The Current State of Online Social Networking for the Health Community: Where Trust Modeling Research May Be of Value

Daniel Ohashi
Computer Science
University of Waterloo
Waterloo, Ontario N2L 3G1
daohashi@uwaterloo.ca

Robin Cohen
Computer Science
University of Waterloo
Waterloo, Ontario N2L 3G1
rcohen@uwaterloo.ca

Xiaotian Fu
Computer Science
University of Waterloo
Waterloo, Ontario N2L 3G1
xfu@uwaterloo.ca

ABSTRACT

In this paper, we discuss the prevalence of misleading information in health-oriented online social networks and discussion boards. With increasing numbers of patients and caregivers browsing online for insights into how to address their specific health problems, and with a growing tendency to value the opinions of peers when making choices about healthcare solutions, it is important for computer science researchers to develop strategies that can be introduced to enable each person to be better informed.

We begin with a brief report on some of the activity currently observed in online communities. From here, we advocate the use of trust modeling, an approach examined by artificial intelligence researchers in the subfield of multi-agent systems. In particular, we sketch some specific solutions to integrate, based on frameworks that we have developed which have been validated as effective in presenting beneficial messages to users. We conclude with a view to the future, both with respect to refinement of our trust modeling solutions, and with respect to engagement of government, healthcare providers and individuals.

CCS CONCEPTS

• **Social and professional topics → Health information exchanges**; • **Computing methodologies → Multi-agent planning**;

KEYWORDS

online health communities, social media, personalized recommendation, trust modeling

ACM Reference format:
Daniel Ohashi, Robin Cohen, and Xiaotian Fu. 2017. The Current State of Online Social Networking for the Health Community: Where Trust Modeling Research May Be of Value. In *Proceedings of DH '17, London, United Kingdom, July 02-05, 2017,* 10 pages.
https://doi.org/http://dx.doi.org/10.1145/3079452.3079462

The authors have declared that no competing interests exist.

DH '17, July 02-05, 2017, London, United Kingdom

ACM ISBN 978-1-4503-5249-9/17/07…$15.00
http://dx.doi.org/10.1145/3079452.3079462

1 ONLINE SOCIAL NETWORKING FOR HEALTH

1.1 Introduction

With the rise of the Information Age, many industries have experienced a shift to the digital world. Health and medicine are no exception. Today, doctor consultations, diagnostic tools and even health education exist online. As the trend of digital health becomes more prevalent, an increasing number of patients and caregivers choose to use online social networks and discussion boards in order to connect with each other. Through social networks, people provide inspiration, exchange information and dispense advice. While these networks can be empowering for users, inaccurate content can also arise, spreading rapidly and causing harm.

In an ideal context one could imagine the existence of moderators who are medical professionals, with sufficient capacity to monitor all the activity in the network, resulting in a suitable reduction of content. Failing this, users could perhaps simply check background information shared by posters, and come to their own conclusions about whom to believe. Navigation of these waters can, however, be quite difficult.

We begin this paper with a section introducing two online social networks[1] for health, PatientsLikeMe and HealthTap. We observe how participants interact within each of the networks, and then discuss typical reactions to fraudulent content. Ultimately, we contrast the two networks, acknowledge some other valuable health forums, and then reach a conclusion about the possible value of introducing an automated intelligent trust modeling system into these communities.

1.1.1 PatientsLikeMe Community. PatientsLikeMe is an online social network of patients. Founded in 2004, this community provides a platform for patients with same disease or similar to conditions to connect with each other and and to and share their experience. The participants in this community are encouraged to give treatment advice, contribute their health data and share their resources, as they believe such information would help others avoid the repetitive trial-and-error treatment approach, and will speed up the pace for researchers. Since 2016, PatientsLikeMe has grown into a network of more than 400,000 members [1].

The three primary methods for a user to interact are to contribute their own health data, to participate in open forums or to publish

[1]By online social networks we mean platforms hosted on the web in order for peers to connect with each other, posting and viewing comments. The term online social networks is used in various research papers including [3] which connects the term with an "interactive portal" for "user engagement" resulting in "a community" to "discuss and share information" on the web.

journals. Contributing their own health data is an effective way for patients to communicate with each other. PatientsLikeMe provides a public profile for each user, so that the user can fill in their own information. The profile covers a broad range of areas including their basic information, their condition and symptoms, their treatments and their opinions on effectiveness of the treatments. Since the profiles are public to other users (except for the sensitive parts of the personal information section), patients have an opportunity to learn about diseases and treatments from others by viewing their health data. These profile data have also benefited researchers by accelerating clinical discovery, as claimed by a paper from 2011 [24]. Exchanging ideas in open forums is another effective way, and also, the most direct way participants interact with each other. PatientsLikeMe uses the information on your disease and symptoms to auto-enroll you into a forum with all other patients having similar distress. In the forum, a user can view a number of different topics and express their ideas by posting threads. In these threads, some comments are found to be non-informative, as greetings or inspirations (EX1 of Appendix A, while others are found to be facts about themselves, like what have they been experienced (EX2 of Appendix A). Some comments make suggestions on a general approach, like what to do to temporarily relieve the pain (EX3 of Appendix A), while others make recommendations on a specific medicine (EX4 of Appendix A). See Appendix A for examples of each type.

Another way patients interact with each other is via journals, as a place for patients to express their personal feelings. While responses in forums are intended for everyone in the community, commenting on others'ĂŹ journals is intended for a specific patient. By having a conversion in a personal journal, patients can also exchange ideas and make suggestions, just like posting threads in a forum.

1.1.2 Credibility of participants of PatientsLikeMe. PatientsLikeMe does not require one to sign up or sign in in order to see a patient's profile or general data on a type of disease, but it does require one to sign in to view the forums and actively participate in the forums. A user can sign in as a patient, a researcher or a clinician .

To sign up as a patient, PatientsLikeMe requires only an email address. There is no validation check on the email address (an example of a modern validation check would be to enclose a link in an email to the email address, and instruct a user to click on the link to complete the new user registration), and no further information on the identity of the user is required. Therefore, the email address serves only as a user name, instead of a type of identification.

