Social Network Analysis of an Online Melanoma Discussion Group
Kathleen T. Durant PhD, Alexa T. McCray PhD, Charles Safran MD
Beth Israel Deaconess Medical Center, Boston, MA
Harvard Medical School, Boston MA

Abstract
We have developed tools to explore social networks that share information in medical forums to better understand the unmet informational needs of patients and family members facing cancer treatments. We define metrics that demonstrate members discussing interleukin-2 receive a stronger response from the melanoma discussion group than a typical topic. The interleukin-2 network has a different topology than the melanoma network, has a higher density, and its members are more likely to have a higher intimacy level with another member and a lower inquisitiveness level than a typical melanoma user. Members are more likely to join the interleukin-2 network to answer a question than in the melanoma network (probability = 0.2 ± 0.05 p-value=.001). Within the melanoma network 20% of the questions posed to the community do not get an answer. In the interleukin-2 network, 1.3% of the questions (one question) do not get a response.

Introduction
Online medical discussion groups give members who are dealing with a particular disease, an opportunity to obtain medical information, coping strategies, as well as support from a community experiencing a similar circumstance. The benefit and utility of a medical forum is potentially twofold: educational and social. The social networks developed through the use of medical forums focus on the sharing of information. Normally, relationships do not exist among the members prior to joining the forum; the relationships are created through the social interactions on the forum. These types of networks have been referred to as knowledge-sharing online social networks (OSN).\(^1\) By sharing knowledge and discussing experiences, members create relationships and form a social network as well as increase their knowledge of the disease. This paradigm is different from other social networking sites that exploit already existing personal relationships. Clinical research has shown that cancer patients who join in-person discussion groups experience a significantly improved quality of life, a significantly reduced pain level\(^2\) as well as a decrease in the three most significant stressors for cancer patients: unwanted aloneness, loss of hope and loss of control.\(^3\) One study reports a decrease in depression and reaction to pain for online support group members;\(^4\) while another reports a decrease in depression as well as perceived stress.\(^5\) Our research defines metrics to quantify different response levels provided by an OSN. It complements the previously cited research.

We construct social networks based on the directed communication between OSN members in order to understand the typical interactions between members. We focus on cancer forums and potentially terminal illness to better understand the informational needs, both met and unmet in a narrow time frame.

We target interleukin-2 (IL-2) since it is prescribed to patients with a more advanced degree of melanoma (stage iv). We wish to determine if the melanoma forum reacts differently to these seriously ill patients. We hypothesize that users posing questions on IL-2 will receive a stronger response from the network.

41% of e-patients have read and 6% have posted medical information or questions to an online web log or discussion group;\(^6\) this usage shows that OSNs are an important informational source to a subset of online patients. We hypothesize applying social network analysis to an OSN will yield insight into the met and unmet needs of its members.

Methodology
Data source: The website www.cancercompass.com, is an online cancer data source sponsored by the Cancer Treatment Centers of America™. Cancer compass has over thirty cancer forums and dates back to 2001. We harvest the posts, threads and users’ data from the melanoma forum using html parsers. Table 1 displays statistics for the corpus collected Sept. 2009.

| User Type  | Care giver | Patient | Survivor | Doctor | Nurse | Member | Total |
|------------|------------|---------|----------|--------|-------|--------|-------|
| Users      | 397        | 310     | 49       | 8      | 87    | 851    |
| Thread Creator | 276   | 178     | 20       | 3      | 62    | 539    |
| Post       | 1175       | 858     | 97       | 122    | 156   | 2408   |

Table 1. The number of users, threads and posts within the melanoma network.
A *user* is a member within the melanoma forum; there are five different *user types*: caregiver, patient, survivor, doctor/nurse and member. Users self-assign a user type when they register at the cancercompass website. A user type typically describes the relationship the user has with melanoma. A *thread* is a discussion found on the melanoma forum; it is a collection of inter-related posts. We assume all threads should have responses. A user who poses a question to the forum is the creator of a thread. When other users respond to the question, their responses are called *posts* and are appended to the thread. This contributing post is a connection to the thread’s creator.

