Sustainability of Social Media

Design, Engagement, and Business Models
Social Media

Computer mediated social interactions

Symbiosis between the technology and humans

Users interactions and relationships are facilitated and expressed through digital technology
People are attracted to other people

+500 M registered users

Role playing in the virtual world: practicing exchange and reciprocity
Between 1.4M and 2.6M players at any moment in time.
User profile. Can see online status for all the friends. Chat capability. Form a social and organizational structure.
Real time communication through live audio.
Engineering Social Platforms

Computing Infrastructure
http://www.vaqueronet.com/colocating.php
http://www.toptechreviews.net/tech-news/facebook-users-to-delete-their-fb-account/

Service Operations
http://www.alpheuscommunications.com
Engineering Social Platforms

Social Network

- Cameron Marlow
  - Facebook
  - co-workers

- Leah Pearlman
  - Facebook
  - co-workers

- Meredith Chin
  - Facebook
  - co-workers

Technology Infrastructure and Support

Computing Infrastructure
- http://www.toptechreviews.net/tech-news/facebook-users-to-delete-their-fb-account/

Service Operations
- http://www.alpheuscommunications.com
How do we ensure healthy and sustainable online community?

Without clear indicators of the community health it is not possible to measure the effectiveness of the designs and increase the utility of the service.

We need methods for domain experts who can use computational tools to analyze and interpret social phenomena.
SOCIAL INTERACTION DESIGN
Community Question-Answering

Online Communities

- Blogs
- Forums
- Newsgroups
- Web Boards
- Distribution Lists
- Question Answering

- Live Search QnA Beta
- askville by amazon
- Yahoo! Answers
- answerbag
- Google Answers
- NAVER 지식인

- 2008
- 2006
- 2005
- 2003
- 2002-06
- 2002
Community Question-Answering

Question

Denis asked

Are dogs really color blind, how do we know that?

Posted 3 days ago
dogs, dog, pets, Animals

Answers

qb666 answered

Dogs do see in color, but their perception of color is not the same as it is for people. They cannot distinguish between red, orange, yellow or green. They can see various shades of blue and can differentiate between closely related shades of gray that are not distinguishable to people. This was copied from Petplace.com.

Posted 13 hours ago

Add comment
Welcome to QnA
Show the world what you know — and find out what you don’t. Get answers to your most burning questions… share info on your area of expertise… earn fame and glory — all compliments of the QnA community.

Featured Question
What do you think of our switch to digital television?
Posted 3 days ago | 2 days left for answering | 7 Answers
Tags: People, by, politics, software, digital, conversational, digital switch, digital television

Featured Question
What do you think of our switch to digital television?

Latest questions
- Does your boss check up on your Facebook or MySpace pages?
  - Open | Asked by Orbebleu 5 minutes ago | 2 Answers | 4 days left
  - Tags: Internet, Business, opinion, face book, my space, ...

- How many of you have a "PRIVATE" drawer in your bedroom?
  - Open | Asked by Chef K 8 minutes ago | 5 Answers | 4 days left
  - Tags: People, Entertainment, fun, conversational, grown-ups

- How do you type with the tilde so that it appears over the n?
  - Open | Asked by steve 12 minutes ago | 4 Answers | 4 days left
  - Tags: Spanish, characters

- Video of CES Steve Ballmer Keynote?
  - Open | Asked by Kay 9 minutes ago | 0 Answers | 4 days left
  - Tags: Internet, software

- Is Lisa Foropoulos being searched for on the internet?
  - Open | Asked by livyanes29 9 minutes ago | 0 Answers | 4 days left
  - Tags: Internet, software
Is running still the best cardiovascular health exercise? I do a lot of gym and fitness weight training, but running is a fantastic cardio but stay at home exercise. I've been using a pair of running shoes and go running with my dog 3-4 times a week. It is great, but I am not a runner. I just started running in January and I'm still learning. The best part is the fresh air and view of the ocean. The only downside is the smell of sweat. It's a must if you have a dog! I love running! I'm 37, and I'm getting healthier with every run. Need more votes, please. Thank you.
Social Interactions in Q&A Systems

Entities:
• Users, questions, answers, comments, votes

Types of social interactions:
• Answer to a question
• Comment to answer
• Vote on best answer
Attracting and Retaining Users is Key to Sustainability

Support novice users in becoming active community participants
Support frequent users in increasing the volume and quality of their content contributions
Promote high quality contributions, for external exploitation—through search.