Signing up as a researcher is much more restricted. Besides an email address, the researcher "must have a PhD and be affiliated with an academic institution" [2]. The sign up won't succeed unless the researcher provides a link to the research and evidence of credentials. Signing up as a clinician is also restricted. The clinician needs to provide a valid professional license in order to complete the registration. Compared to patients, the identification of researchers and clinicians is checked more carefully.

Once a user has signed in, he/she can comment freely anywhere in this community. The platform does not perform any relevancy checks, spam checks or truth checks on the comments. A user can potentially post a link to a malicious website, embed an advertisement in a comment, claim a misleading statement or give some incorrect information, intentionally or unintentionally. Besides a metric of profile stars in a range of zero to three showing the amount of health profile information a user has shared, the platform does not give any indicators for credibility[2] measurement on an account based on one's identity and past activities. It might be difficult for a participant to decide whom to believe or what to believe, because all other participants and their postings appear to be equally trustworthy[3], or perhaps equally untrustworthy. Although some participants might claim a user who participates more might generally have a higher credibility and immunity from deliberate frauds, it is hard to believe that they are not suffering from any misleading information and won't unintentionally share that information in the network. Foreseeably, introduction of a trust modelling system to this platform would help users make better decisions.

For patients in a health community, the most important postings are ones that suggest to them a specific product, practice or therapy. Luckily, not all patients use online social networks as their primary information source. They will likely consult a doctor before they take an action, so doctors can filter out some unwise suggestions before they create harm to patients. In addition, the most lethal fault of wrongful usage of prescription drugs is mitigated by a well-developed drug regulatory system, so the impact of misleading information on patients is reduced.

Researchers, on the other hand, focus more on the data generated from the online social network. Since any registered user can publish his/her health data, an intentional contribution of misleading data can create noise for medical research and data compilations.

1.2 HealthTap Community

HealthTap is an online social network of patients and doctors. Founded in 2010, it provides a platform where patients can get help from a large group of professional doctors. A patient can get immediate, personalized answers from a bank of doctors via sending online messages, or a patient can consult through a video chat. By moving health services online, HealthTap aims to reduce inefficiency in healthcare delivery and to improve healthcare access. HealthTap is a commercial organization, and some of its services have fees associated with it.

In HealthTap, interactions are between patients and doctors. A patient cannot see the profile of another patient, or even know the existence of another patient. The isolation of patients limits the variety of method of interaction in this community. The two primary ways patients can interact with doctors are receiving programs about health issues and asking private questions.

Some doctors in HealthTap have a job of compiling health-related readings. Depending on a patient's interest and health condition, they will release the programs to a patient from time to time. At the same time, HealthTap shares news on health related topics to

[2]By credibility we mean the extent to which a statement made by a peer is considered to be believable by the recipient, where a highly credible peer is one whose statements are expected to be believable most of the time.
[3]Trustworthiness refers to an expectation of not being misleading or deceptive. Trust modeling is a topic of study within the artificial intelligence subfield of multiagent systems. We expand on this introduction in Section 2.

patients on a regular basis. However, a user can only pick a topic he is interested in viewing, but not to make a comment.

The most predominant way of interacting in HealthTap is to ask questions to a doctor. The platform links a patient to a team of real world doctors, who are specialized in the patient'ĂŹs condition, and a patient can ask them questions at any time. Patients enclose their questions in an online dialogue form, and they will get notified once answers to the questions are received, just like how email works. Notably, answers are all provided by professional doctors, and are usually endorsed by several other doctors. All questions on this platform are private, unless a question is considered by HealthTap to be a frequently asked and model question and then makes it visible to the public. Not surprisingly, HealthTap also does not allow other patients to make comments under a public question.

Similar to PatientsLikeMe, signing up as a patient in HealthTap requires only an email address, with no validation on the email address, while signing up as a doctor requires a professional medical license. Doctors are well identified and authenticated during the registration steps. The response-and-endorse mechanism makes the answers to the questions more robust. Due to the limitation of interaction between patients, the credibility issue in HealthTap is more controlled and less of a concern.

1.3 Some other online social networks on health

There are many other social networks of health communities on the Internet besides PatientsLikeMe and HealthTap. Two of them include SubredditHealth and Patient.info. Notably SubredditHealth is an open forum network used by 124,000 users on Reddit, and Patient.info is the number one ranking by Google if searching with keyword "patient forum". Both of them are forum-styled networks, more similar to PatientsLikeMe than HealthTap.

Our research has in fact revealed a fairly rich and varied environment for patients and caregivers to connect in social media. In Appendix B we display a chart of our current findings. We also expand Appendix A to present additional examples of exchanges we discovered in these networks, drawing attention to the fact that concern over misleading information arises in a variety of current contexts.

1.4 Conclusion

In some of the existing online social networks of health communities like HealthTap where patients only interact with doctors, credibility issues are more controlled, and an introduction of trust modelling system is perhaps not a priority. In other networks like PatientsLikeMe where patients can interact among themselves, some serious concerns have been raised on credibility of people and trustworthiness of information. Foreseeably, research on trust modelling can be of great value.

2 TRUST MODELING TO IMPROVE ONLINE HEALTH NETWORKS

Trust modeling researchers in the artificial intelligence subfield of multiagent systems are examining whether one agent (a trustor) will trust another (a trustee), of value in directing the decision making of groups of agents. This has been expressed as estimating the probability that a trustee will perform as expected on a future obligation. A predominant paradigm for performing this reasoning is to use a beta probability density function [14] to anticipate future behaviour based on prior knowledge. Trustors can be informed by peers known as advisors, who supply opinions about whether another peer is trustworthy or not, typically as a binary value, necessitating a step where advisor trustworthiness is then also modeled as well [23, 26].

We now introduce a few specific trust modeling frameworks which show promise in assisting users of online social networks for health.