Melanoma social network: A social network consists of nodes and arcs. We represent the melanoma forum as a social network where the nodes represent the members of the forum and the arcs are the directed communication between the forum’s members. The directed communication is from the answerer to the questioner. The relative thickness of an edge represents the number of directed communication between the two members. Non-directed communication such as questions posed to the community-at-large is modeled as an attribute of the user. It is the inquisitiveness level of the user and is represented by the relative size of the user’s node within the network. In general, questions provide a mechanism for members to potentially connect, while answers are the manifestation of a connection between two members. Within the network, a node is created because of a member’s interest in the forum’s topic, whereas an arc is created because of a user’s willingness to actively contribute to a thread’s discussion.

|          | Caregiver | Patient | Survivor | Doctor | Nurse | Member | Total |
|----------|-----------|---------|----------|--------|-------|--------|-------|
| Users    | 133       | 81      | 9        | 2      | 10    |        | 235   |
| Thread Creator | 44       | 23      | 3        | 0      | 5     |        | 75    |
| Post     | 294       | 187     | 15       | 31     | 12    |        | 539   |

**Table 2. The number of users, posts and threads within the IL-2 network.**

IL-2 Network: We identify the threads, posts and users that have discussed IL-2 within the melanoma OSN. Since the network is defined by the concept IL-2, the IL-2 network is not a mere extraction of a sub-network from the melanoma network. The nodes and the arcs are a subset of the melanoma network; however the edge weights and node weights represented by the node size and arc thickness respectively, will vary between the two networks since the number of questions and answers on IL-2 versus all topics may be different for the users within the two networks.

We use the presence of the terms: interleukin, Interleukin-2, IL-2, aldesleukin, proleukin®, as well as the regular expression interleukin? to identify threads discussing IL-2 treatment. The size of the network is described in Table 2.

Procedure: We define metrics that quantify the response provided by a network to its members. We compare a specific topic’s network response metrics to the response metrics found within the total network. We compare the social interactions within the IL-2 threads to the social interactions within the melanoma threads.

Node classes: There are four classes of nodes: nodes that receive information, nodes that provide information, nodes that receive and provide information and nodes that do neither. Nodes that provide information are producers; nodes that receive information are consumers and nodes that provide and receive information are facilitators. The network may service a consumer’s request or not. We define *p-satisfied consumers* as consumers whose needs have a higher probability of being satisfied.

Hub/authority analysis: We identify the influential producers and the *p-satisfied consumers* using hub/authority analysis. A hub node is a node that has many outgoing arcs but very few incoming arcs. An authority node is a node that has many incoming arcs but few outgoing arcs. Since our network models data transfer, the hubs are the users providing the most information, the most active producers. Authorities are the users who ask for information and receive a large response from the network, the *p-satisfied consumers*. Users identified as both a hub and an authority are nodes that have many incoming and outgoing arcs, the most successful *facilitators* in communication. Facilitators pass information and are passed information, they encourage communication.

Metrics: The responsiveness metrics are: the average length of a thread (measured in the number of responses), the average number of days the thread is active (from question posted to last response) and the percentage of unanswered questions.

Visualization: We visualize the networks using the open-source visualization tool: Pajek. All networks were drawn using the Kamada-Kawai algorithm. Kamada-Kawai is a force-directed layout algorithm that attempts to position nodes on the space so that the geometric (Euclidean) distance between the nodes.
is as close as possible to the graph-theoretic (path) distance between them. Nodes at the center of the network are more highly connected to other nodes than nodes found in the periphery of the network.

**Results**

Using the defined response metrics, we show IL-2 threads receive a stronger response from the melanoma network than a typical topic. We show there is a significant difference in the method in which users join the two networks. We also show the topography and the social structures of the two networks are different; however the same influential users are involved within both networks.

Melanoma network: Figures 1 and 2 use both shape and color to distinguish the user types: red square = patient, blue circle = caregiver, salmon circle = survivor, green diamond = doctor/nurse, and purple triangle = member.

**Figure 1. Melanoma Network.**

The melanoma network has two distinct unequal-sized groups that are aligned along the y-coordinate. The left side group has patients centered and caregivers in the periphery. The right side group has a prominent doctor/nurse surrounded by caregivers with patients in the periphery.