85% of new users start with a question
72% never ask a question again
5% will engage in answering

61% of questions from new users don’t get more than 1 answer
(23% get 0 answers)
Incentives for Best Answer Voting

• Live Q&A rewards for participation: authoring and voting
• Points are used for reputation scoring

<table>
<thead>
<tr>
<th></th>
<th>Voter</th>
<th>Author</th>
</tr>
</thead>
<tbody>
<tr>
<td>Any answer</td>
<td>1 point</td>
<td>5 points</td>
</tr>
<tr>
<td>Best answer</td>
<td>4 points</td>
<td>20 points</td>
</tr>
</tbody>
</table>
Selecting and Rewarding the Best Answer

Common practice: plurality vote. In some designs, votes from the person asking for answer counts more.

Problems:

Ambiguity — What consists a good answer? Something that is correct or useful or entertaining?

Social bias — Both the asker and answerer are identifiable. The vote may reflect a social tie rather than answer quality.

Personal gain — One may attempt to exploit the incentive scheme related to voting participation:

• Gain points by assigning superficial votes to answers, randomly
• Coordinate voting with other participants to get the best answer reward.
From the QnA Data Logs

Evidence of collusion.

“Microsoft or Apple? Feel free to argue and point out their good and bad points. Also feel free to rebut or debate on other people's standpoint. Best argument/answer will get my friends’ and my "best answer" reward.”

Individuals’ aspirations to excel in their social status can adversely affect the quality of their contribution to the community.
Truth and quality determined by a collective awareness and acceptance:

*Information is good if it is provided by trusted sources.*

*Sources are trusted if they provide good information.*

How to define the measures of goodness and trustfulness in a way that cannot be easily tempered with?
Problem Formulation

• We consider a set of questions $Q$. Each question can have one or more answers.

• Each answer may get a number of votes (towards the award for the best answer)

• Define a **voter score function** that assigns a score $r_i$ to each voter so that for each 2 voters their relative score depends on their track record across all the answers in the following way:

The relative scores correspond to the ratio of the total number of votes $\mu$ that the other voters gave to the answers they voted for:

$$r_i : r_j = \mu (a_i) : \mu (a_j)$$
Voter Score Function

It turns out that the function satisfies the condition

$$r_i = \sqrt{\frac{\sum_{\{v_k \in V_q\} \{r_k : a_k = a_i\}}}{\sum_{\{v_k \in V_q\}} r_k}}$$

The voter score is essentially a normalized sum (average) of the scores of all other voters who cast the vote on the same answer. The reliability scores represent a fixed-point for the function $\mathcal{F}$. Thus, we can apply the Brouwer Fixed Point Theorem.
Experiment

• Random Voting Simulation
  – Model increased frequency but superficial voting by individual voters to get more participation points

• Ballot Stuffing Simulation
  – Model coordinate voting to influence the best answer outcome and get points for the best answer vote.

• Data: from MSN Q&A service data, we collect
  – 488,760 questions, 1,330,819 answers
  – 1,599,994 votes, 256,950 users

• Experiments with both sample data and synthetic data
  – Synthetic data was generated based on the best fit distribution.
Random Voting

Simulate Random Voting by uniform distribution in place of Zipf’s Law
We vary the percentage of affected questions (from 1% to 10%) and the percentage of voters who voted randomly (from 1% to 10%).

The number of “best answer awards” changed due to spamming is lower for fixed point score (right) than for plurality voting (left)
Coordinated Voting (Ballot Stuffing)

- Simulate the collusion: fix the number of involved voters (‘stuffers’, here 4 and 10) and the percentage of questions affected (here 50%)
- Both majority voting and fixed point scoring are susceptible to ballot stuffing
- Fixed point scoring flags out the outliers and helps identifying collusion
engagement analysis

SOCIAL ANALYSIS AND SELF-AWARENESS
What are Social Networks?