2.1 Modeling Credibility of Messages

The CRED-Trust model of Sardana and Cohen [19, 20], aims to address an issue that arises when peer advice is considered, within contexts such as online social networks. If the credibility of a peer is explicitly modeled and learned, over time, then this factor can be combined with a representation of peer similarity, in order to reason about whether a message should be shown to a user, or not. The overall goal therefore is to decide, for each message, whether to display it or not, as a kind of filter which may enable deflection of misleading information. And this decision for showing messages is predicated on the predicted benefit of that message, for this user. This in turn is computed on the basis of α and β factors which are integrated into a beta reputation function. The expected value of what is referred to as a Beta density function is generated on the basis of:

$$E[p] = \frac{\alpha}{\alpha + \beta}$$

The probability of some event occurring in the future is determined on the basis of the number of positive previous occurrences and the number of negative previous occurrences. The metric then for trust modeling is to have high trust in an entity if one expects positive outcomes in the future. Note as well that this allows trust beliefs to evolve over time.

In certain online social networks, peers may react to messages that have been posted. Predicted benefit of an annotation or commentary on a message a peer is handled neatly by what Sardana terms the LOAR model of Champaign and Cohen [5]: integrating global reputations for an annotation with predicted local benefit of the message, tempered by factors of similarity between the raters and the user who is being assisted (the one for whom some messages may be filtered). The CRED-Trust model above respects the same desire to reason about the contributions of peers when determining message benefit. It also continues to offer a heuristic solution that focuses on the factors of similarity and credibility. But CRED-Trust allows for misleading "herd' mentality to be overcome: simply being similar and popular does not cause a message to be shown. High credibility and low similarity (for example, the voice of an established medical professional within a network) can now be valued, for instance. This is achieved by CRED-Trust's solution of adjusting the α and β calculations as in the formulae shown in Algorithm 1. Note that the Hamming ratio is a calculation derived on the basis of the differences between the common ratings of two users, where a Hamming ratio of 0 means that the two strings of values are identical and a ratio of 1 means totally different.

Algorithm 1: Delivering a predicted benefit using similarity and credibility (CredTrust)

Input: The current user, u, his set of peers, P, their credibility scores, $c_p \in \{0, 1\}$, and their corresponding ratings for the annotation in focus, $r_p \in \{0, 1\}$

Output: Parameters α^* and β^* to a beta distribution describing trust in the current annotation

$\alpha^* = \beta^* = 1$ `// At the start, user has a uniform expectation about the message`
foreach $p \in P$ **do**
  $h_{up} \longleftarrow computeHammingRatio(u, p)$
  `// Perform a Bayesian update after discounting heuristic`
  if $r_p == 0$ **then**
    `// Adjust the similarity weight by credibility:`
    $\alpha^* += h_{up} \cdot (1 - c_p)$
    $\beta^* += 1 - h_{up} \cdot (1 - c_p)$
  else
    `// Dampen update by credibility`
    $\alpha^* += c_p \cdot (1 - h_{up})$
    $\beta^* += c_p \cdot h_{up}$
  end
end

As a quick summary, appealing to the CRED-Trust model in order to evaluate the suitability of each message in an online social network for a user would amount to: using similarity and credibility in order to determine the α and β values, using those values in order to calculate the predicted benefit of a message (on a scale of 0 to 1, where 1 is very likely to be beneficial) and then applying some threshold in order to decide whether the message should be shown (is above the threshold) or not, to that user.

The small example included in [19] sheds further light on this process. Essentially, one can predict whether a new commentary added to the network should or should not be shown to the user, by performing an aggregate calculation of message benefit, based on a series of messages already posted to the network by peers.

The CRED-Trust model is a valuable proposal to consider when reasoning about whether a message is above a threshold of value in order to display it for a user. This is known in part because that model has been validated through extensive simulations which measure Matthew's Correlation Coefficient (a metric which aims to ensure true positives, false positives, true negatives and false negatives are all handled well). The plots from these simulations are conducted in a setting where the ground truth (whether a user likes or does not like a message) is known; converting a set of these to test data instead of training data then allows one to determine whether the algorithms produce desired results.

All of this is done, however, in a context where one is viewing the actions of a peer, namely their rating of messages that have been authored by others. It is the rating behaviour that the framework ultimately models and reasons with, as it offers both a representation of the reputation of the new commentary and a prediction about whether this kind of message will be of value to the user at hand.

The context in which this framework may be of value is therefore one where a user is able to view both messages from peers and also ratings of those messages. If the social network does not support a kind of thumbs-up or thumbs-down reaction from readers, then in order to be reasoning with the ratings of others in the network, one would have to imagine further processing, such as analysis of messages written by readers in reaction to other posts, to conclude whether the sentiment is predominantly negative or positive. Once done, the rating of that peer can be inferred and Sardana's model can be applied in full. We return to a discussion of future directions in Section 3.

2.2 Beyond Heuristic Approaches

Sardana critiques his own solution for message recommendation based on credibility, labelling the approach as one that is derived from "heuristics" for combining a representation of similarity and credibility. While the CRED-Trust system is demonstrated to be an improvement over other heuristic peer-based trust modeling solutions such as the one derived for intelligent tutoring contexts by Champaign and Cohen [5], that competitor is also designed as a heuristic solution. One may be able to imagine additional value to solutions inspired by intelligent tutoring, when used for discussion boards aimed more at enabling a patient or caregiver to progressively master a course of material on a health topic. However, there is also important improvement when one moves forward to a solution predicated on deeper learning about the peers in the environment. This is the intention of Sardana's second framework, referred to as Bayes-Trust [21].

Sardana himself explains this novel direction well, as one that [18]: "... uses Bayesian learning to derive an observation function that combines user features in a way that is statistically correct, given the environment" and "incorporates a notion of user utilities in order how to act on the basis of evaluated trust".