IL-2 Network: The star relationship is prominent in the IL-2 network; it resembles a flower in the Kamada-Kawai algorithm. The star relationship identifies nodes that are central in information flow. In the IL-2 network, the star’s central nodes are consumers of information and the non-central nodes are the providers of information. Figure 2 shows many consumers receiving information from multiple sources. Many of the producers are central but others are even more peripheral than the consumer node.

Comparing networks: We differentiate the networks by comparing: 1) the density, 2) the arc weights, 3) the node weights, 4) the initial user action and 5) the effect of user type on activity level and on membership duration in the two networks.

**Figure 2. IL-2 network.**

Using user type as a factor, a nonparametric Kruskal-Wallis ANOVA test determines both number of interactions and length of membership is affected by the user type in the melanoma network (p-value = .01, p-value=.007 respectively). However, within the IL-2 network, the number of interactions and the length of membership are not influenced by the user type. We use the nonparametric Kruskal-Wallis test because the data is not normally distributed.

We consider the initial action a user performs when joining both networks. A user may choose to pose a question or answer a question. Users are more likely to join the IL-2 network to answer a question than in the melanoma network (.2 ±.05 p-value=.001).

The density of a network is defined as the number of lines, expressed as a proportion of the maximum possible number of lines. The density of the IL-2 and the melanoma network is .006 and .002 respectively. The IL-2 network is three times as dense as the melanoma network.

**Figure 3. Frequency distribution of the arc weights, arc weights represent intimacy level.**

Arc weight represents the number of times a particular user communicated to another user. A
higher weight means more communication instances between two users. More communication instances increase the intimacy of the two users. Figure 3 shows the IL-2 network has 8.16% of its arcs with weight > 2 whereas the melanoma network has 4.94% of its arcs with weight > 2. Proportionally, the IL-2 network has more intimate pairs.

Node weight represents the number of questions asked by a user and corresponds to the inquisitiveness level of the user. Members of the melanoma network, in general, are more inquisitive than the IL-2 members. 90.95% of the melanoma consumers ask 2 or fewer questions whereas 97.44% of the IL-2 consumers ask 2 or fewer questions. Only 29% of the IL-2 consumers ask more than 1 question compared to 49% of the melanoma consumers.

Inquisitive level of users

![Inquisitive level of users](image)

Figure 4. Frequency distribution of the node weights, node weight represents inquisitive level.

Response Variables: Our fundamental measurement of response is receiving at least one answer from a community member for a question broadcasted to the community. Within the melanoma network 20% of the questions posed to the community do not get an answer. In the IL-2 network, 1.3% of the questions (one question) do not get a response (Table 3, row 1).

|                      | Melanoma | IL-2 |
|----------------------|----------|------|
| No response percent  | 20%      | 1.3% |
| Users per thread     |          |      |
| Median, Mean         | 3, 4.468 | 6, 7.18 |
| Standard Deviation   | 3.906    | 4.8314 |
| Threads’ active days |          |      |
| Median, Mean         | 12, 83.29 | 55, 165 |
| Standard Deviation   | 205.3466 | 308  |

Table 3. Response variables for the networks.

The number of users per thread (Table 3, row 2) shows twice as many users respond to questions on IL-2 than on a typical topic. The number of days a thread is active (Table 3 row 3), shows that IL-2 threads are a source of communication for a longer period of time.

Influential users: We identify influential users using hub/authority analysis of the network. The influential producers are colored red, the p-satisfied consumers are colored yellow and the facilitators are colored green. In both networks, the same one facilitator and three influential providers were identified. Out of the eight p-satisfied consumers found in the IL-2 network, three were also identified as p-satisfied consumers in the top 15 p-satisfied consumers of the melanoma network.

![Influential users](image)

Figure 5. A grayscale magnification of the center of the melanoma network with red, most active provider nodes, a green facilitator node and yellow, p-satisfied consumer nodes.

![Influential users](image)

Figure 6. A grayscale magnification of the center of the IL-2 network (nodes as defined in Figure 5).

Discussion

High-dose IL-2 therapy is a very difficult process; its patients experience a variety of side effects ranging from tachycardia to peeling skin. It has been shown to be extremely successful in a small percentage of stage iv melanoma patients; its 5-year survival rate is 10% versus 2% for chemotherapy. While IL-2 provides hope for patients with an almost terminal illness, this treatment is a great source of distress for patients and their families. Users experiencing IL-2 treatment need more support than the norm and may receive it from...
an OSN. Users seeking IL-2 information get a stronger response from the OSN in terms of the number of people reacting to the questioner, the likelihood of receiving an answer, and the length in time the thread stays active.