**Sociologists**
Network of individuals connected through social ties.

**Computer scientists**
Graph structure that emerges from the social interaction among individuals.

The graph may be related to explicit or implicit relationships between individuals and objects involved in the interaction.
Network graphs are not for everyone?

Facebook Social Graph
Explore your social graph to discover social clusters and photo friends on Facebook.
NodeXL: Network Overview, Discovery and Exploration for Excel

• Leverage spreadsheet for storage of edge and vertex data

• http://www.codeplex.com/nodexl
NodeXL Workbook First
From UCINET Full Matrix DL File...
From GraphML file...
From Pajek File...
From Open Matrix Workbook...
From Open Edge Workbook...
From Email Network...
From Flickr Related Tags Network...
From Flickr User’s Network...
From Twitter Search Network...
From Twitter User’s Network...
From Web 1.0/Blog Network (via VOSON)...
From YouTube User’s Network...
From YouTube Video Network...
From NodeXL Workbook Created on Another...
TWITTER NodeXL Graph of #facsumm
Data import
Data post-processing
Data layout
Metrics computation
Mapping metrics onto visual properties
Filtering, Motifs, Clustering

GRAPH METRICS
Closeness centrality
Betweenness centrality
Eigenvalue centrality
Clustering coefficients
Graph readability metrics
TWITTER NodeXL Graph

“Bing” at 2:30 AM Monday, July 12, 2010
Recognition or Definition of Social Roles?

- Experts and “Answer People”
- Discussion starters, Topic setters
- Discussion people, Topic setters

Bellagio Fountains in action

The animated fountains synchronized with music in front of the Bellagio Hotel

Comments and faves

Nancy Baym
Next time invite me to THIS one! :)

eszter pro
Nice colors.
Human Super-Connectors

Flickr users who commented on Marc_Smith’s photos (more than 4 times)
Human Super-Connectors

Flickr users who commented on Marc_Smith’s photos (more than 4 times)
Flickr – Network Analysis
Flickr – Network Analysis

User-photo-tag network
user A assigns tag T to photo P
Filtering: US Senate Voting

NodeXL – Open source project. Extension of Excel for network visualization.
http://nodexl.codeplex.com/
Filtering: US Senate Voting

NodeXL – Open source project. Extension of Excel for network visualization.
http://nodexl.codeplex.com/

Include only links that reflect more than 65% between two senators.
Filtering: US Senate Voting

NodeXL – Open source project. Extension of Excel for network visualization.  
http://nodexl.codeplex.com/

Include party affiliation: Red for Republicans, Blue Democrats, 2 other parties.
teaching network analysis

ACCESSIBILITY OF TOOLS AND METHODS
They were enjoined to aspire to the four principles of NetViz Nirvana:

1. Every node is visible
2. Every node’s degree is countable
3. Every edge can be followed from source to destination
4. Clusters and outliers are identifiable.
Social Network Analysis by Novice NodeXL Users

Study: Information Systems Students

Over 5 weeks, 15 Information Science (IS) graduate students were introduced to SNA and NodeXL NodeXL Tutorial – three weeks, one 3-hour lecture per week.

SNA was considered as one of the ways to study social interactions.

We closely monitored 7 students as they complete their SNA assignments over several sessions.

D. L. Hansen, D. Rotman, E. Bonsignore, N. Milić-Frayling, E. Mendes Rodrigues, M. Smith, B. Shneiderman, T. Capone (2009) Do you know the way to SNA?: A process model for analyzing and visualizing social media data”Submitted Journal CSCW


Example: Listserv community

“My initial hypothesis: A leader should possess a high betweenness centrality and a high eigenvector centrality”

Green: Leader
Brown: Regular
Red: Novice
Larger nodes are further along in the membership lifecycle

Vertex Color = Betweenness Centrality
Vertex Size = Eigenvector Centrality
(1) identify if the correct individuals are currently serving as administrators in the community
(2) would other individuals would perhaps be a better fit for those roles

Alternative ‘Admin’ Role

Vertex Color = Betweenness Centrality
Vertex Size = Eigenvector Centrality
NodeXL supports clustering of nodes. In the graph, distinct groups are assigned different colours.
‘Group-in-a-box’ layout combines graph clustering with the tree-maps layout
The same data set in different layouts:

1. Fruchteman-Reingo
2. Grid
3. Spiral.
Fundamental issues
Understanding of the tools that are applied
Understanding of the applied analysis
Understanding the relationship between the analysis and the layout algorithm used in visualization.