The overall proposal is depicted as a Partially Observable Markov Decision process where

- S = {good,bad} :: each message can be either "good" or "bad"
- A = {recommend, reject, poll} :: the agent can choose to recommend a message, to reject it, or to poll advisors
- O = {(rating, similarity, credibility)} :: if an agent polls for advice, it receives an observation tuple consisting of the received rating, the advisor similarity (as it appears in LOAR), the advisor credibility, etc.
  - In general, each observation tuple consists of features that are (hopefully strongly) correlated with the underlying message state. The similarity feature, for example, measures the degree to which a user's past ratings agree with a given advisor's past ratings in the hopes that this feature has predictive value for future ratings as well.

   - An assignment of values to an observation tuple corresponds to a row in a conditional probability table (CPT) that reflects the probability of seeing a certain trio of (rating, similarity, credibility) values given the ground truth that a message is good/bad.

- $T : P(s'|s)$ :: the state transition probabilities (an identity function, since the underlying message state does not change)
- Discount factor $0 \leq \gamma < 1$
- Infinite horizon h
- $\Omega : P(o'|s', a)$ :: the probability of an observation given the state and action
- $R : A \times S \rightarrow \mathbb{R}$ :: a reward function that encodes the desirability of each state for a particular agent

To briefly summarize how this framework can be used to determine whether or not to show a message to a user, one imagines that there is an underlying state for each message (that it is either bad or good to show the user) and we are progressively learning this. As one receives advice from peers about a message, evidence builds about whether this message truly yields a positive reward for this user, and thus belief updates can be performed. Decision making agents can then reason about whether the expected utility of showing this message is greater than the expected utility of rejecting it, and with a suitably high prediction of benefit to the user, that message might be shown. The process is described in greater detail in Sardana's thesis [18].

The Bayes-Trust model has also been validated, this time employing data from real online social networking environments, Reddit and Epinions, two settings where peers explicitly indicate their ratings for messages. With truly vast numbers of messages and peers, and even with only small subsets of peers actively involved in posting or rating, the Bayes-Trust model works well, using repeated random subsampling. The plots measure MCC for the Reddit context and mean average error for that of Epinions (where ratings are not binary but instead on a scale of 1 to 5). Bayes-Trust's performance in both contexts is excellent, in comparison with coded competitors (LOAR and BLADE (another Bayesian trust model, developed by Regan et al. [17])).

A primary goal for future work with Bayes-Trust is to integrate additional features, beyond similarity and credibility of raters. In fact, being able to progressively learn the similarity and credibility of message authors is most desirable. Here, we might imagine that certain users have a stronger interest in certain kinds of content or a more favourable reaction to certain kinds of message authors. Future work can examine which features to represent and various ways in which to derive values, through implicit modeling.

Readers interested in learning more about the various approaches being developed within the multiagent trust modeling community are referred to such general references as [25], which includes an excellent overview of multiagent trust paradigms, or a longer exposition of various popular models [11]. Seminal models were originally derived from examining the context of electronic marketplaces [26] though there has concurrently been an interest in delving into the many facets required for modeling the peers [13]. The challenge to properly define the notion of trust is also made clear through ongoing research into developing formal definitions of the term [8, 9].

3 A LOOK TO THE FUTURE FOR TRUST AND ONLINE SOCIAL NETWORKS FOR HEALTH

In this paper, we have provided some examples of online social networks currently being used to manage healthcare, by both patients and caregivers. We have provided examples of postings in these networks which suggest that it would be beneficial to reason more carefully about which messages to show to each user. We have then explained some solutions for the management of messages in social networks which are based on trust modeling and which propose to show only those messages that exceed a threshold of benefit to the user. The models that we outlined were primarily intended to function in environments where it is possible to be modeling similarity, credibility and rating behaviour. A fairly lightweight but easily constructed approach grounded in heuristic formulae is available (CRED-Trust). An alternate framework which reasons from first principles, in a Bayesian interpretation, is also an option (Bayes-Trust).

For online social networks where healthcare is the topic at hand, a few challenges remain in order to make progress. If we are to imagine that not all messages will be shown to each participant, then some kind of filtering algorithm will need to be in effect, before a user is able to view any of the content. This arrangement would need to be agreeable, at the very least, to the user who is about to join the network and who is invested in being informed more carefully about possible treatment options. User acceptance of automated filtering options is therefore a topic for further study.

It is worth noting that an approach where credibility is modeled simply on the basis of the role of the poster (e.g. medical professional vs. caregiver vs. patient) can turn out to be too simplistic, so that progressive modeling of trustworthiness really is of value. This conclusion was reached in papers such as [4, 6] where at times even those with professional stature turned out to be less expert in certain topic areas. Thankfully, as explained in our description of current social networks, at least some environments take important steps to ensure that if a peer is labelled as a professional, that standing is at least true. Examining more closely the challenges in blanket acceptance of advice from medical professionals is also a valuable topic for the future, for contexts such as HealthTap (certainly when designing automated solutions).

In order to repurpose the specific trust models introduced in this paper for the contexts of healthcare, the other primary hurdle to overcome is to adjust those models to function well if the social network has not been set up to support ratings. The ideal goal would be to be modeling each message author and then each message as suitably trustworthy, without any representation of ratings. Certainly there are many multi-agent trust models that do not assume ratings of messages; these still focus primarily on assessing whether each peer is suitably reliable or not, based on past behaviour. An interesting approach that proposes to carefully combine private and public information is that of [26]. This model was designed to operate in other contexts, however, such as electronic marketplaces. Perhaps the most profitable step forward would be to continue to ground the solution in the models of Sardana (designed for social networks), but to imagine integrating additional calculations into the algorithms. Examining how to move towards richer calculations, inspired by these starting points, is a suitably challenging avenue for future exploration.

We are encouraged to learn, as well, that caregivers are trying to leverage social networking as well, in order to support each other and to learn from each other [10]. One could imagine a role for trust modeling specifically aimed at these contexts, in order to present the most relevant messages to each user. This is yet another direction for future research.

