] F. Morstatter, J. Pfeffer, H. Liu, and K. Carley, "Is the sample good enough? Comparing data from Twitter's streaming API with Twitter's firehose," in *Seventh International AAAI Conference on Weblogs and Social Media*, 2013, pp. 400–408.

[27] A. Mislove, S. Lehmann, Y.-y. Ahn, J.-p. Onnela, and J. N. Rosenquist, "Understanding the Demographics of Twitter Users," *Artificial Intelligence*, pp. 554–557, 2011.

[28] C. A. Hidalgo, N. Blumm, A.-L. Barabási, and N. A. Christakis, "A Dynamic Network Approach for the Study of Human Phenotypes," *PLoS Computational Biology*, vol. 5, no. 4, p. e1000353, apr 2009.

[29] J. M. Schaffer, P. Chiu, S. K. Singh, P. E. Oyer, B. A. Reitz, and H. R. Mallidi, "Combined Heart-Liver Transplantation in the MELD Era: Do Waitlisted Patients Require Exception Status?" *American Journal of Transplantation*, vol. 14, no. 3, pp. 647–659, mar 2014.

[30] ——, "Heart and Combined Heart-Kidney Transplantation in Patients With Concomitant Renal Insufficiency and End-Stage Heart Failure," *American Journal of Transplantation*, vol. 14, no. 2, pp. 384–396, feb 2014.

[31] C. L. Sistrom and C. W. Garvan, "Proportions, odds, and risk." *Radiology*, vol. 230, no. 1, pp. 12–19, jan 2004.

[32] Y. Cao, K. Stewart, and R. Kalil, "Geographic patterns of end-stage renal disease and kidney transplants in the Midwestern United States," *Applied Geography*, vol. 71, pp. 133–143, jun 2016.

[33] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and . Duchesnay, "Scikit-learn: Machine Learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2012.

[34] T. Kailath, "The Divergence and Bhattacharyya Distance Measures in Signal Selection," *IEEE Transactions on Communication Technology*, vol. 15, no. 1, pp. 52–60, 2 1967.