Users providing the strongest response (facilitator and producers) are the same in both networks (Figures 5 and 6), revealing the importance of these users within the community. The facilitator node supports a large portion of the melanoma network and has no other influential neighbors. Since it is a facilitator it not only provides information, it also receives information from its surrounding non-influential nodes. The facilitator is a patient and most of its neighbors are also patients, this similar characteristic may explain its neighbors’ willingness to perform directed communication. The highest-ranking producer (doctor) has two neighbors that are also influential producers (both caregivers). All three nodes provide information but have not successfully encouraged conversation between themselves and the surrounding consumers.

Despite the support provided by the influential providers, 20% of the questions posed to the melanoma community do not get an answer. In the IL-2 network, 1.3% of the questions (one question) do not get a response. We determined the unanswered question asks for opinions on the difference between IL-2 treatment and TK1258 treatment for ocular melanoma. Ocular melanoma is a rare form of melanoma with an estimated incidence of 6 per one million per year.\(^\text{11}\) This is a question whose answer may not be known within the community.

IL-2 consumers receive more responses because IL-2 network members are less dependent on receiving a response from an influential user. The IL-2 network is more densely connected than the melanoma network; this means the users responding to the questions are more varied. Many of the providers are found in the periphery (non-central nodes in the star relation Figure 2), showing that many consumers are more central to the IL-2 network than some providers. A user is more likely to join the IL-2 network to answer a question; this action provides support to an existing member of the IL-2 community.

Conclusion
We have demonstrated that SN tools can visualize the interactions within an OSN and provide insight into its communication patterns. We defined novel metrics and applied well-established SN analysis techniques to demonstrate the differences in an OSN’s responsiveness to a particular topic (IL-2). We plan to characterize threads receiving a low or no response from the melanoma forum.

We have shown that one topic (IL-2) in our OSN elicits a strong response when compared to the overall network. Understanding the relative strengths of a response for different topics within an OSN may indicate targets of opportunity for information interventions on behalf of patients and their families.

References
1. Guo L, Tan E, Chen S, Zhang X, Zhao Y. Analyzing patterns of user content generation in online social networks. Proceedings of the 15th ACM SIGKDD International conference on knowledge discovery and data mining. 2009; 369-378.
2. Goodwin PJ, Leszcz M, Ennis M, et al. The effect of group psychosocial support on survival in metastatic breast cancer. N Eng J Med 2001, Dec;345:1719-26.
3. Norton TR, Manne SL, Rubin S, et al. Prevalence and predictors of psychological distress among women with ovarian cancer. J Clin Oncol, 2004 Mar 1;22(5);919:26.
4. Lieberman M, Golant M, Giese-Davis J, et al. Electronic support groups for breast carcinoma: a clinical trial of effectiveness. Cancer 2003; 97(4):920-5.
5. Winzelberg AJ, Classen, C, Alpers GW, et al. Evaluation of an internet support group for women with breast cancer. Cancer. 2003; 98(5):1164-1173.
6. Fox S, Jones S. The social life of health information. Washington DC: Pew Internet and American Life Project. 2009.
7. Nooy W, Mrvar A, Batagelj V. Exploratory Social Network Analysis with Pajek. New York:Cambridge Press; 2005. 334 p.
8. Kleinberg JM. Hubs, authorities and communities. ACM Computing Surveys. 1999;31(4).
9. Kamada T, Kawai S. An algorithm for Drawing general undirected graphs. Information Processing Letters, 1988;31:7-15.
10. Dwyer T, Koren Y, Marriot K. Digi-CoLA: directed graph layout through constrained energy minimization. IEEE Symposium on Information Visualization. 2005;65-72.
11. Hearle N, Damato BE, Humphreys J, et al. Contribution of germlike mutations in BRCA2, P16(INK4A), p14(ARF) and P15 to uveal melanoma. Investigative Ophthalmology & Visual Science. 2003 Feb 44(2):458-462.