Regardless, we feel that social networks are becoming second nature for many online users and because of this we do need to examine what kinds of healthcare information is being shared today, and to look towards offering frameworks that can improve those experiences for users. Simply having each user self policing and using their own judgment when reaching conclusions from the information that is present leaves the door open for some truly undesirable consequences. As such, we would like as well to see more investment from governments to be running their own moderated discussion boards, as part of the solution.

Other related work that is of value to examine, for efforts in the future, include that of researchers who have been studying other outcomes derived from online social networking for healthcare. For example [15] are interested in mapping the topology of online social networks, identifying leading influencers as part of the analysis of opinion dynamics. The work of [7] is also interested in how the healthy behaviour of users may be affected by the posts that are prevalent within their online social networks. It would certainly be interesting for future work to integrate a number of different analyses of the network postings and the network participants in order to continue to provide solutions that assist users in receiving the most effective information for their healthcare management. Finally, the research of [12] acknowledges the problem of conflicting advice appearing in online health communities but suggests that designing better approaches for interweaving patient and clinical expertise may be the best step forward. They propose recognizing long threads as possible flags of discord amongst participants, if patients are expressing opinions on medical discussions. These may then become a signal that health professionals should participate. It may be useful to also incorporate these alternate methods for identifying problematic scenarios.

There has also been a surge of interest, of late, in analyzing the content posted to online social networks, with a view to addressing flaming or even hateful comments, which are considered to be unacceptable behaviour [16, 22]. This is a different concern with the kinds of messages which may surface within the healthcare discussion boards, as all such social gatherings are susceptible to unacceptable content, if unmoderated. One conclusion is, again, that there is important motivation for governments to offer some moderated forums. Another is that it may be possible to imagine expanding the analysis currently proposed here with respect to trust modeling, to include other intelligent reasoning about the set of posts, as part of an overall filtering proposal.

REFERENCES

[1] 2017. About us. (8 February 2017). https://www.patientslikeme.com/about

[2] 2017. Doctor researcher registrations. (8 February 2017). https://www.patientslikeme.com/users/doctor_researcher_registrations/new

[3] Nilufar Baghaei, Jill Freyne, Stephen Kimani, Greg Smith, Shlomo Berkovsky, Dipak Bhandari, Nathalie Colineau, and Cecile Paris. 2009. SOFA: An Online Social Network for Engaging and Motivating Families to Adopt a Healthy Lifestyle. In *Proceedings of the 21st Annual Conference of the Australian Computer-Human Interaction Special Interest Group: Design: Open 24/7 (OZCHI '09)*. ACM, New York, NY, USA, 269–272. DOI:http://dx.doi.org/10.1145/1738826.1738871

[4] John Champaign, Robin Cohen, and Disney Yan Lam. 2015. Empowering Patients and Caregivers to Manage Healthcare Via Streamlined Presentation of Web Objects Selected by Modeling Learning Benefits Obtained by Similar Peers. *ACM Trans. Intell. Syst. Technol.* 6, 4, Article 54 (July 2015), 41 pages. DOI:http://dx.doi.org/10.1145/2700480

[5] John Champaign, Jie Zhang, and Robin Cohen. 2011. Coping with Poor Advice from Peers in Peer-based Intelligent Tutoring: The Case of Avoiding Bad Annotations of Learning Objects. In *Proceedings of the 19th International Conference on User Modeling, Adaption, and Personalization (UMAP'11)*. Springer-Verlag, Berlin, Heidelberg, 38–49. http://dl.acm.org/citation.cfm?id=2021855.2021860

[6] R. Cohen, N. Sardana, K. Rahim, D. Y. Lam, M. Li, O. Maccarthy, E. Woo, J. Zhang, and G. Guo. 2013. Recommending Messages to Users in Social Networks: A Cross-Site Study. In *Proceedings of the 2013 12th International Conference on Machine Learning and Applications - Volume 02 (ICMLA '13)*. IEEE Computer Society, Washington, DC, USA, 445–450. DOI:http://dx.doi.org/10.1109/ICMLA.2013.160

[7] Javid Ebrahimi, NhatHai Phan, Dejing Dou, Brigitte Piniewski, and David Kil. 2016. Characterizing Physical Activity in a Health Social Network. In *Proceedings of the 6th International Conference on Digital Health Conference (DH '16)*. ACM, New York, NY, USA, 123–129. DOI:http://dx.doi.org/10.1145/2896338.2896349

[8] Falcone and Castelfranchi. Trust and transitivity: a complex deceptive relationship. In *Proceedings of 12th AAMAS workshop on trust in agent societies; 2010.*

[9] Rino Falcone and Cristiano Castelfranchi. 2001. The Socio-cognitive Dynamics of Trust: Does Trust Create Trust?. In *Proceedings of the Workshop on Deception, Fraud, and Trust in Agent Societies Held During the Autonomous Agents Conference: Trust in Cyber-societies, Integrating the Human and Artificial Perspectives*. Springer-Verlag, London, UK, UK, 55–72. http://dl.acm.org/citation.cfm?id=646674.701821

[10] J. Fast. 2014. Caregiving and employment in the Canadian context. In *XVIII ISA World Congress of Sociology; 2014.*

[11] Jones Granatyr, Vanderson Botelho, Otto Robert Lessing, Edson Emílio Scalabrin, Jean-Paul Barthès, and Fabrício Enembreck. 2015. Trust and Reputation Models for Multiagent Systems. *ACM Comput. Surv.* 48, 2, Article 27 (Oct. 2015), 42 pages. DOI:http://dx.doi.org/10.1145/2816826

[12] Jina Huh, Rupa Patel, and Wanda Pratt. 2012. Tackling Dilemmas in Supporting 'the Whole Person' in Online Patient Communities. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '12)*. ACM, New York, NY, USA, 923–926. DOI:http://dx.doi.org/10.1145/2207676.2208535

[13] Trung Dong Huynh, Nicholas R. Jennings, and Nigel R. Shadbolt. 2006. An integrated trust and reputation model for open multi-agent systems. *Autonomous Agents and Multi-Agent Systems* 13, 2 (2006), 119–154. DOI:http://dx.doi.org/10.1007/s10458-005-6825-4