Visualization using Harel-Koren layout algorithm.

Sierpinski triangle is a fractal named after Waclaw Sierpinski who described it in 1915.

It is an example of self-similar sets, a mathematically generated pattern that is replicated at every level of reduction or magnification.
Propagation Phenomena in Social Networks

Public Timeline

atemolina: Hoy empieza la preventa de Luis Miguel y yo pobre!!! #tache
less than 10 seconds ago via UberTwitter

RobyBerto: saben algo?
less than 10 seconds ago via web

aarongleeman: Sounds like trade with Orioles is all but complete: J.J. Hardy and Brendan Harris for James Hoey and Brett Jacobson.
less than 10 seconds ago via twitterfeed

catheric9: Jessica Simpson's Fashion Line Worth $1 Billion. For fall 2011, the singer is launching a contemporary sportswear...
less than 10 seconds ago via web

Giisa_Torres: you sai!
less than 10 seconds ago via web

MrOTG: If you have exposure, and you have a single out... and your album don't sell... seems like I would be in the lab making better music.
less than 10 seconds ago via TweetCaster

“Rumour” Tweets

FreePavarotti: I don’t know who started this rumor but I’ve gotten a lot of replies asking if I’m not a canary, people! #freePavarotti
about 18 hours ago via web - Reply · View Tweet
10+ recent retweets

TheQueenBritney: EXCLUSIVE/RUMOR - “Don’t Hold It Against Me” Lyrics http://spearsnibbe.com/?p=350 #britney
1 day ago via web - Reply · View Tweet
10+ recent retweets

breakfastcereal: @angclayee: Rumor Report on right now in the middle of the People’s Choice Mix! Listen live on power1051fm.com breakfastcereal
1 day ago via Twitter for iPhone - Reply · View Tweet · Go to Show Conversation
10+ recent retweets

anusfrance: #Apple Siri. Apple Rumor is Beneficial. No Matter How It’s Cut - Seeking Alpha http://chilp.it/1eb0a5 #hair #best www.HAIRY.com
half a minute ago via TweetDeck - Reply · View Tweet

thesixthnewkidd: @DonnieS_D: I guess Brian said in an interview that some of the nk’s are cruisin. (Could be a rumor) I don’t know! I Know NKs cruiser is n may
half a minute ago via web - Reply · View Tweet

less than a minute ago via TweetMeme - Reply · View Tweet
Analysis of Rumours by Keywords
Rumour Propagation Traces
Star
Analysis and Visualization of Propagation

Green edges connect the central node to neighbors activated before it
Red edges connect the central node to neighbors activated after it.
Static graph insufficient visualization of string--Need time!
Unified Temporal and Spatial Encoding

Encode propagation steps and network attachment

Explore commonality of dynamic motifs:
Identify frequent maximal repeats in the strings.
Identify K-common substrings – substrings that occur in at least K traces.
Analysis and Visualization of Propagation

\[ P(\text{inverse star}) = \langle S_0^0, C_0^1, C_0^2, \ldots, C_0^{14}, O_1^1, B_1^2 \rangle \]

\[ P(\text{star}) = \langle S_0^0, H_0^1, B_0^1, B_0^1, B_0^1, B_0^1, \ldots, B_0^1 \rangle \]
Effective and sustainable online community requires expertise from non-technical fields: sociology, psychology, behavioural economics, and related.

Provide tools for domain experts to enable them to analyze and assist with shaping the social ecosystem.
Empowering the users

In technology mediated social environments, the users rely on technology to express themselves and make sense about the community they are participated in.

Empower users with tools that help them increase their self-awareness and support them in their decisions and engagement.

Applications and services leave traces of personal data that the user should be aware and can analyze, at least to the degree that others are doing already.
Thank you

Natasa Milic-Frayling
natasamf@microsoft.com