[14] Audun Josang and Roslan Ismail. 2002. The beta reputation system. In *In Proceedings of the 15th Bled Conference on Electronic Commerce.*

[15] Patty Kostkova, Vino Mano, Heidi J. Larson, and William S. Schulz. 2016. VAC Medi+Board: Analysing Vaccine Rumours in News and Social Media. In *Proceedings of the 6th International Conference on Digital Health Conference (DH '16)*. ACM, New York, NY, USA, 163–164. DOI:http://dx.doi.org/10.1145/2896338.2896370

[16] Amir H. Razavi, Diana Inkpen, Sasha Uritsky, and Stan Matwin. 2010. Offensive Language Detection Using Multi-level Classification. In *Advances in Artificial Intelligence: 23rd Canadian Conference on Artificial Intelligence, Canadian AI 2010, Ottawa, Canada, May 31 – June 2, 2010. Proceedings*, Atefeh Farzindar and Vlado Kešelj (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 16–27. DOI:http://dx.doi.org/10.1007/978-3-642-13059-5_5

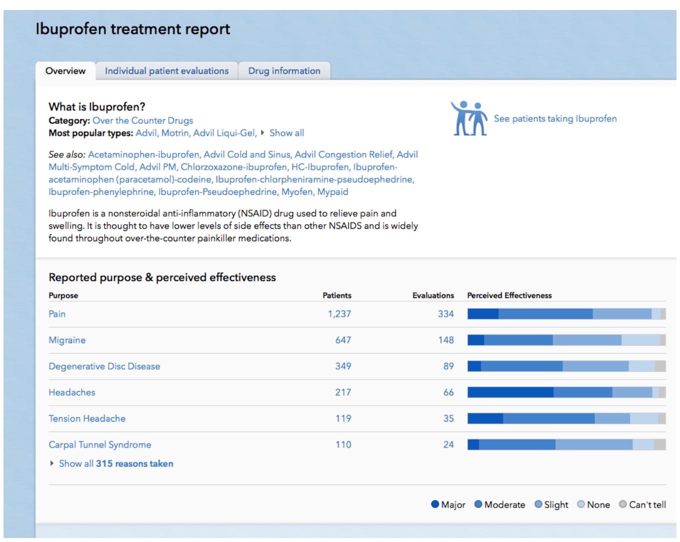
[17] Kevin Regan, Pascal Poupart, and Robin Cohen. 2006. Bayesian Reputation Modeling in E-marketplaces Sensitive to Subjecthity, Deception and Change. In *Proceedings of the 21st National Conference on Artificial Intelligence - Volume 2 (AAAI'06)*. AAAI Press, 1206–1212. http://dl.acm.org/citation.cfm?id=1597348.1597380

[18] Noel Sardana. 2014. *Recommending messages to users in participatory media environments: a Bayesian credibility approach.* Master's thesis. University of Waterloo. http://hdl.handle.net/10012/8311

[19] N. Sardana and R. Cohen. 2014. Demonstrating the value of credibility modeling for trust-based approaches to online message recommendation. In *Proceedings of PST 2014*. 363–370. DOI: http://dx.doi.org/10.1109/PST.2014.6890961
[20] N. Sardana and R. Cohen. 2014. Modeling agent trustworthiness with credibility for message recommendation in social networks. In *Proceedings of AAMAS 2014*. International Foundation for Autonomous Agents and Multiagent Systems, Richland, SC, 1423–1424. http://dl.acm.org/citation.cfm?id=2617388.2617504
[21] N. Sardana and R. Cohen. 2014. Validating trust models against realworld data sets. In *2014 Twelfth Annual International Conference on Privacy, Security and Trust*. 355–362. DOI: http://dx.doi.org/10.1109/PST.2014.6890960
[22] Leandro Silva, Mainack Mondal, Denzil Correa, FabrÃŋcio Benevenuto, and Ingmar Weber. 2016. Analyzing the targets of hate in online social media. In *Proceedings of the 10th International Conference on Web and Social Media, ICWSM 2016*. AAAI Press, 687–690.
[23] W. T. Luke Teacy, Jigar Patel, Nicholas R. Jennings, and Michael Luck. 2006. TRAVOS: Trust and Reputation in the Context of Inaccurate Information Sources. *Autonomous Agents and Multi-Agent Systems* 12, 2 (2006), 183–198. DOI: http://dx.doi.org/10.1007/s10458-006-5952-x
[24] Paul Wicks, Timothy E. Vaughan, Michael P. Massagli, and James Heywood. 2011. Accelerated clinical discovery using self-reported patient data collected online and a patient-matching algorithm. *Nat Biotech* 29, 5 (May 2011), 411–414. http://dx.doi.org/10.1038/nbt.1837 Computational Biology.
[25] Jie Zhang. 2011. A Survey on Trust Management for VANETs. In *Proceedings of the 2011 IEEE International Conference on Advanced Information Networking and Applications (AINA '11)*. IEEE Computer Society, Washington, DC, USA, 105–112. DOI: http://dx.doi.org/10.1109/AINA.2011.86
[26] Jie Zhang and Robin Cohen. 2008. Evaluating the Trustworthiness of Advice About Seller Agents in e-Marketplaces: A Personalized Approach. *Electron. Commer. Rec. Appl.* 7, 3 (November 2008), 330–340. DOI: http://dx.doi.org/10.1016/j.elerap.2008.03.001

A EXAMPLES OF EXCHANGES IN CURRENT SOCIAL NETWORKS

The following are some examples found in the forums in PatientsLikeMe community, with a focus on heart healthcare. We begin with a screenshot showing a report (EX0) and then display below some excerpts from postings viewed on the forum.



EX0: A report on Ibuprofen purposes and perceived effectiveness

<u>EX1: Non-informative comment:</u>

"Hope you're doing better today. Soft hugs and prayers."

<u>EX2: Comment on facts about themselves:</u>

"I have been on amlodipine for quite a few years, typically 10mg. Went down to 7.5 for quite sometime, but now back to 10. I do not have any of those side effects. However, it is also not working real well. I get a lot of BP readings around 150/70 etc. Can?t get the top number down unless I go workout, run or swim distance; or sit and relax for a long period of time, that sometimes work."

<u>EX3: Comment making suggestion on a general approach:</u>

"I have been rolling it on a wooden foot massager and it seems to be helping."

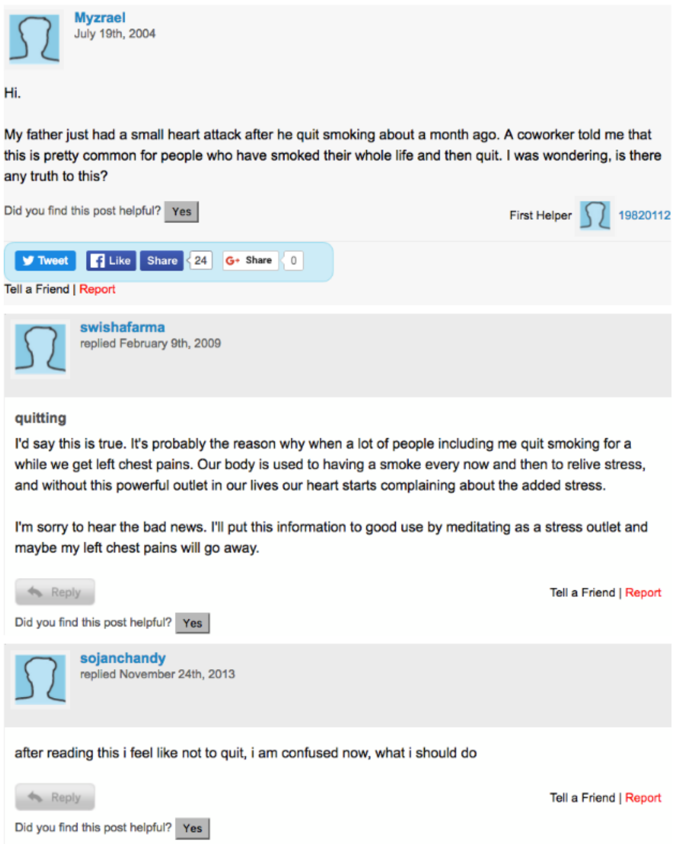
<u>EX4: Comment making suggestion on a specific medicine:</u>

"Have you considered myoqinon CoQ10 to lower your blood pressure naturally to avoid side effects? It also strengthens your heart, increase heart function and reduces pro-bnp."

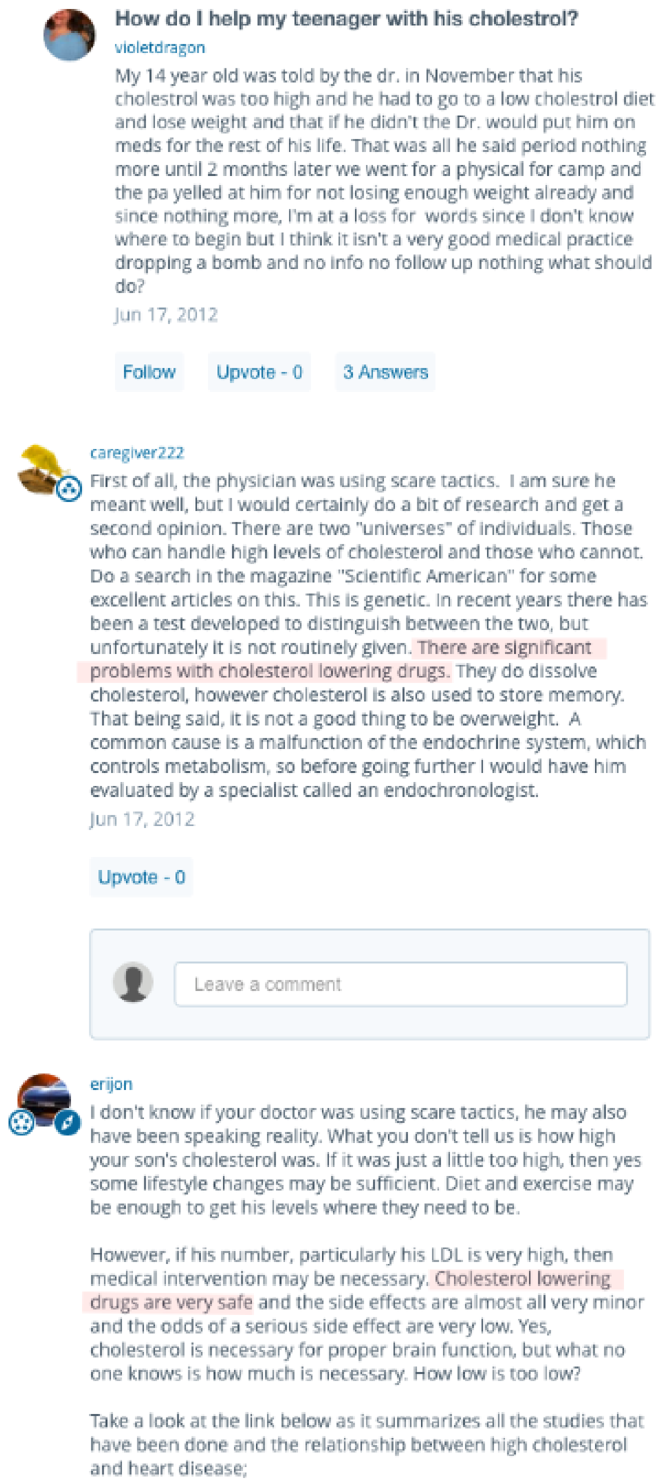
We also found other interesting examples in different forums. EX5 shows a thread on eHealthForum where users were building a case, based on anecdotal evidence, that quitting smoking leads to heart attacks. The example also shows a confused reader, who cannot decide whether or not they should quit smoking. EX6 from MedHelp shows two responses, to the same question about the safety of cholesterol lowering drugs, that offer conflicting information. EX7 was observed on Patient where a user can be seen accepting advice from a stranger about taking an extra beta blocker dose. Finally EX8, shows a user, worried about their health, receive differing advice, which may result in different reactions of concern.

B STATISTICS ON VARIOUS HEALTH FORUMS

Table 1 shows the results of our investigation of online social networks for health. It reports on the age of the forum, its activity level, its most popular topics, and the different trust-related features available.

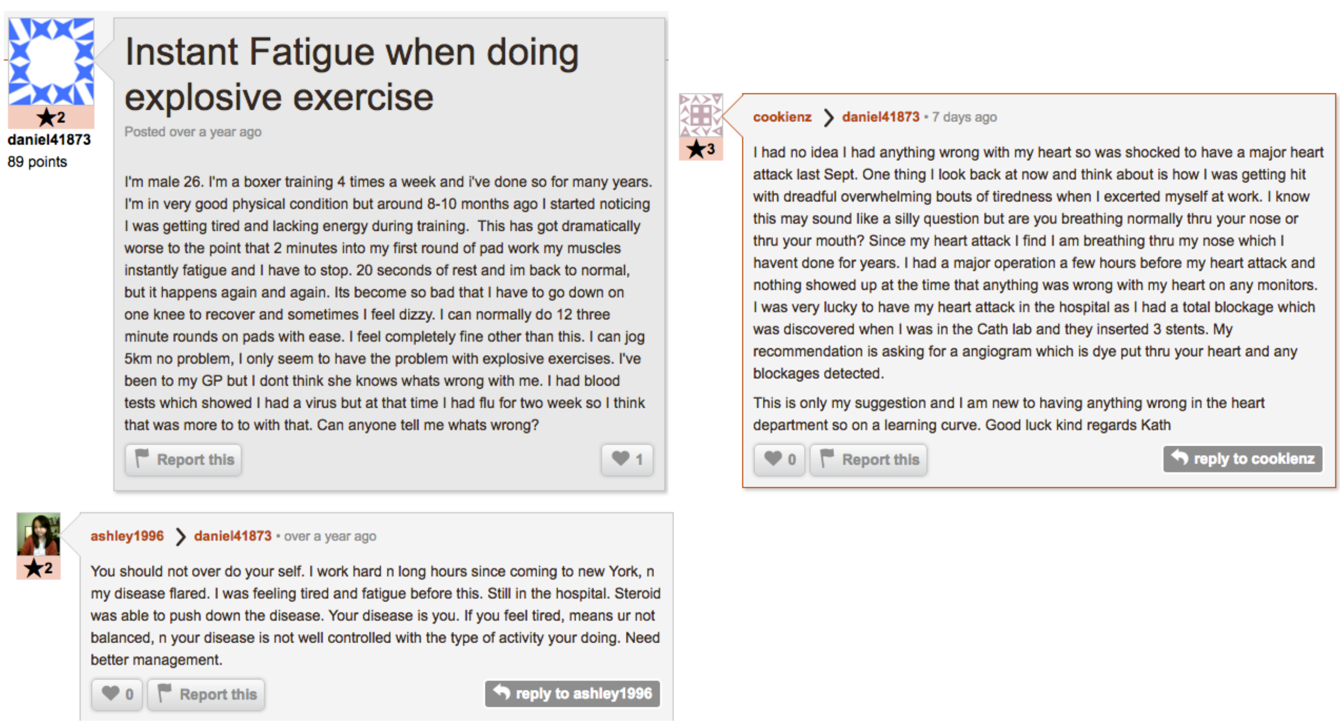


EX5: Anecdotal Evidence that Quitting Smoking Leads to Heartattacks



EX6: Conflicting Advice on Safety of Cholesterol Lowering Drugs

EX7: Accepting Dosage Advice from Stranger



EX8: Differing Advice about Heart Attack

	eHealth forum	Patient	HealthBoards HEALTH MESSAGE BOARDS	talkhealth	MedHelp	patientslikeme	HealingWell.com Community. Support. Resources.
Website	ehealthforum.com	patient.info	healthboards.com	talkhealthpartnership.com	medhelp.org	patientslikeme.com	healingwell.com
Earliest post	2003	2005	2000	2008	2005	2011	2003
Activity	High	Medium	High	Low	High	Low	High
Average Posts/Thread	3.89	N/A	5.48	3.78	N/A	~3.8	9.08
Average Posts/Thread (10 most active topics)	5.57	10.38	6.11	3.86	N/A	N/A	9.67
Total Threads	450,613	>49,558	902,827	3,098	~800,000	~5000	302,067
Total Posts	1,755,034	>514,520	4,954,503	11,696	N/A	~19,000	2,744,150
Most Popular Topics	Pregnancy, Birth Control, Sexual Health, Anxiety and Stress	Anxiety, Hip Replacement, Menopause, PR and GCA	Relationship Health, Thyroid Disorder, Acne, and Back Problems	Eczema, Acne, Allergies, Women's Health	N/A	N/A	Ulcerative Colitis, Lyme Disease, Crohn's Disease,Prostate Cancer
Doctor Verification	✔	✘	✘	✔	✔	✔	✘
				Limited Time Expert Forums	Limited Time Expert Forums	Doctors do not participate in forum	
Ask a Doctor Feature	✔	✘	✘	✔	✔	✘	✘
Account Age Tracker	✔	✔	✔	✔	✔	✔	✔
# of Posts Tracker	✔	✔	✔	✔	✔	✔	✔
Post History	✔	✔	✘	✔	✔	✔	Last 10
Helpful Post Tracker	✔	Blended Point Metric	✘	✘	Best Answers	✔	✘
Report-a-User System	✔	✔	✘	✘	✔	✘	✔

Table1: Statistics on Popular Health Forums